Theory of Mind and Rule Use in Individuals With Down’s Syndrome: A Test of the Uniqueness and Specificity Claims

Philip David Zelazo
University of Toronto, Canada

Jacob A. Burack
McGill University, Canada

Elizabeth Benedetto
Ontario Institute for Studies in Education, Canada

Douglas Frye
New York University, U.S.A.

The relationship between Theory of Mind (ToM) and rule use was explored in adults with Down’s Syndrome (DS) and in non-handicapped pre-schoolers. Twelve low-functioning individuals with DS (mean mental age = 5.1 years, mean chronological age = 22.7) performed worse than 12 MA-matched non-handicapped children (mean MA = 5.1 years) on several standard ToM tasks and on a color-shape card-sorting task in which subjects were required to switch between two incompatible sets of rules. On the ToM tasks, people with DS tended to focus on a single state of affairs (e.g. the present situation). Likewise, on the card sort, these subjects tended to use a single set of rules on all trials. Performance in the two types of task was positively correlated when MA was partialled out. The results are inconsistent with the claim that ToM reflects a domain-specific psychological function and the notion that deficits in ToM are unique to individuals with autism.

Keywords: Theory of Mind, Down’s Syndrome, specificity, rule use

Psychologists have long recognized the importance of the ability to consider one’s own mental states in relation to those of other people. For example, Baldwin (1897) wrote extensively on the topic, now called ‘Theory of Mind’ (ToM) and, as early as 1923, Piaget (1923/1955) characterized the growth of logical thinking as a movement away from unreflective egocentrism towards a ‘de-centered’ view of reality that accommodates multiple perspectives. These psychologists recognized that a ToM permits widespread increases in self-understanding, self-control and social interaction.

A renewed interest in ToM has arisen in part because of the suggestion that autism may be due to deficits in social and self-understanding. Baron-Cohen, Leslie and Frith (1985) found that high-functioning children with autism failed a false belief task devised by Wimmer and Perner (1983), whereas non-handicapped 4-year-olds and people with Down’s Syndrome (DS) did not. Based on these and related findings (e.g. Baron-Cohen, Leslie & Frith, 1986; Leslie & Frith, 1988), two main claims have been made — one regarding the nature of various clinical syndromes, and one about the structure of ToM in general (i.e. not just in autism).

The first claim, referred to here as the Uniqueness Claim, is that deficits in ToM are unique to people with autism and can account for the pattern of pervasive impairments that characterize these individuals (Baron-Cohen et al., 1985; Leslie & Thaiss, 1992). However, as ToM has not been studied systematically in other clinical groups, the veracity of the Uniqueness Claim remains to be assessed.

The second claim, called the Specificity Claim, is that ToM (in both typical and atypical individuals) is a domain-specific psychological function. Several theories of ToM embrace one or another form of this claim. For example, in Leslie’s accounts (e.g. Leslie, 1991), ToM reflects an innate module that develops independently from other types of understanding. The Specificity Claim has received its strongest support from studies indicating that people with autism have impairments in ToM that are unrelated to their functioning in other domains — even domains that would seem prima facie to be similar (e.g. Baron-Cohen et al., 1986; Leslie & Thaiss, 1992).

The Specificity Claim has not gone unchallenged. A number of theorists have argued that problems with ToM derive from general impairments in flexible reasoning.
and the execution of strategic behavior (Frye, Zelazo & Palfai, in press; Hughes & Russell, 1993; Ozonoff, Pennington & Rogers, 1991). One of these domain-general approaches is the cognitive complexity and control theory described by Zelazo and Frye (in press), according to which ToM depends on the ability to use rules of a particular complexity. On this account, the ToM tasks that 3-year-olds find difficult require children to reason first from one perspective and then from an incompatible one. Thus, two conflicting judgements about a single, unchanged situation must be nested or embedded under a general rule governing which judgement to make. Failure to consider the perspective as well as the type of judgement will result in perseveration on the more salient perspective. This theory makes specific predictions about the range of tasks with which someone who fails ToM will have difficulty. The finding of specific deficits in ToM could be due to failure to control for the complexity of the tasks assessing reasoning in different domains.

