Child Development, March/April 2005, Volume 76, Number 2, Pages 502 – 517

Steps in Theory-of-Mind Development for Children With Deafness or Autism

Candida C. Peterson
University of Queensland

Henry M. Wellman and David Liu
University of Michigan

Prior research demonstrates that understanding theory of mind (ToM) is seriously and similarly delayed in late-signing deaf children and children with autism. Are these children simply delayed in timing relative to typical children, or do they demonstrate different patterns of development? The current research addressed this question by testing 145 children (ranging from 3 to 13 years) with deafness, autism, or typical development using a ToM scale. Results indicate that all groups followed the same sequence of steps, up to a point, but that children with autism showed an importantly different sequence of understandings (in the later steps of the progression) relative to all other groups.

Theory of mind (ToM)—the awareness of how mental states such as memories, beliefs, desires, and intentions govern the behavior of self and others—is ''one of the quintessential abilities that makes us human'' (Baron-Cohen, 2000, p. 3). As a cornerstone of social intelligence and satisfying social interaction, ToM develops rapidly during the preschool period. At age 3, most children fail standard ToM tests, prototypically tests of an understanding of false belief. By age 4 or 5, performance by typically developing children is so adept as to indicate that ''understanding of belief and, relatedly, understanding of mind, exhibit genuine conceptual change in the preschool period'' (Wellman, Cross, & Watson, 2001, p. 655). However, children with autism and deafness continue to fail ToM tasks through middle childhood and adolescence.

In an early demonstration of these delays, Baron-Cohen, Leslie, and Frith (1985) found that 80% of a sample of high-functioning children and adolescents with autism failed a standard false belief test that was passed by almost all typically developing preschoolers as well as by a control group with Down syndrome. This finding has been influential and widely replicated. Based on a review of 28 studies that included more than 300 autistic participants ages 4 to 30 years, Happé (1995) concluded that autistic participants consistently performed more poorly on false belief tasks than did children matched for chronological or verbal mental age. Indeed, Happé estimated that a verbal mental age of at least 11 years was needed before a child with autism had an 80% chance of passing a standard false belief test of ToM, in contrast to just 5 years for typical preschoolers. Similar findings were reported by Yirmiya, Erel, Shaked, and Solomonica-Levi (1998) in a meta-analysis of 22 studies that included typically developing control groups.

There have also been reports of delayed ToM among the 90% of deaf children who have hearing parents (Marschark, 1993). In an early study of ToM in deaf children, Peterson and Siegal (1995) administered Baron-Cohen et al.'s (1985) standard false belief test to a group of severely and profoundly deaf children ages 5 to 13 years who, although able to communicate easily in sign at the time of testing, had mastered sign language belatedly at school. Delays emerged that were on par with those seen in the original autistic sample. The finding of seriously delayed ToM development in signing deaf children from hearing families (late signers) has been widely replicated (see Peterson & Siegal, 2000, for a review of 11 studies; see also Jackson, 2001; Lundy, 2002; Peterson, 2002; Woolfe, Want, & Siegal, 2002). Deaf children of normal intelligence and social responsiveness from many countries (exposed to different sign languages and educational philosphies) are consistently behind younger hearing children, and on par with autistic children of similar age, in developing a ToM.

Some deaf children from hearing families do not acquire sign language but are instead trained in a purely oral modality to perceive and express speech

This work was partially supported by National Institutes of Health Grant HD-22149 to Henry Wellman. David Liu was supported by a National Science Foundation graduate fellowship. The authors wish to thank Margaret Gore, Shana Grozdanovic, Pam Jokic, Anna McAlister, Sara Parker, James Peterson, and Kim Peterson for assistance with data collection, and all participating teachers, interpreters, parents, and children for their excellent involvement, interest, and cooperation.

Correspondence concerning this article should be addressed to Candida C. Peterson, School of Psychology, University of Queensland, Queensland, Australia 4072. Electronic mail may be sent to candi@psy.uq.edu.au.

 0009-3920/2005/7602-0013

with the assistance of lipreading, amplifying hearing aids, or cochlear implants. Results of eight false belief studies, sampling a total of 223 oral deaf children with hearing aids or cochlear implants, have consistently revealed ToM delays similar to those observed among high-functioning autistic children and late signers (Peterson, 2004). Thus, ToM problems in the context of deafness cannot be ascribed simply to use of sign language but are general among both oral and signing deaf children whose families are hearing and whose hearing losses are severe to profound.

Deaf children with signing deaf parents or siblings can be described as native signers because of having grown up in families with fluently signing conversational partners. During early and middle childhood, these children have been found to do much better on standard ToM tests than their late-signing or oral deaf age peers (Courtin & Melot, 1998; Peterson & Siegal, 1999; Remmel, Bettger, & Weinberg, 1998). Furthermore, native signers' ToM superiority over late signers' ToM persists even after differences in executive functioning, nonverbal mental age, and language ability have been taken into account (Woolfe et al., 2002). Some results even suggest that native signers may master false belief concepts at a slightly younger age than hearing children (Courtin & Melot, 1998). Thus, delayed ToM development in late signers and oral deaf children is evidently not a function of deafness per se but rather of having grown up deaf in a linguistically deprived environment.

The vast majority of these studies of ToM development by children with typical development, autism, or deafness have employed false belief tasks, the so-called litmus tests for ToM. In these tasks, the child is asked to predict the beliefs or actions of human or puppet protagonists who are not privy to crucial pieces of privileged information, such as the changed location of hidden objects (Baron-Cohen et al., 1985) or the contents of misleadingly marked containers (Gopnik & Slaughter, 1991). Such tasks offer a critical test of children's understanding of mental representation by contrasting beliefs against reality and have netted a wealth of valuable information (Wellman et al., 2001). A focus on a single task or achievement, however, is limited (Astington, 2001, Bloom & German, 2000). In fact, most researchers now believe that, at least among typically developing children, understanding false belief is just one of many aspects of ToM (e.g., Flavell & Miller, 1998) that may emerge within a consistent developmental progression where, for example, understanding of desire precedes understanding of true belief and false belief (see Wellman, 2002).

The data are much less clear for deafness and autism. As one example, consider understanding of desire. Some investigators have concluded that understanding of desire is unimpaired in children with autism or deafness, and some have claimed it is delayed, just as is understanding of belief. For example, Rieffe and Meerum-Terwogt (2000) tested a group of late-signing deaf children ages 6 to 11 years in the Netherlands on a story task requiring explanation of characters' emotional reactions, and Rieffe, Meerum-Terwogt, and Stockmann (2000) did the same with a group of high-functioning autistic children of similar age. Results of both studies showed that deaf and autistic children used desire just as often as did age-matched typical children to explain story protagonists' emotional reactions, even though there were subtle differences among the desire explanations given by each group. Tager-Flusberg (1993) likewise found normal levels of spontaneous reference to desire in natural mother–child conversations involving children with autism, in contrast to depressed levels of spontaneous references to belief.

On the other hand, results from a face-reading task devised by Baron-Cohen, Campbell, Karmiloff-Smith, Grant, and Walker (1995) to study children and adolescents with autism, which was also used by Scott, Russell, Gray, Hosie, and Hunter (1999) to test late-signing deaf children, suggested impaired desire understanding in both groups. The task required attribution of feelings of desire ("What does Charlie want?") or desire-based behavior ("What will Charlie take?") to cartoon faces with neutral expressions whose eyes were directed at different things. The children with deafness and autism were each highly accurate at identifying the object of the eye's focus (when asked, "What is Charlie looking at?") but both groups were significantly outperformed by typically developing 4-year-olds on each of the desire questions.

