NIH Public Access

Author Manuscript

Dev Psychol. Author manuscript; available in PMC 2014 December 01.

Published in final edited form as:

Dev Psychol. 2013 December ; 49(12): . doi:10.1037/a0032419.

Deafness, Thought-Bubbles and Theory of Mind Development

Henry M. Wellman and
University of Michigan

Candida C. Peterson
University of Queensland

Abstract

The processes and mechanisms of theory of mind development were examined via a training study of false belief conceptions in deaf children of hearing parents (N = 43). In comparison to two different control conditions, training based on thought-bubble instruction about beliefs was linked with improved false belief understanding as well as progress on a broader theory-of-mind scale. By combining intervention, microgenetic, and developmental-scaling methods the findings provide informative data about the nature and mechanisms of theory-of-mind change in deaf children, as well as an initial demonstration of a useful intervention for enhancing social cognition in deaf children of hearing parents. The methods and results also point to possible avenues for the study of conceptual change more generally.

Keywords

social cognition; theory of mind; deafness; intervention research; developmental delays

---

Typically developing children's explicit understanding of mental states develops rapidly in the preschool years. This everyday theory of mind (ToM) includes increased insights into agents' beliefs, desires, emotions and perceptions as well as awareness of how such mental states shape the intentional actions of self and others. Initial implicit understanding of some of these concepts begins in infancy (Brandone & Wellman, 2009; Onishi & Baillargeon, 2005; Woodward, 1998) but developments in the years from 3 to 6 are particularly pronounced and important. During this period, typically-developing children make dramatic gains in hallmark achievements of a representational theory of mind, including mastery of explicit false belief tasks. Progress on these tasks has real-world application. ToM mastery predicts preschool children's social popularity with peers (Peterson & Siegal, 2002; Slaughter, Dennis & Pritchard, 2002), teacher-rated social competence (Astington, 2003; Peterson, Slaughter & Paynter, 2007; Watson, Nixon, Wilson & Capage, 1999), and skilled interactions with peers (Dunn, Cutting & Demetriou, 2000) including abilities to play games like hide-and seek (Peskin & Ardino, 1999) and social pretend play (e.g., Astington & Jenkins, 1999). Preschool acquisition of explicit ToM understanding, as indexed by performance on standard false-belief tasks, constitutes an important social-cognitive achievement.

However, not all children of normal intelligence develop ToM understanding as preschoolers. Happé's (1995) comprehensive summary revealed that for individuals with autism passing false-belief tasks was extremely rare before age 13 and continued to be rare after that. Poor ToM performance for individuals with autism is often taken to reflect specific neurocognitive differences and deficiencies (Frith & Frith, 2010). Yet even in the absence of neurocognitive impairment, severe delays exist for severely and profoundly deaf children of hearing parents (e.g., Courtin & Melot, 1998; Meristo, Falkman, Hjelmquist, Tedoldi, Surian & Siegal, 2006; Peterson, 2002; Peterson & Siegal, 1995, 1999; Schick, de

Villiers, de Villiers & Hoffmeister, 2007). A review of over 20 studies revealed that only a minority of profoundly deaf children with hearing parents passed standard false belief tests even at mean ages of 10 or 11 years (Peterson, 2009). This was true whether children had cochlear implants or communicated in sign versus speech. The deaf children in these studies are free of disabilities apart from hearing loss and are typically of normal intelligence and sociability.

A conversational-communicative (Astington & Baird, 2005; Harris, 2005, 2006) interpretation of the deaf data seems reasonable for several reasons. Deaf children who grow up in hearing families do not usually acquire sign language until school entry and, until then, typically have no one at home who can converse freely with them about unobservable thoughts, feelings and other mental states (e.g., Vaccari & Marschark, 1997). Controlled comparisons reveal that hearing mothers talk less about mental states with deaf than hearing children, despite equivalent amounts of talk about non-mentalistic topics (Moeller & Schick, 2006). In typically developing children, rich exposure to conversational-communicative input about persons’ minds via parental mental-state talk and sibling dialogue correlates with rapid ToM development (e.g., Dunn & Brophy, 2005; Ruffman, Slade & Crowe, 2002), so the paucity of such experiences for deaf preschool children is a potential barrier to false belief understanding.

Indeed, nature has provided a controlled comparison. Although most deaf children are born to hearing parents, about 5% have a deaf parent and are able to converse as early and as naturally in sign with parents and siblings as hearing infants and toddlers do in speech. Such native signers achieve false belief understanding (e.g., Courtin & Melot, 1998; Schick et al., 2007) and other milestones of ToM development (e.g., Peterson, Wellman & Liu, 2005) on the same early timetable as hearing children. Thus the severe ToM delays for deaf children with hearing parents are not a consequence of deafness per se, but rather of growing up deaf in the closed conversational world of a hearing family.

With these considerations as background, we report a training study with deaf children of hearing parents that attempts to engender gains in ToM understanding via systematic intervention. Several questions inspire this research. Primarily and theoretically, what is the basic nature of ToM progress? How is it best characterized, how enabled via environmental input, and how constrained via neurobiological maturation, including, perhaps a critical period (Siegal & Varley, 2002)? Training interventions for ToM-delayed children with deafness could help address these issues. Training interventions with young typically-developing children have boosted success on false belief tests (e.g., Lohmann & Tomasello, 2003; Slaughter & Gopnik, 1996). But this sheds no light on whether interventions are successful later in development, for those with known ToM delays.

High-functioning school-age individuals with autism can be coached to pass specific forms of the false belief task via ingenious and extensive training (e.g., McGregor, Whiten & Blackburn, 1998; Swettenham, Baron-Cohen, Gomez & Walsh, 1996; Wellman, et al., 2002), but serious questions remain as to how generalizable such increases are beyond the particular materials, examples, and questions used in training. For example, Fisher and Happé (2005) implemented a brief (4 to 10 day) training in which 10 children with autism learned to slot pictures into a doll’s head to depict her true and obsolete beliefs. They found significant gain immediately after training on the trained false belief task but no transfer to another equivalent one. Of course, the specific neurocognitive impairments associated with autism could account for lack of progress or lack of generalization in a manner that would apply only to this group.

Deaf children could provide a more informative and generalizable test of training's impact in the context of ToM delay. Deaf individuals have been especially informative about critical periods for language learning (e.g., Newport, 1991); they could be equally informative about ToM learning. Longitudinal evidence of ToM progress by deaf children of hearing parents (e.g., Falkman, Roos & Hjelmquist, 2007; Peterson, 2009; Wellman, Fang & Peterson, 2011) hints that ToM progress is possible for some children even as late as mid-adolescence. But these longitudinal data are sparse, and moreover training interventions supply an important complement to longitudinal investigations (Bradley & Bryant, 1983; Harris, 2005).

In the preschool years, transition to a representational theory of mind is well documented for typically developing children, but extant research mostly yields "before" and "after" snapshots of development. Cross-sectional studies depict typical 3-year-olds who fail false belief and 4- and 5-year-olds who pass a variety of explicit, inferential tasks. Training studies similarly identify preschoolers who fail false belief and then provide a posttest snapshot of whether or not false belief was mastered (e.g., Lohmann & Tomasello, 2003). Standard longitudinal studies, because re-testings have employed intervals of 6, 9, or 12 months long (e.g., Dunn & Hughes, 1998; Ruffman et al., 2002) capture performance before and after the preschool transition to false-belief success but typically fail to provide rich information about how the transition unfolds. One way to capture developmental change more richly is through multiple sessions with the same children over a protracted transition, thereby achieving a more microgenetic record of change (Kuhn, 1995; Siegler, 1995). We adopt that approach here using a quasi-microgenetic method with weekly training sessions over 2 months. Moreover, we couple this approach with use of a developmentally-ordered ToM Scale (Wellman & Liu, 2004) which provides a more extended, and genuinely progressive developmental assessment of children's ToM competence.