One purpose of the present study was to test the Specificity Claim by comparing performance on ToM to performance on an equally complex nonsocial rule-use task. Another aim was to test the Uniqueness Claim by examining ToM in people with a syndrome other than autism. We chose to study adults with DS because they are an easily identifiable group with cognitive and behavioral profiles distinct from persons with autism. Further, several previous findings encourage direct investigation of ToM in people with DS. Poor performance on ToM has been documented in people with mental retardation of unknown (Benson, Abbeduto, Short, Nuccio & Mass, 1993) and mixed etiologies (Yirmiya & Shulman, 1994). These findings are inconsistent with the Uniqueness Claim, but the heterogeneity of the subjects involved makes comparison between groups difficult and hinders replication. On the other hand, investigation of ToM in people with DS will permit a straightforward assessment of the Uniqueness Claim. If people with DS exhibit impairments in ToM relative to non-handicapped people who are matched on mental age, then a version of the claim must be false: deficits and/or delays in ToM are not unique to autism. That is, if people with DS are impaired in ToM, then difficulties with ToM per se cannot account for the particular sequelae of autism.

Comparison of ToM and nonsocial rule use in people with DS will also allow a further test of the Specificity Claim. A positive relation between these tasks would suggest that difficulties in ToM are not restricted to understanding mental states, but derive from general constraints on the reasoning involved. Perhaps the Specificity Claim only applies in the case of particular disorders, such as autism.

Method

Subjects

Twelve subjects with DS (six males, six females) were recruited from special schools and residential centers and 12 non-handicapped subjects (seven males, five females) were drawn from a public school in a middle-class section of Montreal. The mean chronological age (CA) of subjects with DS was 22.7 years (SD = 4.6, range: 16–30.9); for non-handicapped subjects it was 5.9 years (SD = 0.6, range: 5.2–6.8). Subjects were selected on the basis of mental age (MA), first language (English) and willingness to participate. The two groups were matched on receptive verbal MA as assessed by the Peabody Picture Vocabulary Test — Revised (PPVT-R; Dunn & Dunn, 1981). For subjects with DS, the mean MA was 5.1 years (SD = 0.9, range: 3.9–6.3); and it was 5.1 years (SD = 0.9, range: 4.0–6.3) for non-handicapped subjects. The PPVT-R was used because picture vocabulary tests have been used extensively as a basis for matching in previous studies (e.g. Leslie & Frith, 1988; Leslie & Thiass, 1992; Sodian & Frith, 1992).

Procedure

All subjects were tested individually at their school or center in two 30 min sessions that were no more than 2 weeks apart. Subjects sat across from a primary examiner while a second experimenter scored performance. In the first session, subjects received the PPVT—R. Then they received the card sort and three ToM questions (false belief, appearance–reality and representational change) for each of two items. Phase 1 of the card sort was presented and then there was a break wherein subjects received one of the ToM items (the crayon box). Subjects were then given Phase 2 of the card sort, which was followed by the other ToM item (the horse). Phases 3, 4 and 5 of the card sort were then presented with no breaks between them. The pretend–reality task was administered during the second session along with additional tasks not relevant to the present paper.

The card sort. This task was taken directly from Frye et al. (in press). Eighteen drawings (cards) were sorted into two trays. A red triangle and a blue circle served as standards and were displayed continuously. The remaining cards (test cards) included blue triangles and red circles. The task was divided into five phases: two were learning phases and three phases comprised the card sort proper. In Phase 1, subjects were shown the two standard cards, labeled by color and shape (e.g. ‘Here’s a red triangle’). One dimension (e.g. color) was highlighted and the corresponding rules were stated at least three times, in four different forms (namely: ‘If X, then Y’; ‘All X ...’; ‘Only X ...’; ‘No X ...’). Then, subjects were shown two demonstration cards (one from each category), which the experimenter sorted interactively. At this point, the rules were stated again: ‘If it’s a red one, then it goes there, but if it’s a blue one, it goes here,’ and subjects were shown five randomly selected test cards. For each card, the experimenter stated the rules, labeled the card by both dimensions and asked, ‘Where does this go in the color game?’ Subjects then either placed the card into a tray or indicated a tray and were either told: ‘That’s right ... (e.g. the blue ones go here; you’re good at the color game)’, or ‘No, that’s not right ... (e.g. put all the blue ones here; that’s how you play the color game).’ The relevant dimension (color or shape) in this phase was matched.

Phase 2 was exactly like Phase 1, except that the other rule (e.g. shape) was used. Phases 3, 4 and 5 comprised the card sort proper. In Phase 3, the experimenter reminded subjects of the last game played (i.e. in Phase 2). Then, for each of five test cards, she stated the rules, labeled the card by color and shape, and asked subjects: ‘Where does this go in the shape game?’ Subjects were not told whether or not they sorted cards correctly.