However, methodological problems may have limited the validity of this task as a measure of desire understanding for children with deafness or autism, who may be more reliant than typical developers on pictorial information (Peterson, 2002). Indeed, Peterson (2003) altered the pictorial stimuli used and found that, under these conditions, signing deaf children of similar age and background to those tested by Scott et al. (1999) proved highly adept at reading desires from facial cues, with 88% earning perfect scores. In addition, the deaf children tested by Peterson did not perform any worse on the desire test than did a 4-year-old hearing control group, even though only 27% of the older deaf children passed a battery of standard false belief tests.

The data, despite mixed evidence (see also Baron-Cohen, 1991; Steeds, Rowe, & Dowker, 1997), are intriguing. Yet, to our minds, the most important questions are not so much what various groups know about desire, or about false belief, but rather about overall sequences of developing understanding. For example, does acquisition of an understanding of desire (delayed or not) regularly precede understanding of belief? In this regard, reconsider the now-classic finding that high-functioning autistic individuals and late-signing deaf children routinely fail standard false belief tasks. Although failure rates are high in both groups, some older individuals do pass (e.g., across the 18 autistic samples reviewed by Happé, 1995, 31% of the high-functioning individuals passed, and in Russell et al.'s, 1998, study, 17% of deaf late signers passed false belief at age 6 and 60% passed at age 15). These data raise intriguing questions. Are individuals with autism (or late-signing deaf individuals) distinctively impaired in ToM development, or only significantly delayed? More precisely, to the extent that older deaf or autistic individuals do achieve social cognitive understandings (e.g., understanding of desire, understanding of false belief), do they demonstrate delay in a consistent developmental trajectory or do they demonstrate some other developmental paths or patterns?

If developmental sequences were found to be similar or different across groups (e.g., between typically developing children and those with autism, or between autistic children and late signers), this would be important and informative. Data addressing such issues should assist in pinpointing the processes responsible for ToM impairments and delays while having implications for theoretical accounts of ToM development generally. Consider the theoretical divergence among nativist; modular; and experiential, socio-cultural accounts of ToM. Nativist positions argue that autistic children's difficulties have "a specific innate basis" (Scholl & Leslie, 2001, p. 697) such that "a specialized cognitive mechanism which subserves the development of folk psychological notions is dissociably damaged in autism" (Leslie & Thaiss, 1992, p. 229). An observation of contrasts in the sequence of ToM acquisition between late-signing deaf children and those with autism might, from a nativist perspective, be drawn on to elucidate the nature and functioning of the postulated neurological ToM module thought to be damaged for the latter, but not the former, group. Experiential, sociocultural accounts collectively view social experiences and conversational interactions through family talk and play as building blocks for ToM (e.g., Dunn, 1994). From this perspective, equivalent sequences in deaf and autistic groups could indicate that both groups encounter similar opportunities or restrictions (e.g., comparably curtailed family conversational interactions) whereas different sequences might reflect different kinds of social and conversational involvements. Data on sequences of acquisition are unlikely to adjudicate conclusively between such complex positions. From any perspective, however, information about sequences, and especially sequences amid delay, should illuminate our understanding of how ToM is achieved or constructed.

The current research empirically addressed these questions of sequence amid delay. A promising methodology for doing so has recently emerged from Wellman and Liu's (2004) investigation of different aspects of mental state understanding in typically developing preschoolers. In that study, tasks assessing the understanding of (a) diverse desires, (b) diverse beliefs, (c) perceptual access to knowledge, (d) false belief, and (e) hidden emotion were all similar in procedural methodology, linguistic structure, task demands, and materials. Yet results revealed a clear order of difficulty, indicating a consistent developmental progression among these five tasks. Typically developing children who passed a later item generally passed all earlier items as well, and the significance of the sequence was confirmed by Guttman scale and Rasch model analyses. The major question for the present study was whether the same order of difficulty also applies to children with deafness and with autism.

Method

Participants

The total sample of 145 Australian children consisted of 88 boys and 57 girls ranging in age from 3 years 7 months to 13 years 7 months. They comprised four groups. Table 1 displays some of the characteristics of each group. Group 1, native signers, consisted of 11 children who had been either severely deaf (losses from 71 to 90 dB) or profoundly deaf (losses more than 91 dB) since birth and who, since that time, had been reared in a household with deaf family member(s) who had native-speaker proficiency in sign language. Ages in Group 1 ranged from 6 to almost 13 years. Group 2, late signers, consisted of 36 children who had been severely or profoundly deaf since birth and who came from hearing families with no deaf members. Their ages ranged from $5\frac{1}{2}$ to 13 years. The deaf native and late signers were one another's classmates in one of

Table 1
Characteristics of Children in Each Diagnostic Group

	Deaf native signers	Deaf late signers	Autistic children	Typical preschoolers
No. of children	11	36	36	62
Mean age	10.67	10.01	9.32	4.50
(*SD*)	(1.83)	(2.46)	(1.88)	(0.59)
Age range (years; months)	6;4 to 12;6	5;5 to 13;2	6;3 to 14;2	3;10 to 5;9
Ratio of boys:girls	6:5	16:20	33:3	33:29
Mean language proficiency	3.73[a]	2.96[a]	94.31[b]	
(*SD*)	(0.65)	(0.70)	(27.21)	
Range of language proficiency scores	3 to 5[a]	2 to 4[a]	48 months to 156 months[b]	
Ratio of profoundly deaf: severely deaf	9:2	30:6		

[a]Denotes teachers' sign language ratings (0–5).
[b]Denotes verbal mental age (months) from Peabody Picture Vocabulary Test.

several government-funded Total Communication units for hearing-impaired children attached to primary schools located throughout the state of Queensland, Australia. In these Total Communication units, the main conversational medium, both in class and on the playground, was Signed English, supplemented by lipreading, fingerspelling, and Auslan (Australian Sign Language). All the deaf children were proficient in Signed English at the time of testing, and many, especially the native signers, also had proficiency in Auslan. Some in both deaf groups had some expressive speech, but not to the level of being able to dispense with the assistance of a signing interpreter when taking our tests, and all members of Groups 1 and 2 used sign, rather than speech, as their greatly preferred medium of communication according to both teachers' reports and direct observation during the testing session.

To avoid confounding deafness with other potentially complicating variables, no deaf child with a known or suspected severe disability apart from hearing loss (e.g., autism, mental retardation, visual impairment, cerebral palsy) was tested. All children in Groups 1 and 2 had been pupils in a Total Communication primary school for more than 12 months, and all were able to communicate effectively in sign according to their teachers' reports. This was confirmed by their good comprehension and control-question performance on our tasks (see Results). We screened the children for language proficiency by asking their highly experienced, fluently signing classroom teachers to rate each child's command of sign language vocabulary on a 6-point scale in which 0 = minimal or nonexistent; 1 = a few basic signs and gestures only; 2 = somewhat smaller than the average signing child this age, but adequate for everyday communication; 3 = about average for children this age who communicate freely and fluently in sign; 4 = somewhat larger than for the average signing child this age and enabling highly skilled communication; and 5 = exceptionally large signing vocabulary used with great fluency and versatility (top 10% of children of this age). All the children in Groups 1 and 2 we tested had earned teacher language ratings of at least 2 on this measure, indicating all had signing vocabularies that were perceived as adequate for everyday communication. Means, ranges, and standard deviations for scores on this measure are shown in Table 1.