False belief understanding is one milestone amid an extended progression of ToM insights, as is clear from research using a ToM scale (Wellman & Liu, 2004) that encompasses carefully constructed tasks assessing understanding of (1) Diverse Desires (people can have different desires for the same thing), (2) Diverse Beliefs (people can have different beliefs about the same situation), (3) Knowledge-Access (something can be true, but the uniformed do not know that), (4) False Belief (something can be true, but someone might believe something different), and (5) Hidden Emotion (someone can feel one way but display a different emotion). These tasks are similar in procedures, language, and format, yet U.S. (Wellman & Liu, 2004; Wellman, et al. 2011), Australian (Peterson, Wellman, & Liu, 2004; Shahaeian, et al. 2012), and German (Kristen, et al., 2006) preschoolers evidence a clear order of difficulty (just as listed above). A similar 5-step developmental progression, at the same rate, characterizes Chinese (Wellman, et al. 2006) and Iranian children (Shahaeian, et al., 2012).

Using this scale in combination with microgenetic training research can provide especially informative data concerning the nature of conceptual changes. One crucial example concerns accounting for variability in children's responses to cognitive interventions. To illustrate, Lohmann and Tomasello (2003) compared several training conditions with young preschoolers all of whom failed false belief at pretest. In the most successful condition (*full training*) on average children performed substantially better at posttest on a set of three false belief tasks, however, post-test success ranged from 0 to 3 among individual children. In Amsterlaw and Wellman's (2006) microgenetic study, children received scripted theory-of-mind experiences (over multiple sessions across 6 weeks) and this enhanced false belief performance relative both to a pretest and control groups. Nevertheless there was again considerable individual variation (see also Flynn, O'Malley, & Wood 2004). Why do some children gain but others, exposed to the same conditions, do not? Novel mixes of scaling, training, and microgenetic-longitudinal methods could better inform us about this question which is fundamental to any theoretical account of the nature of cognitive change.

Our training approach used pictorial input in cartoon-like format to try to convey mental-state insights to deaf children. Specifically, we used a thought-bubble program that has yielded proven, albeit limited, success for children with autism (Wellman, et al., 2002). Previous research shows that children with deafness (and autism) may be more responsive to pictorial stimuli than hearing preschoolers (e.g., Peterson, 2002). For hearing children, reference to thoughts as representational mental states typically occurs via language and conversation (Astington & Baird, 2005; Dunn & Brophy, 2005; Harris, 2005); for deaf children pictorial input specifically focused on thinking could usefully augment and clarify such language. Moreover, everyday media that feature pictures along with print--comics, storybooks, cartoons and graphic novels (e.g., Manga)--are very popular with deaf children in the age range we tested. Thought-bubbles could thus provide a useful tool for helping deaf children to understand representational mental states, including differences among persons in the mental representations they may hold. As tangible, readily comprehensible, depictions of the mental states (like false belief) that hearing parents have difficulty talking about with their deaf child (Moeller & Schick, 2006), thought bubbles may give deaf children new means for thinking about these states and discussing them with others.

Overview

School-age deaf children in a thought-bubble training group were compared to two control groups of deaf children. The baseline control group had pretests and posttests as a control for any spontaneous developmental gains over the study's time frame. To control for general benefits from practice in discussing and working on visual-representational projects, a non-ToM training group received an extended program of training that was similar in procedural format to the thought-bubble training but with content involving artistic representation of visual information rather than thought bubbles representing mental states.

Method

Participants

A total of 43 Australian signing deaf children of hearing parents aged 5 to 13 years took part: 13 children (mean age: 9.8; range: 7.8–13.0) in the ToM training group; 16 (mean age: 9.8; range: 6.8 –13.2) in the baseline control group; and 14 (mean age: 8.7; range: 5.7–12.2) in the non-ToM intervention control group. Table 1 summarizes background information on these children. There was no significant difference among the groups' mean age at pretest, $F(2, 40) = 1.61$, $p = .21$, and gender balance was also equivalent: $\chi 2\ (2) = 1.16$, N = 43, $p = .56$. All children had severe or profound prelingual hearing losses and were being educated in special units for hearing impairment within large public schools. In these units Signed English (91%) or Auslan (9%) was the primary mode of instruction. The same educational program, delivered by teachers with identical training, qualifications and language approaches, was being followed by all groups though at different school campuses.

Twenty-one children (49%) had a cochlear implant, 6, 7 and 8 in ToM training, baseline, and non-ToM training groups, respectively: $\chi 2\ (2) = .34$, N = 43, $p = .84$. All children were academically capable, as evidenced not only by their uniform placement in the primary school grade that would be expected for hearing children their age, but also by their teachers' reports that they were coping satisfactorily with that grade's national academic curriculum. All children had acquired good everyday communication skills in their preferred sign language modality, according to both their teachers' reports and their uniformly high success on our comprehension control questions.

Two school units for hearing impairment, essentially the same apart from geographical location, supplied all children in the sample. Children in the ToM training group came from School 1, and the baseline control children came from School 2. The non-ToM training group came equally from Schools 1 and 2 (n= 6 and 7, respectively), consisting of a later cohort who had had no involvement with either of the other groups. Teachers at Schools 1 and 2 were closely matched in age, experience, training and qualifications; all carefully adhered to a shared, government-mandated approach to language-teaching and classroom-communication. Indeed, the academic curriculum was the same in both units and teachers from each school met regularly to compare notes and exchange ideas. Given that the two schools offered essentially the same program, most parents simply chose between them on the basis of proximity. This was unlikely to have introduced any systematic bias since the neighborhoods around Schools 1 and 2 were highly similar in all major demographic respects (Australian Bureau of Statistics, 2011). For example, their median weekly household incomes were $1472 and $1464, respectively (both above the national average of $1234), they had similar household composition and family sizes (1.7 and 1.8 children, respectively), similarly low levels of unemployment (6.7% and 5.8% respectively) and similarly high levels of private home ownership (66% and 69% respectively).

In sum, although assignment of children to groups was school-based not random, the close similarity between Schools 1 and 2 argued against school differences having systematic effect on our findings. Moreover, a final indicator of the similarity between Schools 1 and 2 was the absence of any pretest group differences, as we detail in the results, including no differences between the non-ToM-training subgroups at School 1 and 2.

Experimental Design and Procedures

For testing and training, two adults were present: an experienced male experimenter (the same for all children), and a professionally-qualified female interpreter of sign language. The interpreter differed according to the school attended. Yet, all were equivalent in being familiar school staff, well-known to children, and highly fluent in the child's preferred sign language (Signed English or Auslan). All interpreters were professionally-qualified sign language interpreters and all had full certification by NAATI, the Australian "National Accreditation Authority of Translators and Interpreters" (NAATI Ltd, 2011). In each case accreditation had been achieved at the highest level specified by NAATI (namely "Translator and Interpreter"—formerly known as Level 3) as deemed optimal for complex work in educational settings.

The same female interpreter was used for all of any one child's testing and training sessions. However, no interpreter was involved with more than one of the three experimental conditions; thus all were blind both to the hypotheses of the study and to the other types of training being used. The interpreter sat beside the experimenter, directly opposite the participant in full view, and provided an accompanying translation of the experimenter's speech in the child's preferred mode of sign language, using a style of interpretation very familiar to these children from everyday school routines. The interpreter and experimenter paused while critical bits of stories were acted out (e.g., a doll's entry onto the scene) to avoid problems of divided attention, and both adults monitored that the child's gaze was directed at the props or the interpreter, as appropriate.