In Phase 4, subjects were told: ‘OK, now we’re going to switch and play a different game, the ___ (e.g. color game). Remember the color game? In the color game, if it’s red it goes here, but blue ones go there. You have to pay attention’. Five test cards were presented exactly as in Phase 3 (although
the experimenter stated appropriate rules). In Phase 5, subjects were required to switch between rules in each of four trials. Each switch was marked explicitly, and the experimenter stated the relevant rules in every trial.

**Appearance-reality, false belief and representational change.** These tasks were administered as in previous studies (see Gopnik & Astington, 1988). Two ‘deceptive’ items were used: (1) a crayon box containing straws; and (2) a blue drawing of a horse that could be covered by red cellophane, which made it look black. Each item was presented first in its deceptive state. After its real state was revealed, the item was returned to its deceptive state and put out of reach, at which point three ToM questions — appearance-reality, false belief and representational change — were asked in a counterbalanced order. For example, for the first item, subjects were shown the crayon box, which was then opened to reveal straws. A straw and a crayon were placed in front of subjects who were asked: ‘What’s in the box?’ Subjects could respond either verbally, or by pointing to the straw or the crayon. The box was then closed and placed out of reach and the test questions were administered.

The test questions (illustrated using the first item) were as follows. For false belief, subjects were shown a puppet and asked: ‘What does she think is in it, straws or crayons?’ For representational change, subjects were asked: ‘When you first saw this, before you opened it up and looked inside, what did you think was in it, straws or crayons?’ The appearance and the reality questions were counterbalanced across items: ‘What’s really and truly in the box, sticks or crayons?’ (reality) and: ‘What does this look like it has in it right now, straws or crayons?’ (appearance).

**Pretend-reality.** This task was taken directly from Flavell, Flavell and Green (1987). After a training task, the experimenter introduced two items, a piece of string and a pair of sun-glasses, one at a time. Subjects were prompted to pretend that the string was a snake and the sun-glasses were a telephone. While subjects were pretending with each item, the experimenter provided them with two appropriate pictures (string and snake, or glasses and telephone) and asked them: (1) what the item really was; and (2) what they were pretending it was. For example, for the string, the experimenter asked, ‘What’s that really and truly, a string or a snake? OK, and what are you pretending it is right now, are you pretending it’s a string or are you pretending it’s a snake?’ Questions were repeated until subjects made an unambiguous verbal response or pointed to one of the pictures.

**Results**

The mean number of cards correct in each phase is presented separately for each group in Fig. 1. Although it would seem to provide a good index of the ability to use rules, the number correct is a poor measure. First, group results based on number correct are misleading because, in the first four phases, most subjects (81 of 96 cases, i.e. 84%) were either correct or incorrect in all five post-switch trials. Second, because non-handicapped children were nearly always correct, there was little or no variance in many of the cells created by crossing the independent variables. This situation compromises the validity of an ANOVA (Wilcox, 1987).

An alternative approach is to interpret each subject’s pattern of responses and to classify that subject according to whether he or she (a) failed to use either set of rules; (b) used only a single set; (c) employed both sets but was unable to switch rapidly between them; or (d) was able to switch rapidly between rule sets. Subjects were given a 0 if they failed to learn both judgements in Phases 1 and 2. They were given a 1 if they learned both, but were unable to use them during Phases 3 and 4. A 2 was assigned when subjects switched between Phases 3 and 4, but could not alternate during Phase 5. A score of 3 indicated success on all switches. In practice, failures in Phases 1 or 2 were scored as 0 because there was no strong evidence for the use of rules at all. Success in the first two phases followed by failure in Phases 3 or 4 was a 1 because it signaled an ability to learn both rules but not to switch between them. Correct performance in the first four phases, showing an ability to switch between rule sets, defined a 2. Success in Phase 5 plus the other phases was scored as a 3. For each of the first four phases, success was defined as responding correctly in four out of five trials, or in the last three. Success in Phase 5 was defined as responding correctly in all four trials.

Table 1 summarizes the number of subjects in each group who were assigned to each rule category. Success on the card sort was related to diagnostic group
Table 1
Classification of Subjects by Response Category on the Card Sort

<table>
<thead>
<tr>
<th>Diagnostic group</th>
<th>Response category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down’s Syndrome</td>
<td>4 5 2 1</td>
</tr>
<tr>
<td>Non-handicapped</td>
<td>0 0 4 8</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, most subjects with DS (nine out of 12) scored 0s or 1s, indicating either that they failed to learn both sets of rules, or that they learned both sets but were unable to use them during Phases 3 and 4. Seven of these subjects used the same set of rules in both Phases 3 and 4, despite the fact that they were told the relevant rules on every trial. Five of these subjects were able to learn both sets of rules in Phases 1 and 2. In contrast, all of the non-handicapped children were able to employ both rule sets in Phases 3 and 4. Eight out of 12 were successful in all five phases, including the last, which required different rules on each trial.