Group 3 consisted of 36 children with autism who ranged in age from 6 to 14 years. Their mean age (see Table 1) did not differ significantly from the mean ages of the deaf children in Groups 1 and 2, $F(2, 141) = 1.90$, $p > .15$. All the children in Group 3 had been diagnosed as meeting *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM–IV*], American Psychiatric Association, 1994) criteria for autism by a team of experienced psychiatrists and clinical psychologists, and all were attending government-funded schools, or units, specialized for autism. To ensure adequate language skills to take our tests, we chose children who were classed as high functioning, and we administered the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1981) as a screening measure. No child with a verbal mental age (VMA) below 4.00 years was included. (Two children initially suggested by teachers as high functioning scored below this minimum level of language proficiency on the PPVT; therefore, they were not tested further or included in our sample descriptions.) The mean VMA for Group 3 was 7 years 10 months (range = 4;0 to 13;0) and, as is typical of children with autism, most VMAs were below the child's chronological age, the mean gap being 18 months behind. None of the autistic children we tested had codiagnoses of additional disabilities (e.g., mental retardation, deafness) and they all came from families whose sole, or first, language was English. For the children with autism, as for the deaf children in Groups 1 and 2, social skills lessons were a regular part of the school curriculum. However, to the best of our knowledge, these lessons did not include topics of ToM, false beliefs, or feigned emotion.

The remaining 62 children constituted Group 4, typically developing preschoolers ages $3\frac{1}{2}$ to $5\frac{1}{2}$ years who attended government-funded preschools, either adjacent to, or in neighborhoods similar to, the schools and units the children in Groups 1 to 3 attended. This was the only (admittedly approximate) indication available to us of similar socioeconomic background across the sample. All of the preschoolers in Group 4 had English as their sole, or primary, language and none had any known or suspected serious disabilities according to their teachers' reports.

Procedure

Each child was tested individually. For the signing deaf children, two adults were present: an experienced male experimenter and one of three professionally trained interpreters of sign language who were highly familiar with the style of total communication used in each deaf child's classroom as well as with each child's language preferences (e.g., for signed English vs. Auslan). Each interpreter was well known to the tested child and was employed in some capacity in the child's school. The interpreter, who was seated beside the experimenter and directly opposite and in full view of the participant, provided an accompanying translation of the experimenter's speech in the child's preferred mode of sign language, using a style of interpretation that was a familiar part of these children's everyday school routines. The interpreters paused while critical bits of stories were acted out (such as a doll's entry onto the scene), and both adults monitored that the child's gaze was directed at the props or the interpreter, as appropriate, before continuing each part of the procedure. Both adults independently recorded the child's pointing responses, and subsequent matching of their records revealed complete agreement. In addition, the interpreter supplied an ongoing oral translation of all the child's signed communication, which was recorded by the experimenter on the data sheets. We developed task and question wording (see the Appendix) in close consultation with native speakers of Auslan and Signed English to ensure the feasibility of uniform translation into these languages. Our interpreters were well practiced both in signed translation and in the importance of adhering to the script exactly, and this was monitored by the experimenter.

The children with autism, and the typical preschoolers, were tested individually in speech by one of four experimenters (3 female and 1 male). The procedure, tasks, wording of the narratives and questions, and all other experimental features were identical for children in each of the four groups. Details appear in the Appendix. At the time of the testing, none of the interpreters or experimenters was aware of our main hypotheses, of the sequence of performance reported in Wellman and Liu (2004), or of our interest in the deaf or hearing status of the deaf pupils' families.

Tasks and Scoring

The five tasks we used—diverse desires, diverse beliefs, knowledge access, false belief, and hidden emotion—were closely modeled on the five tasks Wellman and Liu (2004) had found to produce a coherent Guttman scale for typical preschoolers in the United States. A few minor changes were made to wording or stimuli for some of the tasks, either to make them familiar to Australian children (e.g., changing cookie to biscuit) or to simplify the task language and ensure that a directly equivalent translation could be made into Signed English and Auslan. Our modified versions of the tasks were then given in an identical manner to all the children in all groups in our sample (see the Appendix).

All tasks included a focal test question as well as at least one other preliminary question or a control question or both. Like Wellman and Liu (2004), we ensured that children responded to the preliminary questions sensibly and attentively. In addition, we required that children pass any associated control questions, as well as test questions, to count as passing a task. This ensured that children comprehended and remembered all the relevant vocabulary, syntax, and story information on which a meaningful, rather than random, response to a test question could be based. Most children in all groups in our study passed most of the control questions (see Table 2).

Our only notable change from Wellman and Liu's (2004) procedures involved the control question for the hidden emotion task. We dropped their control question: ''In the story, what would the other children do if they knew how Matt felt?'' because its length, conditional syntax, and embedded phrase structure posed sign-translation and comprehension

Table 2
Percentage of Children Passing Each Item by Responding Correctly to All Test and Control Questions

	Deaf native signers ($n = 11$)	Deaf late signers ($n = 36$)	Autistic children ($n = 36$)	Typical preschoolers ($n = 62$)
Diverse desires	100	92	86	95
Diverse beliefs	91	92	86	85
Knowledge access	82	53	75	82
(Control only)	(100)[a]	(94)	(89)	(97)
False belief	82	33	47	32
(Control only)	(100)	(86)	(92)	(90)
Hidden emotion	54	28	64	19
(Control only)	(91)	(70)	(83)	(79)

[a]Numbers in parentheses denote percentages of children responding correctly to control questions, irrespective of test question accuracy.

problems. As an alternative control question, after children had responded to the final test question by either naming (or pointing at a picture of) the protagonist's apparent emotion we asked, ''Why did he try to look [child's test question response]?'' (see the Appendix).

To be scored as passing this control question, children could give any reason that demonstrated awareness of relevant story information. Control reasons did not have to justify a correct test question response to demonstrate story comprehension. For example, children who had chosen the wrong (sad) emotion in response to the test question often gave control answers such as, ''He was sad that the children all teased him'' or ''His friends were being nasty to him'' and these were scored as correct. Also scored as correct were justifications for accurate (e.g., happy) test question choices such as, ''So the kids wouldn't tease him,'' ''To try to stop them laughing,'' and ''To try to trick them.'' Responses that gave no indication of story comprehension were scored as incorrect. These included, ''Don't know,'' empty answers (e.g., ''Because that's why''), and reasons that were irrelevant to the story (e.g., ''The boy likes to play football''). The first author coded all these control question responses, both for correctness and for type of correct or incorrect reason given. An independent rater, who was unaware of the respondents' diagnostic groupings, then coded 50 transcripts and agreed with the primary coder on 94% of them. As was the case for control questions on other tasks, most children in all the groups passed the comprehension control question on the hidden emotion task (79% correct overall, see Table 2). For the main statistical analyses (reported next and in the tables) children were only counted as passing a task when they passed both the test question and any associated control questions. (We also repeated all analyses after ignoring responses to control questions. Both approaches resulted in equivalent results for the current study, just as in Wellman and Liu's, 2004, study.)

Males were disproportionately represented in Group 3, as they are in the population of children with autism as a whole (Frith, 1989). Therefore, before conducting the main analyses, we tested for gender differences on all our main variables in Groups 2 and 4, each of which had large enough numbers of children of both sexes to make such comparisons valid. No significant gender differences in pass rates emerged in any group on any of the individual tasks, all χ^2s < 1, all ps $> .35$. Also, there was no significant difference between the frequencies of boys and girls whose complete response pattern across the five tasks either exactly matched, versus departed from, the Guttman scale sequence reported by Wellman and Liu (2004), Fisher's exact $p = .11$ for Group 2, and $\chi^2(1) = 1.68$, $N = 62$, $p > .25$, for Group 4. Thus, the gender imbalance in Group 3 was unlikely to have significantly influenced our results or interpretations.

Results

To provide an initial comparison with other research, which has predominantly assessed false belief, first consider children's performance on our single false belief task. In the deaf native signers group, 82% of children passed this task, yet only 33% of late signers and 47% of children with autism did so (see Table 2), $\chi^2(2) = 8.27$, $N = 83$, $p < .02$. There was no statistically significant difference among Groups 2 and 3, $\chi^2(1) = 1.44$, $N = 72$, $p > .20$. But there was a significant difference in false belief understanding between the deaf native signers and their late-signing deaf classmates, $\chi^2(1) = 8.01$, $N = 47$, $p < .01$, similar to significant differences observed by Courtin and Melot (1998), Peterson and Siegal (1999), Remmel et al. (1998) and Woolfe et al. (2002). In short, in line with previous studies (see Peterson & Siegal, 2000, for a review), our native signers outperformed each of the other nontypical groups.