In total, children received fully bilingual input conforming to a well-practiced routine that was a common part of their everyday classroom instruction. This bilingual presentation was preferable to signing alone because many children had some residual (or implant-assisted) hearing, or some lip-reading facility, and so were able to profit from this bimodal exposure to the tasks and training. At the same time, most pretests and posttests permitted a completely nonverbal response (e.g., pointing) and the same was true of many questions asked during training (see on-line Appendices)

At pretest, all children took a battery of ToM tasks and a language measure. Then, for the ToM training group, a two-week school holiday intervened, after which each child took part individually in six thought-bubble training sessions, each roughly half an hour long and conducted at about the same time for each child each week across a training period of exactly six weeks duration. The non-ToM training group also had weekly half-hour training sessions, each child with the male experimenter and a constant, familiar interpreter, over a similarly extended period (M = 6.00 weeks, range = 5 to 7 weeks) but their training addressed artistic production rather than mental representation. Those in the baseline control group had no intervention but took the pretest and posttest after a time lag equivalent to that of the other groups (see Table 1).

ToM and non-ToM Training

The ToM training concepts, materials and procedures were closely modeled on the thought-bubble program that Wellman et al. (2002) developed for children with autism. Training focused on changed-location false belief situations and two-dimensional cardboard objects (cut-out dolls, thought bubbles, and containers and rooms with opening paper flaps) were used throughout training (see on-line Appendix A). All of these were very different from the three-dimensional dolls, props and stimuli used for the ToM pretests and posttests. The two adults conducted the training (as well as all pre-and post-testing) in weekly individual sessions in a quiet room at school.

For the ToM-training children, each week was devoted to one of six sequential stage concepts shown in on-line Appendix A. At each stage, the children continued with the training until a pre-established success criterion was reached (see Appendix A). All children achieved this success criterion for each stage within the specified weekly session, but with somewhat different session lengths for different children (averaging about 30 minutes) depending on the number of repeated trials a child needed to achieve that week's stage concept. Later stages generally demanded more training than earlier ones did. No child required any extra trials at stage 1; most (77%) did so for most later stages. For example, the extra trials needed on stages 5 and 6 ranged from 0 to 8 and 0 to 12. A measure used in statistical analyses was each child's total extra training trials across the six training weeks. Comprehension control questions (e.g. "What is really in the red box?") were interspersed throughout the training and children did uniformly well on these. On rare control-question failures, the preceding demonstration and all questions were repeated.

The non-ToM training group also had weekly sessions of about 30 minutes duration, presented in the same bilingual manner in the child's preferred modality of sign. As illustrated in on-line Appendix B, these children received a different art project each week, always beginning with a demonstration phase (just as for the ToM-training) followed by a phase of individual coaching using questions and corrective feedback (similar to that for ToM training but with no ToM content). Each week's art activity was decomposed into 4 or 5 simple steps, similar to ToM training. The type of art activity changed each week; projects required both two-dimensional picture creation (e.g., making a colored paper collage of a peacock) and three-dimensional representations (e.g., an owl made from clay). Like the ToM training group, all children managed to complete each art project within each week's session, even though the amount of individual teaching needed differed from child to child.

Measures

Language ability pretest—Children's general language comprehension was assessed at pretest using the syntax subscale ("Sentence Structure") of the Clinical Evaluation of Language Fundamentals Preschool (CELF-P) test (Wiig, Secord & Semmel, 1992). This 22-item scale is designed for hearing children aged 5 to 8 years and uses nonverbal (picture-pointing) responses to assess a broad range of semantic, morphological and syntactic concepts including plurality, verb tense, relative clauses and embedded propositions. Children took it in their preferred modality of signing (95% Signed English). We used raw scores as our dependent measure, as recommended by the test manual. Indeed, no age norms are available for individual subscales of the CELF-P and none are available for deaf children for any aspect of the test. Exactly the same approach (i.e., use of only one of the six CELF-P subscales and reliance on raw scores rather than age-equivalents as potential ToM correlates) has been taken effectively in one past study of deaf children (Peterson, Wellman & Slaughter, 2012) and at least eight past ToM studies of hearing children (see Milligan, Astington & Dack, 2007, Table 1, for details). In these, the correlation with false belief has generally been significant, as it was for our sample, r (38) = .71, $p < .001$. Furthermore, score ranges (13 to 22) and the absence of floor or ceiling effects (only 8 children--19%--scored 20 or higher and all scored at least twice what would be expected by chance) support the test's suitability for our sample. Groups did not differ significantly from one another, F (2,39) = 2.18, $p > .125$, indicating comparable general language ability. All children had complete language data except for three in the baseline control group. Child reluctance (2 cases) or a scheduling problem (1 case) precluded administration of the language test to these three children.

Pretest and posttest ToM measures—Theory of mind (ToM) understanding at pretest and immediate posttest was measured for all three groups using ToM batteries comprising (a) a three-item false-belief composite (Total False Belief, TFB) and (b) the five tasks of Wellman and Liu's (2004) broader ToM Scale. The three false belief items consisted of the two changed-location items of Baron-Cohen, Leslie and Frith's (1985) Sally-Ann task and the misleading container (bandaids box) item from Wellman and Liu's scale. All tasks were presented and scored as in the original publications, and included a focal test question about a person's false belief and one or two language/comprehension control questions. As in the original publications, a pass for each item required accuracy on controls as well as on the test.

Wellman and Liu's (2004) five-task ToM Scale was administered and scored exactly as in Peterson et al. (2005). Five sequentially-developing ToM concepts were assessed: (1) Diverse Desires (DD), (2) Diverse Beliefs (DB), (3) Knowledge Access (KA), (4) misleading container False Belief (FB), and (5) Hidden Emotion (HE). These tasks are carefully matched in procedures, stimuli, and questions, and research has demonstrated that contrasts in performance on them cannot be explained away by executive function, or language performance alone (Wellman et al. 2008; Peterson et al. 2012). Each task included a warm up question, a focal test question and a comprehension control question, all of which must be answered correctly to pass. Scale totals (0–5) reflect the total steps passed correctly.

We further examined fine-grained changes from pretest to posttest via several subsidiary variables. Because our thought-bubble training was structured around the concept of false beliefs about object locations, to measure exactly this trained concept (although tested with different stimuli and no thought bubbles) we examined gains on the two Sally-Ann items. The misleading-container false-belief item was also examined separately as an index of "near-generalization". To examine changes in understanding ToM concepts *not* including false belief, we computed two additional subtotals from the ToM Scale. A 3-item pre-false-belief (Pre-FB) total consisted of just the items (DD, DB and KA) that are routinely mastered ahead of false belief by both deaf and hearing children. A 4-item index of “far-generalization” to other ToM concepts besides false belief was the sum of DD, DB, KA, and HE.

Post-tests—The posttest was given to all children in each of the training groups within one week of completing their final training session and after an equal lapse of time from pretest for the baseline control group. The actual intervals separating pretest from immediate posttest did not differ significantly between ToM training and non-ToM training groups, t (25) = 1. 61, p >.10, nor between ToM training and the baseline control group, t (27) = 1. 73, p = .095.

Results

Pretest Comparisons among Groups

As outlined in Table 1, before training began the groups were closely matched in age, linguistic skill, and other key background variables. As shown in Table 2, there were no significant pretest differences on any ToM measure. For total false belief (TFB), the omnibus ANOVA was nonsignificant, F (2, 40) = 0.78, p = .465, and pairwise t-tests showed no significant contrasts between ToM versus non-ToM training groups, t (25) =1.11, p = .279, nor between ToM training and baseline control, t (27) = 0.08, p = .940. The same was true on the specific false belief concept to be trained, F (2, 40) = 0.67, p > .50 (pairwise contrasts were t (25) =1.07, p = .297 and t (27) = 0.15, p = .884, respectively) and on a different (container) false belief test, $\chi 2$ (2) = 1.20, N = 43, p =.548. ToM Scale totals were also equivalent for all three groups at pretest, F (2, 40) = 0.36, p > .60; pairwise contrasts showed no difference between ToM training versus non-ToM training, t (25) =0.32, p = .749, nor between ToM training and baseline control, t (27) = 0.75, p = .462. (Nor were there any significant pretest group differences on any individual ToM Scale task: $\chi 2$ (2) ranged from 1.01 to 3.47, N = 43 (ps ranged from .176 to .602).