Theory of Mind Questions

All ToM questions, including pretend–reality, had a similar design. For each type of question, there were two items. Subjects were given a 1 for each item they responded to correctly, so scores ranged from 0 to 2 for each type of question. For pretend–reality and appearance–reality, two questions actually were asked for each item and subjects had to answer both correctly to score a 1. For example, for appearance–reality, subjects were required to say what the object appeared to be and what it really was to receive credit for that item. Only a single question was asked for each item for false belief and representational change.

The mean number of items correct for each type of question is depicted in Fig. 2. Given the categorical nature of the data, and the fact that the distribution of number of correct items was marked by substantial heterogeneity of variance \( F_{max} (4, 12) = 11.64, p < .01 \), which would compromise the validity of an ANOVA, data from the four ToM tasks were combined and differences between groups were analyzed using both parametric and nonparametric tests. The total score across all four ToM tasks ranged from 0 to 8. A one-way between-subjects ANOVA on total ToM scores revealed a main effect of diagnosis, \( F(1, 22) = 25.87, p < .0001 \). Subjects with DS \( (M = 3.25, SD = 1.42) \) performed worse than non-handicapped subjects \( (M = 6.42, SD = 1.62) \). For the purposes of nonparametric analyses, subjects with scores of 5 or more were classified as passing the ToM questions, whereas subjects who had lower scores were classified as failing. Five was chosen as criterion because it reflected a majority of questions. (More strict criteria produced the same pattern of results.) Among the subjects with DS, nine were classified as failing and three as passing. In contrast, 10 of the non-handicapped subjects passed the ToM questions and two failed them. A chi-squared analysis confirmed that the distribution differed across groups, \( \chi^2 (1, N = 24) = 8.22, p < .005 \).

All subjects who failed pretend–reality items answered one of the two test questions correctly. In the majority (78%) of the cases, this occurred because subjects answered the reality question correctly. Similar results were found for the appearance–reality task; all of the failures reflected errors in one of the two questions and in the majority (83%) of these cases, subjects answered the reality question correctly.

Relation Between Tasks

The simple correlations for all subjects among the total score on all four ToM tasks (out of 8), response category on the card sort and MA revealed that performance on the two types of task was highly correlated \( (r = .59, p < .005) \), although neither task was related to MA \( (r = .02 \) for each of the tasks). The lack of relation to MA is probably due to the fact that subjects’ MAs were within a restricted range, as well as to the fact that non-handicapped children tended to perform well on...
Rule Use and Down’s Syndrome

Discussion

Adults with DS performed worse than MA-matched non-handicapped children on several ToM tasks and a card sort that required them to switch between two incompatible sets of rules. Nine out of 12 subjects with DS were classified as failing the ToM tasks, whereas the majority of non-handicapped children passed them. Likewise, nine out of 12 subjects with DS were unable to use both sets of rules independently in Phases 3 and 4 of the card sort, whereas all of the non-handicapped children were able to do so. Further, performance on the two tasks was positively correlated when MA was partialled out.

The present findings suggest that, compared to MA-matched non-handicapped children, older individuals with DS with MAs ranging from 3.9 to 6.3 years have pronounced difficulty with standard ToM tasks. The people with DS correctly answered a mean of 3.25 out of eight questions and most of this success was on pretend-play questions. The majority of non-handicapped 3-year-olds (Flavell et al., 1987; but see Amsel & Remy, 1992). Together with other recent results showing that people with mental retardation perform worse than MA-matched non-handicapped subjects on ToM (Benson et al., 1993; Yirmiya & Shulman, 1994), the present findings challenge the Uniqueness Claim regarding ToM — that is, the claim that deficits in ToM are unique to people with autism and can account for their pervasive social and cognitive impairments. Deficits in ToM cannot invariably produce autistic behavior, because people with DS present a distinct behavioral and social profile.