On average, the typical preschoolers (Group 4) also failed false belief (32% correct), but that group was considerably younger than any other group. In this group, children who were age 4.5 years and older significantly outperformed younger preschoolers, $\chi^2(1) = 3.94$, $N = 62$, $p < .05$, displaying an age trend for typical children that is also in line with much previous research (Wellman et al., 2001).

Moving beyond false belief, percentages correct for each task for each group are shown in Table 2. We assigned children a total ToM score (from 0 to 5) reflecting how many of the five tests in the battery they had passed (by correctly answering test and control questions). There was a statistically significant difference in this total score among the groups, $F(3, 141) = 3.37$, $p < .05$. A post hoc Scheffé test showed that the native signers ($M = 4.09$) scored significantly higher than their late-signing deaf peers ($M = 2.97$). The autistic children and the preschoolers scored intermediately. Their means, 3.58 and 3.15, respectively, did not differ significantly from one another's, and they did not differ significantly from those of either of the deaf groups. Total ToM scores increased significantly with chronological age for the children in Group 1, $r(9) = .70$, $p < .02$; Group 2, $r(34) = .55$, $p < .01$; and Group 3, $r(34) = .42$, $p < .01$. The correlation with age did not reach statistical significance for the typical preschoolers, $r(60) = .17$, $p = .20$. However, for these children there was a statistically significant difference in the mean ToM scores (3.54 and 2.86) of those who were older versus younger than 4.5 years, $t(60) = 2.86$, $p < .01$, similar to the finding for false belief alone.

The key question for our research, however, concerns not age trends or group differences but rather sequences of understanding. Note that for this question equivalences, or lack thereof, in chronological age across different groups are much less crucial and, thus, less problematic. As long as there is a range of ages in each group, children may demonstrate the exact same patterns, and indeed a range of ages and abilities can be helpful, rather than problematic, for revealing extended sequences.

Table 2 is organized from easiest to hardest according to the sequence reported by Wellman and Liu (2004). The pattern for three of the groups (namely, the hearing preschoolers and both deaf groups) conforms to that same sequence, discounting tied high scores on some adjacent items (possibly denoting ceiling effects). The remaining group, children with autism, follows the same sequence except that for them false belief was the most difficult task, more difficult even than hidden emotion. Their sequence thus reverses the sequence for all other groups for these last items.

To confirm these impressions we conducted Guttman and Rasch scaling analyses on Groups 2 through 4 in the same manner as in Wellman and Liu (2004). The smaller size of Group 1 makes such inferential analyses inappropriate; therefore, we describe their performance only briefly and descriptively.

Scalogram Analyses

Guttman scalogram analyses (Green, 1956) assess how well observed item difficulty orderings match the theoretical prediction of a perfectly ordered scale in which any participant who passes a harder item on the scale will have passed all the easier items, and no participant who passed a more difficult item will have failed any of the easier items. Table 3 shows the Guttman scalogram patterns that emerged in Wellman and Liu (2004), along with the numbers and percentages of children in each group whose correct and incorrect answers fit each of the scale-consistent patterns.

The data for the typical preschoolers in Group 4, with tasks ordered as in Wellman and Liu (2004), yielded an index of reproducibility of .95 (scores above .90 are considered significant; Green, 1956). More important, Green's (1956) index of consistency, a more conservative measure that takes into account patterns that could be expected by chance alone, was .58 (scores greater than .50 are considered significant; Green, 1956). In other words, for typical hearing preschoolers in our sample, the data clearly replicated those of Wellman and Liu. In their study 80% of typical preschoolers fit the pattern exactly (compared with 79% in the current study), and their indexes of reproducibility and consistency were .96 and .56, respectively. Thus, the sequence for typical children seems robust against any cross-national or cross-sample variations that may have existed between the groups of preschool pupils tested in the two studies. Furthermore, the nearly identical pattern of performance indicates that the tasks were validly comparable across studies in spite of the slight variations we made on Wellman and Liu's original tasks in terms of wording, control questions, linguistic complexity, and scoring, as described earlier.

The late signers in Group 2 produced patterns of task performance that closely fit the same sequence. For these deaf children, with the tasks ordered as in Wellman and Liu (2004), the data yielded an index of reproducibility of .98 and an index of consistency of .77, both of which are statistically significant. Thirty-two of the 36 late signers (89%) had response patterns that conformed to the Guttman sequence exactly—a very high level of consistency.

The native signers, although small in number, also closely fit this same scale pattern of performance across the five tasks. Nine of these 11 children (82%) had patterns of performance that conformed exactly to the expected scale sequence as reported by Wellman and Liu (2004).

However, children with autism (Group 3) displayed a clearly different pattern. When their responses were

Table 3
Number of Children in Each Group Whose Pass (+) and Fail (−) Responses Fit Guttman Scale Patterns

Standard pattern								
Diverse desires (DD)	Diverse beliefs (DB)	Knowledge access (KA)	False belief (FB)	Hidden emotion (HE)	Group 1: Native signers ($n = 11$)	Group 2: Late signers ($n = 36$)	Group 3: Autism ($n = 36$)	Group 4: Typical preschool ($n = 62$)
−	−	−	−	−	0	2	1	0
+	−	−	−	−	0	1	0	2
+	+	−	−	−	1	13	5	8
+	+	+	−	−	0	4	3	23
+	+	+	+	−	3	5	2	11
+	+	+	+	+	5	7	12	5
			Total	Guttman	9 (82%)	32 (89%)	23 (64%)	49 (79%)

End-reversed pattern					
DD	DB	KA	HE	FB	Group 3: Autism ($n = 36$)
−	−	−	−	−	1
+	−	−	−	−	0
+	+	−	−	−	5
+	+	+	−	−	3
+	+	+	+	−	7
+	+	+	+	+	12
			Total	End-reversed	28 (78%)

Other observed patterns								
DD	DB	KA	FB	HE	Group 1: Native signers ($n = 11$)	Group 2: Late signers ($n = 36$)	Group 3: Autism ($n = 36$)	Group 4: Typical preschool ($n = 42$)
−	+	−	−	−	0	1	0	1
+	−	+	−	−	0	0	0	3
+	−	+	+	+	0	0	0	2

Note. There were also 11 separate idiosyncratic patterns that are not shown here as none of them applied to more than a single child in the sample. In total, there are 32 possible patterns of +/− responses across the five items. The Guttman scale-consistent patterns thus represent 19% of the logically possible patterns and the remaining other observed patterns represent 81% of the possibilities.

ordered as in Wellman and Liu (2004) and as for the other groups (the ordering shown at the top of Table 3), only 64% fit the Guttman pattern and their index reproducibility was .90 and their index of consistency was .29, the latter being substantially below the .50 level required for significance. However, children with autism did not show random sequences but instead followed a different, yet consistent, sequence of development. If the last two tasks in the sequence, the false belief and the hidden emotion tasks, are reversed, then, as shown in the middle of Table 3, 78% of children with autism fit this new end-reversed sequence. Furthermore, for the children with autism, this alternative pattern of responses yielded an index of reproducibility of .95 and an index of consistency of .55, both of which are statistically significant.