Note in Table 1, however, that children had a wide range of ages at pretest (5 years 8 months to 13 years 2 months) and that mean group ages differed by as much as 1 year (although, to reiterate, they did not significantly differ). Thus, for the main analyses we used ANCOVA with age as the covariate to compare the groups’ ToM scores, followed up with simple planned comparisons as recommended by Keppel (1973). We accepted for training only children who failed multiple ToM tasks at pretest. Thus all children in both training groups had room to progress. However, a few children in the ToM-training group (n= 2), the non-ToM-training group (n = 1), and the baseline control group (n = 4), did pass all three pretest false belief tasks. In most analyses we retained these children both because they had failed other pretest ToM Scale tasks and because they could conceivably do worse at posttest, in line with previous findings both for hearing preschoolers (Flynn, et al. 2004) and older deaf children (Falkman, et al., 2007). In fact, four children two in each control group with pretest TFB scores of 3 earned scores of 2 at posttest. Nevertheless, for several follow-up analyses we looked at a smaller, more conservative subgroup of children, namely the “reduced” sample (n = 36) who failed at least one (and typically most or all) of the pretest false belief tasks.

Group Differences after Training

Table 2 shows children’s performance on the posttest, given without any thought-bubble supports. ANCOVA (with age as a covariate) for the total false belief measure, posttest TFB, revealed a significant overall group difference, F (2, 39) = 5.65, p = .007. The ToM training group performed significantly higher than both the non-ToM training group (p = .

002) and the baseline control group (p = .028). ToM-trained children outperformed the other groups on changed location false belief tasks (similar to the training emphasis) in an overall ANCOVA, F (2, 39) = 5.01, p = .012, and in direct comparison outperformed the non-ToM training group (p = .003). Moreover, there was an identical group pattern on the misleading container task (a task dissimilar to the ones used in training) $\chi^2(2)$ = 6.77, N = 43, p = .034.

Posttest ToM Scale scores likewise differed among the groups, F (2, 39) = 6.33, p = .004; again the ToM training group outperformed the non-ToM training group (p = .002) and the baseline control (p = .007). This difference represented more than merely improvement in the false-belief task embedded in that scale; the ANCOVA for the 4-item “far generalization” scale total (removing FB) yielded a significant overall posttest difference among the groups, F (2, 39) = 5.08, p = .011, with ToM training significantly surpassing both non-ToM training (p = .005) and baseline control (p = .014).

To confirm these results with children who were the most incorrect on false belief at pretest, we examined the three “reduced” groups. ANCOVA (with age controlled) again revealed significant posttest group differences for total false belief, F (2, 32) = 8.25, p = .001, with the reduced ToM training group outperforming both the reduced baseline control and the reduced non-ToM training groups (p = .004 and .001, respectively). This same pattern was evident for both changed-location (F (2, 32) = 6.23, p = .001: planned-comparisons ps = .014 and .002, respectively) and misleading container FB tasks, $\chi^2(2)$ = 7.64, N = 36, $p < .05$. On the scale total at posttest, there was likewise a significant overall group difference F (2, 32) = 8.75, p = .001, and again the thought-bubble training group outperformed both the reduced baseline and reduced non-ToM training groups (p = .002 and p = .001). Identical patterns appeared for the 4-item “far generalization” score, F (2, 32) = 5.45, $p < .01$ (p = .010 and .005, respectively).

We computed individual gain scores for each ToM variable as the difference between a child’s score at pretest versus posttest. Results echoed the patterns reported above. For the TFB and changed-location totals, group differences in posttest gains were highly significant, F (40) = 14.30 and 6.35, respectively, $p < .001$ and .005. The two control groups did not differ (p = .491 and .460), but the ToM-training group significantly outperformed each of them, all ps < .005. Similarly, for the ToM-Scale and the far-generalization score, there were significant group differences in gains, F (40) = 13.54 and 7.97, respectively, both ps < .001. The two control groups did not differ (p = .285 and .316), but the ToM-training group significantly outperformed both (all ps < .006).

Comparisons within the ToM-training Group

As these gain scores indicate, thought bubble training (and it alone) resulted in significant increases over pretest performance. Moreover, these results show training effects were widespread; deaf children in the ToM training group made significant gains on a new type of false belief task not used in training and further on the ToM Scale (where they gained a mean of 1.38 over their pretest levels). For the “reduced” ToM training group, the subgroup most incorrect on false belief tests at pretest, the pattern for all these pre- to post-test gain analyses was very similar. For example, these children recorded significant posttest gains on TFB and in total ToM Scale steps passed, $t(10)$ = 5.16 and 6.25, respectively, ps < .01.

Individual differences in ToM gains made after training—Figure 1 shows ToM Scale scores for individual children. All but two children (85%) in the ToM-training group had upwardly-sloping lines (i.e., gains of one ToM-Scale step or more) and none regressed. Most of the control children (73%) failed to change (flat lines: 43%) or regressed (30%). Also clear in Figure 1 is that although ToM-trained children progressed, they differed in starting points and total progress. What accounts for this variability? Here we consider three factors. First, we considered pretest performance on the ToM Scale, a key measure of how far along a child was in ToM understanding at the start. Additionally, for comparison and control, we considered children 's language competence and the amount of training they received during training sessions themselves. Focally, there was a large, significant positive correlation between posttest total false belief, TFB, and pretest performance on the ToM Scale, *rho* = .80, $p = .001$. There was a similar correlation of ToM Scale scores on posttest and pretest, *rho* = .68, $p = .011$. There were no significant correlations of language ability with either posttest TFB—*rho* = .47, $p = .107$—or posttest ToM Scale—*rho* = .38, $p = .198$. Extra thought-bubble training was significantly negatively correlated with posttest TFB = *rho* = −.73, $p = .004$—although not posttest ToM Scale—*rho* = −.43, $p = .141$. Thus, children who needed more training during the thought-bubble sessions (extra thought-bubble trials) did worse at posttest on false belief understanding than those who did not.

To better assess how close or far children were from *false-belief* understanding at pretest (albeit failing false-belief itself at that point) we scored children on just DD, DB, and KA combined (Pre-FB), those conceptual steps that regularly precede false belief developmentally for deaf and hearing children (e.g., Wellman et al., 2011). There was a highly significant correlation for children who had ToM training between Pre-FB scores at pretest and their TFB performance at posttest, *rho* = .86, $p < .001$. Pretest Pre-FB also correlated significantly with posttest ToM Scale scores, *rho* = .59, $p = .033$.

To assess the independent contributions of separately significant predictors, we conducted two hierarchical regression analyses focused only on TFB. We did not consider ToM Scale scores further because from raw correlations alone, Pre-FB was the only significant predictor of ToM Scale scores at posttest. We restricted our hierarchical regressions to two predictor variables so that our sample size (n=13) would exceed the required minimum (n=10) for a two-predictor hierarchical regression (see Tabachnick & Fidell, 1989, p. 129).

First, consider the relevant contributions of Pre-FB and extra thought-bubble training to post-test TFB. With extra training trials as the Step 1 predictor, the equation was significant, *Mult R* = .71, $R^2 = .51$, *Adj.* $R^2 = .46$, $F\ (1, 11) = 11.42$, $p = .006$. Entry of Pre-FB at Step 2 produced a further significant increment, F (*change*: 1, 10) = 19.52, $p < .001$, and a significant full equation *Mult R* = .89, $R^2 = .80$, *Adj.* $R^2 = .76$, $F\ (2, 10) = 14.07$, $p = .004$. Beta weights in the final model showed that *only* Pre-FB was independently significant β = .87, $p = .004$. (For extra training: β = .−02, $p = .920$).