The subjects with DS also performed relatively poorly on the card sort. Four out of 12 subjects with DS were unable even to learn to use both sets of rules in Phases 1 and 2, which suggests that rather than using rules at all, they may simply have approached the task with a bias toward sorting according to one dimension. Difficulty using even a single set of rules has been observed among non-handicapped 2.5-year-olds (Zelazo, Reznick & Pinon, 1995; Zelazo & Reznick, 1991), indicating that rule use follows a developmental sequence. Five subjects with DS were able to learn both rules in Phases 1 and 2, but then perseverated with one set of rules in Phases 3 and 4 — a pattern characteristic of 3-year-olds (Frye et al., in press). The finding that people with DS had difficulties switching with the card sort supports the suggestion that reasoning about mental states requires a domain-general system of inferences in which incompatible sets of rules are embedded under a more general rule governing which set of rules to use (Frye et al., in press).

The positive relation between ToM and nonsocial rule use extends previous findings and supports the cognitive complexity approach as against the Specificity Claim that ToM is a domain-specific function. A positive relation emerged despite differences in task format. Not only do people with DS exhibit difficulties in ToM and rule use relative to MA, but these difficulties co-vary across individuals. According to the Specificity Claim, ToM should be independent of card sorting regardless of the subjects involved. Difficulties in ToM may depend on more general difficulties in flexible reasoning, such as the ability to use a higher-order rule.

Previous findings supporting the Specificity Claim may have been due to a failure to compare ToM to tasks of equivalent complexity. More generally, however, evaluations of specificity and uniqueness potentially are compromised by difficulties in assessing general functioning in clinical samples. As ToM tasks are largely language-based, and as tests of language among people with autism typically provide MA scores that are low relative to non-verbal MA tests, the prevailing strategy in ToM research is to use a language-based test such as the PPVT-R. However, picture vocabulary tests assess vocabulary recognition rather than general language functioning or overall developmental level. These tests may be sensitive to different aspects of intellectual functioning in different groups — perhaps overestimating MA in some groups while underestimating it in others. In addition, although scores from these tests are correlated with other indices of MA in non-handicapped children, these relations are less likely to be found in people with organic mental retardation because these people show high variability across domains (Burack & Volkmar, 1992; Hodapp & Burack, 1990). Thus, vocabulary scores are unlikely to be an appropriate tool for matching among diverse clinical groups.

Although some impairments in ToM among people with DS have been reported (e.g. Baron-Cohen et al., 1986; Baron-Cohen, 1989) and there is evidence that people with mental retardation perform worse on ToM than MA-matched non-handicapped subjects (Benson et al., 1993; Yirmiya & Shulman, 1994), subjects in the present study were relatively impaired than one would expect based on previous studies (Baron-Cohen et al., 1985; Russell, Mauthner, Sharpe & Tidswell, 1991). Discrepancies among studies could be due to a variety of factors besides the specific samples involved. For example, subjects in the present study were relatively old. The sample of subjects with DS that participated in the studies of Baron-Cohen (1989) and Baron-Cohen et al. (1985, 1986) had CAs ranging from 6 years, 3 months to 17 years. Increased CA may exacerbate the difficulty of matching groups on MA. Perhaps the PPVT overestimates the ability of adults with DS because these people can learn words gradually, although their rate of development declines during adolescence and adulthood. Further work clearly needs to be done to assess reasoning in a wide range of people with DS.

None the less, the results may shed light on the impairments associated with DS, at least for low-functioning adults. For example, children with DS have been shown to have difficulty disengaging from an attractive stimulus (Kopp, Krakow & Johnson, 1983).
redirecting attention (Cicchetti & Ganiban, 1990) and responding to a previously inhibited stimulus (Sersen, Astrup, Floistad & Wortis, 1970). These difficulties may be related to difficulties using rules to guide behavior.

Finally, given that the mean MA of the subjects with DS was well above 4 years, and given that MA was based on a picture vocabulary test as in previous research, the results from the present study raise serious questions about the veracity of the Uniqueness and Specificity Claims. Investigations of ToM should include tasks that are matched for logical complexity but do not involve reasoning about mental states. In addition, ToM should be studied in subjects with a wide range of clinical syndromes.

Acknowledgements—This research was supported by New Researcher Awards to J. Burack from the Social Sciences and Humanities Research Council of Canada and the Fonds pour la Formation de Chercheurs et l’Aide a la Recherche du Quebec. A paper based on data from this study was submitted by E. Benedetto in partial fulfilment of the requirements for the M.A. degree in the Dept of Educational and Counseling Psychology at McGill University. The authors thank B. Randolph for help with the data collection, M. Scharer and J. Brennan for assistance in the preparation of this manuscript and S. Jacques for comments on an earlier draft.

References


Accepted manuscript received 4 October 1995
This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.