Clearly, children's responses are highly scalable and consistent, but a key reversal emerged: False belief is harder than hidden emotion for the autistic children, but hidden emotion is harder than false belief for everyone else, including the late signers. Confirming this reversal, significantly more children with autism failed the false belief task while passing every other task than did children in all the other groups combined, $\chi^2(1) = 5.25$, $N = 145$, $p < .05$. Furthermore, consider the frequencies of children who passed either false belief or hidden emotion but not both. Three fourths of the children with autism passed hidden emotion even though they failed false belief, whereas percentages in the other groups (in numerical order) were only 20%, 37%, and 28%, $\chi^2(3) = 7.93$, $N = 43$, $p < .05$. More specifically, whereas 75% (9 of 12) of the autistic children who passed only one of the tasks passed hidden emotion, 71% (22 of 31) of the children in Groups 1, 2, and 3 showed the opposite pattern, succeeding on false belief but failing on hidden emotion, $\chi^2(1) = 7.53$, $N = 43$, $p < .01$.

The statistical reliability of the contrast in task sequence between the children with autism and those in the other groups was further confirmed by the finding that when we compare the frequencies of children whose responses perfectly fit the standard pattern (diverse desires, diverse beliefs, knowledge access, false belief, hidden emotion; per Wellman & Liu, 2004) versus the end-reversed pattern (diverse desires, diverse beliefs, knowledge access, hidden emotion, false belief), a significant difference emerged between the autistic children in Group 3 and all other groups combined, $\chi^2(1) = 6.05$, $N = 127$, $p < .02$. No other sequence apart from these two was supported by the data; this is clear from the scale-inconsistent patterns produced by children who do not fit either of these sequences. All the scale-inconsistent patterns that emerged in two or more of the 145 children are detailed in the bottom part of Table 3, where frequencies are also shown. Only 3 distinct scale-inconsistent patterns emerged in 2 or more children. In addition, only 11 other nonscale patterns occurred, each being so idiosyncratic that it appeared in just a single child.

When children failed an item they invariably answered sensibly to any preliminary questions (e.g., for the diverse desires and diverse beliefs tasks, all children picked an item or location that was appropriate for the first question, thereby showing attention to the materials, understanding of the questions, and so on). More important, even children who failed an item typically passed the relevant control questions for that item, as shown in Table 2. Of course, less advanced children in any group might fail both the target and control questions for some of the hardest items. The most relevant data thus concern the first task a child failed. That is, for Groups 1, 2, and 4, consider the order at the top of Table 3, the order that best fits those groups. Using that order, on the first task failed (whatever it was for each child), children in those groups were 91% correct on the paired control questions (100% correct for Group 1, 92% for Group 2, and 90% for Group 4). The same finding applies for the children with autism in Group 3 except that here the proper order of tasks is end-reversed, as fitting for this group. Using this order, 93% of the children with autism were correct on the paired control question for the first task they failed (whereas, of course, 0% were correct on the paired target question). Thus, for all groups, task failure (at whatever point in the scale it first arose for any given child) appeared to reflect conceptual problems rather than being an artifact of language difficulties or confusion about the procedure.

Rasch Analyses

Because Guttman scales are stringent—items are scale appropriate only for fitting the exact step functions for increasing difficulty—item response theory approaches to scale analysis have been developed, in part, to allow consideration of less strict scale progressions (Andrich, 1985; Bock, 1997; Embretson & Reise, 2000; Lord & Novick, 1968). The most straightforward item response theory model, the Rasch measurement model, is a one-parameter logistic model for dichotomous items that estimates item difficulty and person ability levels (Rasch, 1960; Wright & Masters, 1982; Wright & Stone, 1979). The Rasch item response theory measurement model is

Table 4
Rasch Analysis Results for Late Signers, Children with Autism and Typical Preschoolers

Late signers Group 2					Children with autism Group 3					Typical preschoolers Group 4				
Order	Measure	Error	Standardized infit	Standardized outfit	Order	Measure	Error	Standardized infit	Standardized outfit	Order	Measure	Error	Standardized infit	Standardized outfit
DD	0.00	1.25	−0.2		DD	0.00	0.60	0.1	0.0	DD	0.00	0.70	0.5	1.1
DB	0.00	1.25	−0.2		DB	0.00	0.60	−0.3	−0.4	DB	1.59	0.50	1.2	2.9
KA	8.28	1.27	−1.2	0.7	KA	1.15	0.50	−0.6	−0.6	KA	1.95	0.41	−2.4	−1.3
FB	13.58	0.69	−0.4		HE	2.08	0.49	0.2	0.2	FB	5.72	0.37	−1.9	−0.8
HE	14.54	0.72	−0.1		FB	3.58	0.61	1.0	0.4	HE	6.99	0.46	0.4	0.4
M			−0.4					0.1	−0.1				−0.4	0.4
SD			0.4					0.5	0.4				1.4	1.5

Note. DD = diverse desires; DB = diverse beliefs; KA = knowledge access; FB = false belief; HE = hidden emotion.

often regarded as a probabilistic model for Guttman-type scaling and has the advantage of using probability estimations, rather than all-or-none comparisons, to identify scale properties (Andrich, 1985; Wilson, 1989).

Data for the five items for Groups 2, 3, and 4 were analyzed with Rasch measurement models using the WINSTEPS/BIGSTEPS computer program (Linacre, 2003; Linacre & Wright, 1994; see Wellman & Liu, 2004, for more details on using the Rasch model for these types of analyses). For numerical simplicity, the item difficulty and person ability measures on the linear logit scale were rescaled so that diverse desires (arbitrarily considered as the anchor task of the five tasks) had an item difficulty measure score of 0.0 on the linear scale. Table 4 shows ordered solutions from least difficult (lowest measurement score) to most difficult (highest measurement score) for the three groups. The order of item difficulty is the same in the Rasch models as in the Guttman scales for each group. Thus, this method confirms the reversal of sequence for children with autism.

Goodness-of-fit statistics, using a Rasch model, evaluate the notion that a person with a given ability level will be likely to respond correctly to less difficult items and will be likely to respond incorrectly to more difficult items. Two types of fit statistics are estimated for each item and each person: (a) infit, which is more sensitive to unexpected probabilities of response near the item or person's measurement level, and (b) outfit, which is more sensitive to unexpected probabilities of response far from the item or person's measurement level (Linacre & Wright, 1994; Wright & Masters, 1982). Standardized infit and outfit statistics for individual items have an expected value of 0. Values greater than 2.0 indicate greater unpredictable variation than expected. We consider standardized fit statistics for individual items greater than 2.0 as indicating misfit (Wright & Masters, 1982). In addition to extracting a sequence of item difficulty, the Rasch measurement model, given its probabilistic assumptions, also calculates a metric for how far apart successive items are. This is accomplished by using variation around the best fit sequence. Thus, a limitation of the Rasch model is that if responses fit the extracted sequence too well (e.g., look almost identical to a deterministic Guttman scale), these calculations of item distances (but not the extraction of the sequence) become uninformative and incalculable because there is insufficient variance to model. When this happens, and thus there are almost perfect separations between certain items, the Rasch model is unable to estimate standardized outfit statistics.

For the typical preschoolers, as shown in Table 4, the standardized infit statistics for all five items and all but one of the outfit statistics fall well short of 2.0, and mean fit statistics are near the expected value of 0. For the children with autism, the standardized infit and outfit statistics for all five items fall well short of 2.0, and mean fit statistics are near the expected value of 0. For the late signers, the standardized infit statistic for all five items fall well short of 2.0, and the mean standardized infit statistic is near the expected value of 0. The Rasch model was unable to calculate the standardized outfit statistic for four of the five items for the late signers because they closely fit a deterministic Guttman scale (as indicated by this group's high indexes of reproducibility and consistency, as reported earlier).