Conceivably, Pre-FB could be just a proxy for children's language competence. As noted earlier, across all three groups false belief performance did correlate with CELF language scores (*rho* = .71, $p < .001$). For children in the ToM training group alone language ability did not correlate significantly with TFB (see above), but it was substantial—*rho* = .47, $p = .107$. Therefore we explored the independent contribution of Pre-FB along with language to TFB. With language ability entered at Step 1, the result was nonsignificant: *Mult R* = .42, $R^2 = .18$, *Adj.* $R^2 = .10$, $F\ (1, 11) = 2.39$, $p > .10$. Pre-FB scores, entered at Step 2, resulted in a significant increment, F (*change*: 1, 10) = 33.55, $p < .001$, and a significant overall equation, *Mult R* = .90, $R^2 = .81$, *Adj.* $R^2 = .77$, $F\ (2, 10) = 21.50$, $p < .001$. Pretest performance on the Pre-FB scale tasks was separately significant in the final model, β = .982, $p < .001$, but language ability was not: β = −.15, $p = .388$. Thus, how advanced children had been before training in understanding ToM concepts that developmentally precede false belief was the best predictor of posttest false belief understanding after thought-bubble training.

In short, children who were equally incorrect on false belief at pretest responded differentially to training depending on how advanced they were on the other concepts comprising the ToM Scale. Our design enabled us to sensitively test this possibility because, at pretest, roughly half the sample as a whole was “closer” and the others “further” from focal false belief understanding (for the focal thought-bubble training group 54% had proceeded as far as KA, as had 43% of the control groups).

Discussion

Although most deaf children growing up in hearing families are severely delayed in ToM development (see Peterson, 2009, for a review), it is clear from our findings that these setbacks are not intractable. An intensive 6-week program of learning to use thought bubbles to talk about and represent beliefs resulted in dramatic improvements on a range of standard ToM tasks. Not only did these deaf primary-schoolers (a) gain understanding of the ToM concepts focal to their training (changed location false belief), but they also (b) generalized their gains to a new type of false belief task (misleading container), and (c) progressed significantly on a broad developmental scale of ToM concepts that are precursors to, and consequences of, false belief understanding. In fact, ToM-trained children gained an average of more than one full step on the ToM Scale after their 6-weeks’ work with thought-bubbles. Longitudinal evidence (Wellman et al., 2011) suggests that, without the benefit of training, the same amount of gain would require at least two years for a deaf child to achieve.

The non-ToM intervention group helps confirm that gains by the ToM- trained group related specifically to our ToM training. It is not definitive that our thought-bubble form of ToM training was crucial, but it was significantly helpful beyond mere passage of time (baseline control) and beyond general, interactive-instructional experiences with the experimenter and interpreters (non-ToM training group). It is important to note that our ToM training not only employed thought-bubbles but used these pictorial devices to provide and elicit talk about mental states, including frequent use of mental-state terms like *think*, *know*, *want*, *see* (see on-line Appendix A). Mental-state language, separate from thought-bubbles themselves, may well have contributed to ToM gains by these children. Indeed, as noted in the introduction, language mastery and opportunities to discuss thoughts with others are clearly bound up in ToM growth generally. Thus we favor a conversational interpretation of the ToM delays typically observed in deaf children of hearing parents.

Interpretation of language effects for ToM typically pinpoint three factors: syntactic structures (e.g., deVillers, 2005), mental-state terms (Bartsch & Wellman 1995; Ruffman, et al. 2002), and the perspective-shifting requirements of conversational interchange (Harris 2006; Lohmann & Tomasello, 2003). For assessment purposes, our standard language measure focused primarily on syntactic competencies and non-mental-state vocabulary. However, given past findings that exposure to mental-state terms is selectively impaired when parents of deaf children are hearing (Moeller & Schick, 2006) plus the fact that conversational interchanges about persons’ minds and perspectives are longitudinal predictors of ToM development in hearing preschoolers (e.g., Ruffman et al., 2002), we believe that the increased talk about mental-states entailed by our thought bubble training regime (see Appendix A) was the most likely language-based contributor to the gains we observed. Making mental-state reference more frequent and concrete via pictorial representations like thought-bubbles could certainly aid in the instigation, clarity, and effectiveness of mental-state discourse. Similarly, it could aid in making salient the different perspectives of different persons within a social situation. This hypothesis about the interplay between linguistically-mediated mental-state conversations and visually-accessible depictions of mental-states is certainly worthy of future research.

From a methodological standpoint, inclusion of Wellman and Liu’s (2004) ToM Scale as a pretest and posttest measure had clear value. At posttest it demonstrated that the gains in

ToM understanding made by ToM-trained deaf children were not confined to false belief alone, but extended to developmentally earlier and later steps in the progressive appreciation of others' minds. This indicates that the novel understanding emerging from ToM training was both genuine and in conformity with a natural developmental trajectory, although at a much faster pace. Pretest scale scores proved even more revealing. Children systematically failed false belief at pretest. Yet their "closeness" to false belief understanding, as measured by their advancement along the ToM Scale, definitely shaped their progress. In regression and correlation analyses this was by far the largest single factor accounting for posttest gains and did so even after language competence and total training trials were controlled.

This pattern of performance helps answer the intriguing question of why same-age children, equally poor initially at understanding false belief, responded so variably to ToM training. Two distinct possibilities could account for such individual variations. One concerns general cognitive factors "external" to the domain of theory of mind —perhaps some children were more attentive, thoughtful, or had better memory and so learned more. A complementary possibility concerns factors "internal" to ToM conceptions themselves—even given consistent failure at pretest, some children may have had a more advanced conceptual foundation about persons and minds and so progressed more, building upon these early insights. This is not a new proposal, it is the basis of Piaget's (1970) constructivist proposals, but direct evidence for it spanning genuine long-term conceptual change, is in surprisingly short supply. Our data clearly exemplify two empirical signatures of such a process of conceptual development: (1) that prior conceptual knowledge influences the presence and amount of learning, and (2) that learning proceeds in orderly progressions. Children's prior understandings both enabled learning (for children closer to false belief on the ToM Scale at pretest) and constrained it (for those further away). Whether closer or further, however, children's progress proceeded through an ordered set of intermediate understandings (see Figure 1). For those further from an understanding of false belief, being trained about false beliefs via thought bubbles did not often engender an understanding of false belief, but it provoked other "earlier" ToM understandings anyway.

It is worth considering how prior understandings worked to facilitate false belief acquisition. Our measure of prior understandings was children's performance on DD, DB, and KA scale tasks prior to training. Higher scores on this measure thus reflect better progress toward understanding knowledge access, KA, in particular and thus better understanding of the relationship between information access and resulting mental states (while still failing FB). We speculate that understanding this relationship, acquired in the case of understanding KA, aids recognition of the role of perceptual evidence in belief formation as well (and this also explains why KA reliably precedes FB understanding in cross-sectional and longitudinal theory of mind scaling research, e.g., Wellman, et al. 2011). Information access was also a focus of our later ToM training sessions. Others have speculated on the formative role of understanding perceptual access as a conceptual prerequisite for understanding FB evidence (see Gopnik & Wellman 1994).

Factors external to theory of mind, rather than prior ToM understandings, might nonetheless constrain ToM progress to a critical period. However, the present findings, along with the longitudinal data from Wellman, et al, (2011) and the follow-up data on adult Nicaraguan signers reported by Pyers and Senghas (2009), are not consistent with a preschool (i.e., ages 2 to 6 years) critical period for ToM development. What our data add, in particular, to this emerging set of findings is needed evidence from a systematic training study.