Nonetheless, the Rasch analyses clearly confirm the findings of the Guttman scalogram analyses as to sequences of task difficulty for each group, confirming that the order of difficulty for the false belief and hidden emotion tasks reverses when the children with autism are considered. This is clear in Table 4 for the focal contrast between late signers and children with autism and for the comparison between typically developing children and those with autism. For typical preschoolers and for children with autism there was sufficient variance to estimate meaningfully the relative distances between items. For preschoolers, the order is false belief then hidden emotion, and the distance between these 2 tasks is sizable (2.3). For children with autism, the order is hidden emotion and then false belief, and the reversed distance is also sizable (1.5).

Discussion

Consistent with much previous research, most of the late-signing deaf children and high-functioning children with autism in our study were still failing a standard false belief test at 8 to 10 years old, whereas most deaf native signers and hearing-normal children older than 5 to 6 years passed such tasks. Bolstered by this replication of past research, our crucial and novel results concern the overall patterns and sequences of understanding across a variety of ToM understandings and tasks. The results of the Guttman and Rasch analyses for typically developing children confirmed the same progressive order of difficulty observed earlier by Wellman and Liu (2004). The results for late signers and children with autism were also significantly scalable. For late signers, the sequence that emerged exactly matched that for typical children (and for native signers), albeit delayed in ages of attainment. For children with autism, however, the sequence was different.

These sequences amid delay include two findings of equal importance. First, for all groups the developmental sequence seems identical in its early steps, albeit delayed in some cases. Second, for individuals with autism alone, the sequence is different in its later steps, specifically with regard to understanding false belief. To set the context for these findings, consider several alternative possible patterns of results. Hypothetically, for one or both of the delayed groups, it was possible that no scalable pattern would emerge. For late signers and for children with autism, exposure to social life, to linguistic and conversational content, and to remedial instructional experiences could have been so idiosyncratic that some children would be good at some concepts and some would be good at other concepts in ways that were unsystematic and thus unscalable across individuals. Yet, results for both delayed groups were highly regular and scalable. Alternatively, it was also possible that the life experiences or neurological processes in one or the other of these delayed groups could have been so atypical that their order of understanding would have been completely different from that of typically developing children. Yet, the order of acquisition for all groups matched the same sequence in its early steps. Finally, it was possible that the order of acquisition could have been identical across all groups. However, the sequences were actually different in one notable respect.

Against this background, the exact patterns we found are informative in multiple ways. That deaf children and hearing Australian preschoolers demonstrate the same sequence of understanding as did Wellman and Liu's (2004) original sample of normally developing American preschoolers confirms the original scale findings with an additional and different sample. These findings add considerably to earlier suggestions that, with variations in absolute ages of mastery, ToM understanding develops similarly across different parts of the world both in typical children (see Wellman et al., 2001, for a review) and in those who are deaf (see Peterson, 2004; Peterson & Siegal, 2000, for reviews). The late-signing deaf children's data are especially noteworthy; even though these children are often as slow as their peers with autism to master ToM, the steps they go through seem to match those of the hearing preschoolers and native signers who are on a faster timetable.

Our scale analyses indicate that late-signing deaf children and individuals with autism are delayed not just on standard false belief tasks but on mental state understanding more generally. For them, like typically developing children, an understanding of desire precedes corresponding understanding in the realm of belief. In addition, an initial understanding of knowledge and ignorance develops ahead of the understanding that someone can hold a belief that the child knows is decidedly false. Yet, at each of these steps (except for diverse desires, where our data are obscured by ceiling effects for these groups), both late-signing deaf children and children with autism come to these understandings at ages that are older, by several years, than their age of acquisition by typically developing children or native signers. Why might these children's developmental lag be widespread across a variety of mental state concepts and tasks? One explanation we favor is that many departures from the normal courses of experience with social interaction, language, and conversation (especially about nonobvious topics such as internal states) are at work in children with deafness or autism. These delay each sequential step in the child's understanding of minds in connected ways, as demonstrable in a delayed, yet consistent, progression of understanding.

These groups' consistent similarity in sequence at the lower end of our scale suggests that conceptual progression, or socioconversational experiences, are remarkably similar across human development when it comes to early, basic understanding of mental states. The progression (most evident in these data for typical preschoolers and native signers) on the first two scale steps (diverse desires to diverse beliefs) is consistent with the postulated progression of naive psychological thinking from an initial implicit theory that human behavior is desire driven to an implicit belief-desire psychology (Wellman, 1993, 2002; Wellman & Woolley, 1990). The fact that knowledge access (i.e., awareness that not seeing leads to ignorance) was easier for all groups rather than false belief is consistent with a theoretical distinction sometimes drawn between Level 1 perspective taking (blocked sensory access precludes seeing or knowing) versus Level 2 perspective taking (different people can see, or think about, the same thing differently; e.g., Flavell, 1999; Reed & Peterson, 1990; Taylor, 1988). That diverse belief was easier than false belief for all groups is similarly revealing. Sometimes children's understanding of false belief is equated with an understanding of thinking per se. However, the fact that all the samples of children in our study mastered diverse belief and knowledge access before false belief argues for a series of conceptual steps in a developing understanding of thinking, with false belief being distinctive and difficult. According to Leslie's (1994) nativist account, the same neurological module that computes false belief also computes mental state attributions of thinking, knowing, desiring, and feeling. Moreover, the same damaged neurological module that blocks false belief understanding is responsible for deficits in representational understanding of desires and knowing. Our data speak against such a proposal. They suggest instead that individuals with autism may have a distinctive, autism-specific difficulty with the sort of mental state understanding that is needed for false belief tasks, above and beyond other sorts of mental state understandings.

This conclusion results primarily from the reversal of the ordering of the last two scale items by the autistic group, relative to the order characterizing the deaf and hearing participants. We do not believe that these findings are due to peculiar tasks or task formats. Both tasks have been used in prior research and were framed simply in equivalently direct language using similar stimuli. Most important, the fact that order of difficulty reverses itself for these tasks across groups argues against the possibility that one task or the other was intrinsically more difficult because of content, task demands, or linguistic complexity.

Why might those with autism demonstrate such a reversed pattern? Probably they are processing the tasks differently from the other groups, but this general explanation admits several possibilities. One possibility concerns the propensity for those with autism to think pictorially and to conceptualize thoughts as pictures in the head. That is, those with autism may be relatively more adept at dealing with mental states that can readily assume a pictorial format (Harris & Leevers, 2000; Peterson, 2002; Wellman et al., 2002), and perhaps it is easier to construe hidden emotions, rather than false beliefs, in pictorial terms. Pictures can distort, omit, or become obsolete, but they do not negate reality in the same way as a false belief. In this way, the apparent emotion in our real-apparent emotion story could be imagined as a picture (say, of a painted, smiling mask) that could coexist with reality (e.g., by being held in front of a sadly drooping real face). Arguably, this sort of visual analogy could make the real-apparent emotion task easier than false belief for the autistic group, in line with the suggestion by Harris and Leevers (2000) that pictures ''may offer children with autism an external and understandable analogue of their primary mode of thinking, namely visual imagery'' (p. 198).

However, such a proposal then fails to account for the opposite pattern (the greater difficulty of hidden emotion than false belief) for the late-signing deaf group. Late-signing deaf children, like those with autism, are also apt to think visually. They are generally better at remembering visual scenes, or pictures, than words or text (Marschark, 1993), have well-developed visuospatial skills (Emmorey, 1998), and perform even more skillfully than age-matched hearing controls on tasks requiring visual identification of subtle similarities and differences in facial photographs (Bettger, Emmorey, McCullough, & Bellugi, 1997). Furthermore, late signers have been found to score similarly to their peers with autism on false photographic and false drawing tests, and both groups typically score higher on these than on standard false belief tests (deVilliers & deVilliers, 2000; Peterson, 2002; Peterson & Siegal, 1998). Thus, a different explanation is needed.

Our hypothesis (for future research) is that for normally developing children both false belief and apparent emotion understandings involve false representations. Both of these tasks portrayed protagonists whose internal states differed from tangible, observable reality, and both involved deceptive appearances (a box with false contents, a face with a false expression). Therefore, the hidden emotion task was similar to the standard false belief task, for typical children, in requiring an understanding of the impact of deceptive appearances on others' mental states, but arguably harder in that it required an understanding of the protagonist's deliberate intentions and active attempts to deceive someone else (e.g., see Harris, Donnelly, Guz, & Pitt-Watson, 1986).

In contrast to this sort of sequenced understanding, our hypothesis is that high-functioning individuals with autism have ''hacked out'' (Happé & Siddons, 1994) rote solutions to certain emotion situations that allowed them to bypass false belief understanding in this instance; they have devised some sort of ''work-around'' for dealing with such situations, without thinking about them in terms of false belief. Emotion understanding certainly has real-world relevance for high-functioning children with autism, who may devise some alternative strategies to take account of their peers' real and apparent emotions to interact partially successfully with them, even though lacking a clear appreciation of their mental states.

This hypothesis might seem unlikely in that it attributes to individuals with autism a relatively enhanced awareness of (some) emotions. However, in line with this proposal, there is emerging evidence to suggest that high-functioning individuals with autism understand complex emotions more readily than false belief and that they may outstrip control groups in the sophistication of certain kinds of emotional understanding. Hillier and Allinson (2002) found that high-functioning autistic adolescents understood others' embarrassment remarkably well on computerized story tasks assessing awareness of a range of different variables bound up with embarrassment (such as the causes for it, the influences on it of presence and type of audience, etc.). Embarrassment is complex, both cognitively and socially (Baron-Cohen, 1995); a mature appreciation entails recognition of a real or imagined other's adverse evaluation of one's own situation or behavior. Yet Hillier and Allinson found that adolescents with autism displayed a sensitivity to subtle cues and performed at least as well as control groups with typical development or learning disabilities, matched on mental and chronological age. They even gave responses and justifications that were more sophisticated than those from control groups in some of the conditions (e.g., for scenarios manipulating levels of authority of the audience). Such an unexpectedly adept understanding of embarrassment resembles the present finding of unexpectedly good performance by children with autism on our hidden emotion task. The tasks and types of emotions used in both contexts were also similar: Protagonists were portrayed with negative emotions arising from social discomfort with peers.

In their everyday lives, older high-functioning individuals with autism seem to seek ways to minimize adverse peer encounters. Dissanayake and Macintosh (2003) found that high functioning children and adolescents with autism spent more time alone on the outskirts of the peer group, and less time in reciprocal interaction, than their peers with typical development. Yet, despite this evidence of the social aloofness that is a diagnostic symptom of autism, the autistic participants did make overtures to interact and seemed to enjoy positive peer contact when it did arise. Possibly, onlooker peer experiences, along with painful direct social encounters with situations such as those portrayed in our tasks, as victims of embarrassment or teasing may heighten the high-functioning autistic child's sensitivity to these specific kinds of emotional situations. Such increased attention, coupled with the relatively sound nonverbal cognitive skills that are associated with a diagnosis of high-functioning autism, may enable these children to bypass false belief and to use alternative cognitive heuristics, or work-arounds. These could be especially evident in controlled experimental settings, such as that of our hidden emotion task, as opposed to the press of ongoing peer interaction.

Late-signing deaf children, though delayed in ToM understanding, do not demonstrate the social aloofness and social inappropriateness in everyday interactions that are defining symptoms of autism (Frith, 1989). Instead, late-signing deaf children in schools where sign is used typically enjoy cohesive relationships and easy communication with their signing classmates. Thus, the motivation to work out nonmentalistic ways of coping with hidden emotion may not arise as urgently as for an autistic child. This hypothesis is speculative, but the consistent reversal of sequences in understanding mental states that we found requires some account.

We do not claim that our findings are empirically definitive. The samples of autistic and deaf children we have tested are substantial for an area of research where (because of practical difficulties of recruitment as well as the infrequent population prevalence of these disabilities) samples are often 12 to 15 in a group, and rarely 20 or more. In absolute terms, however, the sample sizes are modest, and further research with additional samples from other countries and other kinds of educational systems would be both welcome and informative. Nevertheless, we do believe that our methods are exemplary. Especially for comparisons across groups with differing diagnoses and delays, the crucial questions concern not just children's performance at one age on one task (no matter how representative that particular age or task); rather, the crucial questions concern developmental trajectories and sequences within an extended domain of understanding. Longitudinal data might be ideal for such questions. However, even for typically developing children, the achievements we focus on in this research take 3 or more years to unfold. For individuals with delay, in the groups we consider, the same achievements may take 8 or 10 years, or even longer, to emerge across the full sequence of steps. Consequently, the use of a developmental scale of the type employed in the current research provides an important methodological substitute for addressing crucial questions about long-term growth trajectories—in this case, about the sequence of steps, amid delay, that deaf and autistic children pass through to develop an everyday understanding of mind.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Arlington, VA: Author.

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. *Sociological Methodology, 15*, 33–80.

Astington, J. W. (2001). The future of theory of mind research. *Child Development*, *72*, 685–687.

Baron-Cohen, S. (1991). Do people with autism understand what causes emotion? *Child Development*, *72*, 385–395.

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

Baron-Cohen, S. (2000). Theory of mind and autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 3–20). Oxford, England: Oxford University Press.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., & Grant, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology, 13*, 379–398.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have theory of mind? *Cognition*, *21*, 37–46.

Bettger, J. G., Emmorey, K., McCullough, S., & Bellugi, U. (1997). Enhanced facial discrimination: Effects of experience with American Sign Language. *Journal of Deaf Studies & Deaf Education*, *2*, 224–233.

Bloom, P., & German, T. (2000). Two reasons to abandon the false belief test as a test of theory of mind. *Cognition*, *77*, 825–831.

Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice*, *16*, 21–33.

Courtin, C., & Melot, A. (1998). Development of theories of mind in deaf children. In M. Marschark & M. D. Clark (Eds.), *Psychological perspectives on deafness* (pp. 79–102). Malwah, NJ: Erlbaum.

deVilliers, J. (2000). Language and theory of mind: What are the relationships? In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 83–122). Oxford, England: Oxford University Press.

deVilliers, J. G., & deVilliers, P. (2000). Linguistic determinism and the understanding of false belief. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 189–226). Hove, England: Psychology Press.

Dissanayake, C., & Macintosh, K. (2003). Mindreading and social functioning in children with autistic disorder and Asperger's disorder. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind* (pp. 213–240). New York: Psychology Press.

Dunn, J. (1994). Changing minds and changing relationships. In C. Lewis & P. Mitchell (Eds.), *Origins of an understanding of mind* (pp. 297–310). Hove, England: Erlbaum.

Dunn, L. M., & Dunn, L. M. (1981). *The Peabody Picture Vocabulary Test: Revised*. Circle Pines, MN: American Guidance Service.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Emmorey, K. (1998). The impact of sign language use on visuospatial cognition. In M. Marschark & M. D. Clark (Eds.), *Psychological perspectives on deafness* (pp. 79–102). Malwah, NJ: Erlbaum.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50*, 21–45.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology, Vol. 2*, 857–898. New York: Wiley.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, England: Oxford University Press.

Gopnik, A., & Slaughter, V. P. (1991). Young children's understanding of changes in their mental states. *Child Development*, *62*, 98–110.

Green, B. F. (1956). A method of scalogram analysis using summary statistics. *Psychometrica*, *21*, 79–88.

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

Happé, F., & Siddons, F. (1994). Autism and theory of mind in everyday life. *Social Development*, *3*, 109–125.

Harris, P. L., Donnelly, K., Guz, G. R., & Pitt-Watson, R. (1986). Children's understanding of the distinction between real and apparent emotion. *Child Development*, *57*, 895–909.