Clearly our initial study has limitations. Our sample, while sizable for research with delayed populations, was modest. We chose to use different interpreters across the three conditions. While advantageous in keeping interpreters blind to our focal variables and hypotheses, this added a further dimension of difference between the three groups. Further, we did not randomize children to conditions; we arbitrarily assigned different schools to conditions. This raises an intriguing possibility that warrants systematic examination in future research: Did our thought-bubble training achieve its effectiveness in part because an entire group of familiar classmates all received their training simultaneously over a lengthy time frame allowing for ample conversation outside the training sessions? That is, thought-bubble training might have stimulated children in our sample to talk to each other about mental states and pictorial depiction of thoughts as their training progressed, activating for them the known benefits of shared conversation for deaf and hearing children's spontaneous ToM mastery. Indeed, Meristo et al. (2007) found that deaf *native* signers attending oral-only primary schools lagged substantially behind deaf native signers at bilingual (sign-plus-speech) schools in ToM performance, suggesting that conversations at school are helpful even when opportunities for dialogue with natively-signing deaf family members at home are unimpaired. In short, further research is needed to replicate and better understand the intervention effects demonstrated by our results.

Nonetheless, our findings demonstrate that deaf children's ToM delays can be ameliorated via thought bubble-supported conversation. In contrast, more limited success emerged from two previous studies that applied a very similar thought-bubble training program to children with autism (Paynter & Peterson, 2012; Wellman, et al., 2002) including little evidence for generalization to false belief concepts beyond the one focal to training. Our results with deaf children are very different. The gains these children made were not only durable but also widely generalizable to tasks and ToM-scale concepts beyond the training examples.

Pragmatically, one next question is whether it is possible to go beyond boosting social cognition to applied intervention on behalf of deaf children's everyday social skills and social behavior. Can thought-bubble training of the type we implemented ultimately assist deaf children not only to understand others' minds but also to apply this understanding in their everyday interaction? The need here is a real one. As compared with their hearing peers, deaf pupils often have problems with numerous aspects of everyday pragmatic social communication and conversational exchange (e.g., Most, Shina-August & Meilijson, 2010) and are often found to have serious social difficulties both inside and outside the classroom (e.g., loneliness, social exclusion, victimization, and social immaturity) as well as more peer relationship problems than matched hearing children (e.g., lack of mutual friends, low sociometric popularity, social reticence: see Kluwin, Stinson & Colarossi, 2002, for a review). To the extent that these difficulties reflect social-cognitive deficits in understanding their interactional partners' mental perspectives and communicative intentions, ToM-training via thought bubbles might prove a valuable addition to the repertoires of teachers, parents and therapists working to boost deaf children's pragmatic conversational skills and social participation. Of course, the link may well be bi-directional. Until they enter a compatible conversational environment (e.g., a signing school) deaf children of hearing parents are likely to have little practice with the kinds of everyday discourse and interactive use of language to share ideas that can supply mental-state insights (Harris, 2006). As these informal conversational experiences accrue, children may gain increasing readiness to profit from focused ToM training like our thought-bubble program. This, in its turn, may reciprocally open up new opportunities for everyday conversational exchange.

Could thought-bubble training be used with deaf children younger than the school-age children tested here? Typically-developing preschoolers easily understand thought-bubbles by age 3 and 4 years (Wellman, et al. 1996). Moreover, Stages 1 and 2 of our training (see on-line Appendix A) essentially introduced thought bubbles to children and assessed their understanding of them given this minimal introduction. *No* deaf child in our sample needed repeated training or demonstrations on these initial thought-bubble concepts to be correct (and advance on to the further stages). In contrast, 74% of children in the training group did need repeat training and demonstrations to master the ToM concepts introduced in Stages 3, 4, 5 and 6 (e.g., thoughts about unseen changes). However, even at these more advanced stages, none of these necessary additional training trials were prompted by any error by any child in interpreting pictorial thought-bubbles more basically as depictions of protagonists' thoughts. Future research should examine training with younger deaf children.

In sum, the particular kind of ToM training via thought bubbles that we investigated with deaf children, along with the quasi-microgenetic and developmental-scaling methods we used, provide needed theoretical and practical information about ToM development in deaf children of hearing parents. The data additionally offer a number of useful insights about the nature and mechanisms of conceptual change more generally, worthy of further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Preparation of this research was supported by an Australian Research Council award to Peterson and a NIH Merit award (HD022149) to Wellman.

References

Amsterlaw J, Wellman HM. Theories of mind in transition: A microgenetic study of the development of false belief understanding. Journal of Cognition and Development. 2006; 7(2):139–172.10.1207/s15327647jcd0702_1

Astington, JW. Sometimes necessary, never sufficient: False-belief understanding and social competence. In: Repacholi, B.; Slaughter, V., editors. Individual differences in theory of mind: Implications for typical and atypical development. Hove, U.K: Psychology Press; 2003. p. 13-38.

Astington, JW.; Baird, JA. Why language matters for theory of mind. New York: Oxford University Press; 2005.

Astington JW, Jenkins JM. A longitudinal study of the relation between language and theory-of-mind development. Developmental Psychology. 1999; 35(5):1311–1320.10.1037/0012-1649.35.5.1311 [PubMed: 10493656]

Australian Bureau of Statistics. Census quickstats Queensland. 2011. (Downloaded from: http://www.censusdata.abs.gov.au/census_services/get_product/census_2011 on 20 December 2012)

Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a "theory of mind" ? Cognition. 1985; 21(1):37–46.10.1016/0010-0277(85)90022-8 [PubMed: 2934210]

Bradley L, Bryant P. Categorizing sounds and learning to read: A causal connection. Nature. 1983; 301:419–421.

Brandone AC, Wellman HM. You can't always get what you want: Infants understand failed goal-directed actions. Psychological Science. 2009; 20(1):85–91.10.1111/j.1467-9280.2008.02246.x [PubMed: 19037904]

Courtin, C.; Melot, A. Development of theories of mind in deaf children. In: Marschark, M.; Clark, MD., editors. Psychological perspectives on deafness. Malwah, NJ: Erlbaum; 1998. p. 79-102.

Dunn, J.; Brophy, M. Communication, relationships and individual differences in children's understanding of mind. In: Astington, JW.; Baird, J., editors. Why language matters for theory of mind. New York: Oxford University Press; 2005. p. 50-69.

Dunn J, Cutting A, Demetriou H. Moral sensibility, understanding others and children's' friendship interactions. British Journal of Developmental Psychology. 2000; 18:159–177.

Dunn J, Hughes C. Young children's understanding of emotions within close relationships. Cognition & Emotion. 1998; 12(2):171–190.

Falkman K, Roos C, Hjelmquist E. Mentalizing skills of non-native early signers. European Journal of Developmental Psychology. 2007; 4:178–197.

Fisher N, Happé F. A training study of theory of mind and executive function in children with autistic spectrum disorders. Journal of Autism and Developmental Disorders. 2005; 35(6):757–771.10.1007/s10803-005-0022-9 [PubMed: 16283087]

Flynn E, O'Malley C, Wood D. A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. Developmental Science. 2004; 7(1):103–115.10.1111/j.1467-7687.2004.00326.x [PubMed: 15323122]

Frith U, Frith C. The social brain. Philosophical Transactions of the Royal Society B. 2010; 365:165–175.

Happé FGE. The role of age and verbal ability in the theory of mind task performance of subjects with autism. Child Development. 1995; 66(3):843–855.10.1111/j.1467-8624.1995.tb00909.x [PubMed: 7789204]

Harris, PL. Conversation, pretense and theory of mind. In: Astington, JW.; Baird, JA., editors. Why language matters for theory of mind. New York: Oxford University Press; 2005. p. 70-83.