Harris, P. L., & Leevers, H. (2000). Pretending imagery and self-awareness in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 182–202). Oxford, England: Oxford University Press.

Hillier, A., & Allinson, L. (2002). Understanding embarrassment among those with autism. *Journal of Autism & Developmental Disorders*, *32*, 583–592.

Jackson, A. L. (2001). Language facility and ToM development in deaf children. *Journal of Deaf Studies and Deaf Education*, *6*, 158–173.

Leslie, A. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*, 211–234.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity and conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*, 225–251.

Linacre, J. M. (2003). *User's guide and program manual to WINSTEPS: Rasch model computer programs*. Chicago: MESA.

Linacre, J. M., & Wright, B. D. (1994). *A user's guide to BIGSTEPS: Rasch model computer programs*. Chicago: MESA.

Lord, F. N., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lundy, J. (2002). Age and language skills in deaf children in relation to theory of mind development. *Journal of Deaf Studies and Deaf Education*, *7*, 34–56.

Marschark, M. (1993). *Psychological development of deaf children*. New York: Oxford University Press.

Peterson, C. C. (2002). Drawing insights from pictures: The development of concepts of false drawing and false belief in children with deafness, normal hearing and autism. *Child Development*, *73*, 1442–1459.

Peterson, C. C. (2003). The social face of theory of mind. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind* (pp. 171–196). New York: Psychology Press.

Peterson, C. C. (2004). Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *Journal of Child Psychology and Psychiatry*, *45*, 1096–1106.

Peterson, C. C., & Siegal, M. (1995). Deafness, conversation and theory of mind. *Journal of Child Psychology and Psychiatry*, *36*, 459–474.

Peterson, C. C., & Siegal, M. (1998). Changing focus on the representational mind. *British Journal of Developmental Psychology*, *16*, 301–320.

Peterson, C. C., & Siegal, M. (1999). Representing inner worlds: Theory of mind in autistic, deaf and normal hearing children. *Psychological Science*, *10*, 126–129.

Peterson, C. C., & Siegal, M. (2000). Insights into a theory of mind from deafness and autism. *Mind & Language*, *15*, 123–145.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Reed, T., & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective-taking. *Journal of Autism and Developmental Disorders*, *20*, 555–567.

Remmel, E., Bettger, J., & Weinberg, A. (1998, November). *The impact of ASL on theory of mind*. Paper presented at the meeting of TIRSL 6, Washington, DC.

Rieffe, C., & Meerum-Terwogt, M. (2000). Deaf children's understanding of emotion: Desires take precedence. *Journal of Child Psychology and Psychiatry*, *42*, 601–608.

Rieffe, C., Meerum-Terwogt, M., & Stockmann, L. (2000). Understanding atypical emotions among children with autism. *Journal of Autism & Developmental Disorders*, *30*, 195–203.

Russell, P. A., Hosie, J. A., Gray, C. D., Scott, C., Hunter, N., Banks, J. S., et al. (1998). The development of theory of mind in deaf children. *Journal of Child Psychology and Psychiatry*, *39*, 903–910.

Scholl, B., & Leslie, A. (2001). Minds, modules and meta-analysis. *Child Development*, *72*, 696–701.

Scott, C., Russell, P., Gray, C., Hosie, J., & Hunter, N. (1999). The interpretation of line of regard by prelingually deaf children. *Social Development*, *8*, 421–425.

Steeds, L., Rowe, K., & Dowker, A. (1997). Deaf children's understanding of beliefs and desires. *Journal of Deaf Studies & Deaf Education*, *2*, 185–195.

Tager-Flusberg, H. (1993). What language reveals about the understanding of minds in children with autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds* (pp. 138–157). Oxford, England: Oxford University Press.

Taylor, M. (1988). The development of young children's understanding of the seeing-knowing distinction. In J. W. Astington, P. L. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 207–225). New York: Cambridge University Press.

Wellman, H. M. (1993). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami

(Ed.), *Handbook of childhood cognitive development* (pp. 167–187). Oxford, England: Blackwell.

Wellman, H. M., Baron-Cohen, S., Caswell, J., Swettenham, J., Toye, E., & Lagatutta, K. (2002). Thought bubbles help children with autism acquire an alternative to theory of mind. *Autism*, *6*, 343–363.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.

Wellman, H. M., & Liu, D. (2004). Scaling of theory of mind tasks. *Child Development*, *75*, 523–541.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*, 245–275.

Wilson, M. (1989). A comparison of deterministic and probabilistic approaches to measuring leaning structures. *Australian Journal of Education*, *33*, 127–140.

Woolfe, T., Want, S., & Siegal, M. (2002). Signposts to development: ToM in deaf children. *Child Development*, *73*, 768–778.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: MESA.

Yirmiya, N., Erel, 0., 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.

Appendix

Diverse Desires

[Display adult doll and pictures of carrot and biscuit]: Here is a lady. This lady wants her morning tea. Here are two foods, a carrot and a biscuit.

Pretest Question: Which do you like best? That's a good choice. But the lady doesn't like [biscuits]. She likes [carrots]. She loves to eat [carrots] best of all.

Test Question: So now the lady can choose only one food. Which will she choose? [If no answer prompt: Will she choose a carrot or a biscuit?] (*Correct answer = food the adult likes, always opposite to child's own preference)*

Diverse Beliefs

[Display girl doll and pictures of bushes and garage]: This girl wants her cat. The cat is hiding. It could be in the bushes or it could be in the garage.

Pretest Question: Where do you think the cat is? Well, that's a good idea. But the girl thinks the cat is in [opposite of child's choice].

Test Question: Where will the girl look for her cat? [If no answer: Will she look in the garage or the bushes?] (*Correct answer = opposite place to child's own belief)*

Knowledge Access

[Display toy chest with drawer closed]: Here is a drawer.

Pretest Question 1: What do you think is in it? That's a good guess. Let's open it. Oh, look! There is a dog in it! [Display toy dog; then close it inside drawer.]

Control Question 1: So what is in the drawer? [Doll enters]: This girl has never seen this drawer before. She has never opened it.

Control Question 2: So has she looked in this drawer?

Test Question: Does the girl know what is in this drawer? (*Correct answer = no)*

False Belief

[Display closed Band-Aid box]: Here is a Band-Aid box. What do you think is in it? [If no answer, or answer other than "Band-Aids," tester continues: What is usually in a box like this? In the shops, what does a box like this have in it?] Let's look in the box. Oh! There is a pig in it. [Tester closes pig in box.]

Control Question 1: Okay, so what is in the box? [Boy doll arrives]. Here comes the boy. He has never looked in this box.

Test Question: What does the boy think is in the box? *(Correct answer = Band-Aids)*

Control Question 2: Did he look in the box?

Hidden Emotion

Here is a boy [picture of back of head]. The boy and his friends were playing. A girl teased the boy and the others all laughed. The boy did not laugh. He did not think it was funny. But the boy did not want the others to see how he felt. If they saw how he felt, they would call him a baby.

Real Emotion Question: How did the boy really and truly feel when everyone laughed and teased him? [Emotional face pictures are offered for pointing. If no answer, tester points in turn and says: Did he feel happy? Or okay? Or sad?]

Reality Justification Control Question: Why did he feel [sad/okay/happy]?

Apparent Emotion Test Question: How did the boy try to look on his face when everyone laughed at him and teased him? [Emotional face pictures are offered for pointing. If no answer, tester points in turn and says: Did he try to look happy? Or okay? Or sad?] *(Correct answer = less negative emotion for apparent emotion question [e.g., okay] than was given for real emotion question [e.g., sad])*

Appearance Justification Control Question: Why did he try to look [sad/okay/happy]?