Harris, PL. Social Cognition. In: Damon, W.; Lerner, RM., editors. Handbook of child psychology. Volume 2. Cognition, perception, and language. Hoboken, N.J: John Wiley & Sons; 2006. p. 811-858.

Kluwin T, Stinson M, Colarossi G. Social processes and outcomes of in-school contact between deaf and hearing peers. Journal of Deaf Studies and Deaf Education. 2002; 7:200–213. [PubMed: 15451873]

Kristen S, Thoermer C, Hofer T, Aschersleben G, Sodian B. Skalierung von "theory of mind" aufgaben (Scaling of theory of mind tasks). Zeitschrift fur Entwicklungspsychologic und Padagogische Psychologie. 2006; 38:186–195.

Kuhn D. Microgenetic study of change: What has it told us? Psychological Science. 1995; 6(3):133–139.10.1111/j.1467-9280.1995.tb00322.x

Lohmann H, Tomasello M. The role of language in the development of false belief understanding: A training study. Child Development. 2003; 74(4):1130–1144.10.1111/1467-8624.00597 [PubMed: 12938709]

McGregor E, Whiten A, Blackburn P. Teaching theory of mind by highlighting intention and illustrating thoughts: A comparison with three-year-olds and autistic subjects. British Journal of Developmental Psychology. 1998; 16:281–300.

Meristo M, Falkman K, Hjelmquist E, Tedoldi M, Surian L, Siegal M. Language access and theory of mind reasoning: Evidence from deaf children in bilingual and oral environments. Developmental Psychology. 2007; 43:1156–1169.10.1037/0012-1649.43.5.1156 [PubMed: 17723042]

Milligan K, Astington JW, Dack LA. Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. Child Development. 2007; 78(2):622–646.10.1111/j.1467-8624.2007.01018.x [PubMed: 17381794]

Moeller MP, Schick B. Relations between maternal input and theory of mind understanding in deaf children. Child Development. 2006; 77:751–766.10.1111/j.1467-8624.2006.00901.x [PubMed: 16686799]

Morgan G, Kegal J. Nicaraguan sign language and theory of mind: The issue of critical periods and abilities. Journal of Child Psychology & Psychiatry. 2006; 47:811–819.10.1111/j.1469-7610.2006.01621.x [PubMed: 16898995]

Most T, Shina-August E, Meilijson S. Pragmatic abilities of children with hearing loss using cochlear implants or hearing aids compared to hearing children. Journal of Deaf Studies and Deaf Education. 2010; 7:200–213.

National Accreditation Authority of Translators and Interpreters, Ltd (NAATI). Accrediation of interpreters. 2012. (Copyright NAATI). (Downloaded from http://www.naati.com.au/accreditation.html on 27 December 2012)

Newport, EL. Contrasting concepts of the critical period for language. In: Carey, S.; Gelman, R., editors. The epigenesis of mind: Essays on biology and cognition. Hillsdale, NJ: Erlbaum; 1991. p. 111-130.

Onishi KH, Baillargeon R. Do 15-month-old infants understand false beliefs? Science. 2005; 308(5719):255–258.10.1126/science.1107621 [PubMed: 15821091]

Paynter J, Peterson CC. Further evidence of benefits of thought-bubble training for theory of mind development in children with autism spectrum disorders. Research in Autism Spectrum Disorders. 2013; 7(2):344–348. http://dx.doi.org/10.1016/j.rasd.2012.10.001.

Peskin J, Ardino V. Representing the mental world in children's social behavior: Playing hide-and-seek and keeping a secret. Social Development. 2003; 12(4):496–512.10.1111/1467-9507.00245

Peterson CC. Drawing insights from pictures: The development of concepts of false drawing and false belief in children with deafness, normal hearing and autism. Child Development. 2002; 73:1442–1459.10.1111/1467-8624.00482 [PubMed: 12361311]

Peterson CC. Development of social-cognitive and communication skills in children born deaf. Scandinavian Journal of Psychology. 2009; 50(5):475–483. [PubMed: 19778395]

Peterson CC, Siegal M. Deafness, conversation and theory of mind. Journal of Child Psychology and Psychiatry. 1995; 36(3):459–474. [PubMed: 7782409]

Peterson CC, Siegal M. Representing inner worlds: Theory of mind in autistic, deaf, and normal hearing children. Psychological Science. 1999; 10:126–129.

Peterson CC, Siegal M. Mindreading and moral awareness in popular and rejected preschoolers. British Journal of Developmental Psychology. 2002; 20:205–224.

Peterson CC, Slaughter VP. Telling the story of theory of mind: Deaf and hearing children's narratives and mental state understanding. British Journal of Developmental Psychology. 2006; 24(1):151–179.10.1348/026151005X60022

Peterson CC, Slaughter VP, Paynter J. Social maturity and theory of mind in typically developing children and those on the autism spectrum. Journal of Child Psychology and Psychiatry. 2007; 48(12):1243–1250.10.1111/j.1469-7610.2007.01810.x [PubMed: 18093030]

Peterson CC, Wellman HM, Liu D. Steps in theory-of-mind development for children with deafness or autism. Child Dev. 2005; 76(2):502–517.10.1111/j.1467-8624.2005.00859.x [PubMed: 15784096]

Peterson CC, Wellman HM, Slaughter V. The mind behind the message: Advancing theory of mind scales for typically developing children, and those with deafness, autism, or Asperger Syndrome. Child Development. 2012; 83(2):469–485.10.1111/j.1467-8624.2011.01728.x [PubMed: 22304467]

Piaget, J. Piaget's theory. In: Mussen, PH., editor. Carmichael's manual of child psychology. New York: Wiley; 1970.

Pyers JE, Senghas A. Language promotes false-belief understanding: Evidence from learners of a new sign language. Psychological Science. 2009; 20(7):805–812.10.1111/j.1467-9280.2009.02377.x [PubMed: 19515119]

Ruffman T, Slade L, Crowe E. The relation between children's and mothers' mental state language and theory-of-mind understanding. Child Development. 2002; 73(3):734–751.10.1111/1467-8624.00435 [PubMed: 12038548]

Schick B, De Villiers P, De Villiers J, Hoffmeister R. Language and theory of mind: A study of deaf children. Child Development. 2007; 78(2):376–396.10.1111/j.1467-8624.2007.01004.x [PubMed: 17381779]

Shahaeian A, Peterson CC, Slaughter V, Wellman HM. Culture and the sequence of steps in theory of mind development. Developmental Psychology. 2011; 47(5):1239–1247.10.1037/a0023899 [PubMed: 21639620]

Siegal M, Varley R. Neural systems involved in 'theory of mind'. Nature Reviews Neuroscience. 2002; 3(6):463–471.10.1038/nrn844

Slaughter V, Dennis MJ, Pritchard M. Theory of mind and peer acceptance in preschool children. British Journal of Developmental Psychology. 2002; 20(4):545–564.10.1348/026151002760390945

Slaughter V, Gopnik A. Conceptual coherence in the child's theory of mind: Training children to understand belief. Child Development. 1996; 67(6):2967–2988.10.1111/j.1467-8624.1996.tb01898.x [PubMed: 9071768]

Swettenham JG, Baron-Cohen S, Gomez JC, Walsh S. What's inside someone's head? Conceiving of the mind as a camera helps children with autism acquire an alternative to a theory of mind. Cognitive Neuropsychiatry. 1996; 1(1):73–88.10.1080/135468096396712 [PubMed: 16571475]

Tabachnick, B.; Fidell, L. Using multivariate statistics. New York: Harper & Row; 1989.

Vaccari C, Marschark M. Communication between parents and deaf children. Journal of Child Psychology and Psychiatry. 1997; 38:793–801. [PubMed: 9363578]

Watson AC, Nixon CL, Wilson A, Capage L. Social interaction skills and theory of mind in young children. Developmental Psychology. 1999; 35:386–391.10.1037/0012-1649.35.2.386 [PubMed: 10082009]

Wellman HM, Baron-Cohen S, Caswell R, Gomez JC, Swettenham J, Toye E, Lagattuta K. Thought-bubbles help children with autism acquire an alternative to a theory of mind. Autism. 2002; 6(4): 343–363.10.1177/1362361302006004003 [PubMed: 12540127]

Wellman HM, Cross D, Watson J. A meta-analysis of theory-of-mind development: The truth about false belief. Child Dev. 2001; 72(3):655–684.10.1111/1467-24.00304 [PubMed: 11405571]

Wellman HM, Lopez-Duran S, LaBounty J, Hamilton B. Infant attention to intentional action predicts preschool theory of mind. Developmental Psychology. 2008; 44(2):618–623.10.1037/0012-1649.44.2.618 [PubMed: 18331149]

Wellman HM, Fang F, Liu D, Zhu L, Liu G. Scaling of theory-of-mind understandings in Chinese children. Psychological Science. 2006; 17(12):1075–1081.10.1111/j.1467-9280.2006.01830.x [PubMed: 17201790]

Wellman HM, Fang F, Peterson CC. Sequential progressions in a theory of mind scale: Longitudinal perspectives. Child Development. 2011; 82(3):780–792.10.1111/j.1467-8624.2011.01583.x [PubMed: 21428982]

Wellman HM, Hollander M, Schult CA. Young children's understanding of thought bubbles and of thoughts. Child Development. 1996; 67(3):768–788.10.1111/j.1467-8624.1996.tb01763.x [PubMed: 8706525]

Wellman HM, Liu D. Scaling of theory-of-mind tasks. Child Development. 2004; 75(2):523–541.10.1111/j.1467-8624.2004.00691.x [PubMed: 15056204]

Wiig, E.; Secord, W.; Semel, E. Clinical Evaluation of Language Fundamentals (CELF)—Preschool. New York: Psychological Corporation; 1992.

Woodward AL. Infants selectively encode the goal object of an actor's reach. Cognition. 1998; 69(1): 1–34. http://dx.doi.org/10.1016/S0010-0277(98)00058-4. [PubMed: 9871370]

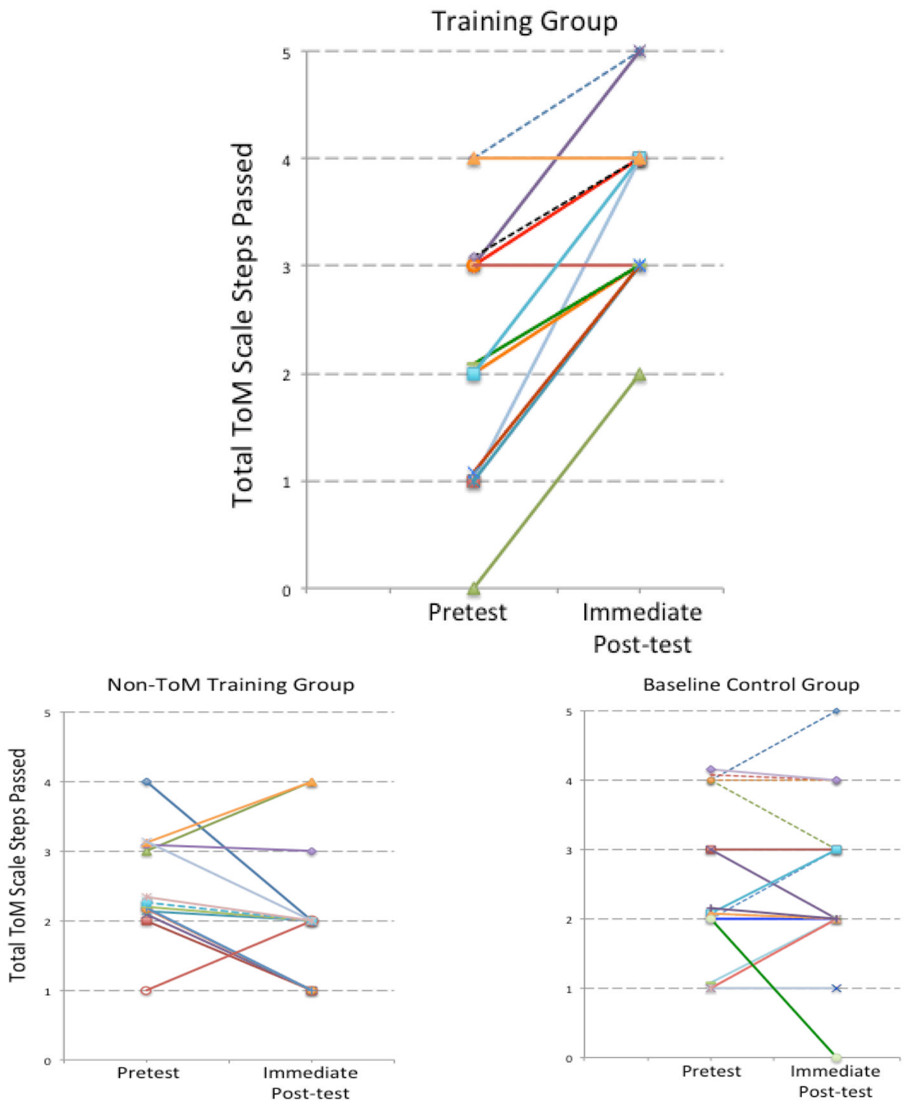

Figure 1.

Individual children's progress on the ToM Scale. In each panel, different colored lines represent different individual children. Solid lines are those children (N=11 in the ToM training group, N = 12 in the baseline control group, N = 13 in the non-ToM training group) in the reduced sample. The dotted lines are the additional children in the full sample (i.e., those who passed false belief at pretest).

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Background characteristics of the three groups

	ToM training group (n = 13)	Baseline control group (n = 16)	Non-ToM training group (n = 14)
Mean age (years) at pretest	9.82 (1.50)	9.80 (2.08)	8.66 (2.20)
Mean language ability score (0–22)	18.31 (1.84)	17.23 (2.56: n = 13)	16.57 (2.06)
Mean interval from pretest to immediate posttest (weeks)	12.23 (1.36)	13.50 (2.34)	11.07 (2.24)
Boy/girl ratio	7/6	10/6	7/7
Percent with cochlear implant	46%	50%	57%

(Standard deviations are in parentheses)

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Performance at pretest and posttest

Pretest	ToM-training group (n = 13)	Baseline control group (n = 16)	Non-ToM-training group (n = 14)
Mean pretest TFB (0–3)	1.15 (1.07)[1]	1.19 (1.28)	.71 (.99)
Mean pretest changed location false belief (0–2)	.92 (.86)	.88 (.88)	.57 (.85)
Percent passing pretest container false belief	23%	31%	14%
Mean pretest ToM Scale total (0–5)	2.23 (1.29)	2.56 (1.15)	2.36 (.74)
Mean pretest “far generalization”[2](0–4)	2.00 (1.08)	2.25 (.78)	2.21 (.80)
Posttest			
Mean posttest TFB (0–3)	2.08 (1.04)	1.19 (1.17)	.57 (.94)
Mean posttest changed location false belief (0–2)	1.46 (.78)	.88 (.88)	.43 (.65)
Percent passing posttest container false belief	62%	31%	14%
Mean posttest ToM Scale total (0–5)	3.62 (.87)	2.63 (1.26)	2.07 (1.00)
Mean posttest “far generalization” (0–4)	3.00 (.58)	2.31 (1.01)	1.93 (.73)

[1]Standard deviations are in parentheses

[2]“Far generalization” = ToM Scale without FB (i.e., DD+DB+KA+HE)