

## Explaining the Unpredictable: The Development of Causal Theories of Mind in Deaf and Hearing Children

Candida C. Peterson  
*University of Queensland*

Henry M. Wellman  
*University of Michigan*

Two studies of 100 children aged 3–12 years examined theory of mind (ToM) understanding via explanations and predictions in hearing preschoolers and ToM-delayed deaf children. Study 1's 75 children (31 deaf; 44 hearing) displayed an "explanation advantage," devising valid epistemic ToM explanations despite failing simpler forced-choice false-belief (FB) prediction tests. This novel discovery for deaf children extended to unexpectedly frequent cognitive ("think" or "know") explanations. Study 2 (with 25 additional deaf children;  $M_{\text{age}} = 9$ ) showed that microgenetic FB explanation practice resulted in significant gains on FB prediction posttests that were absent in a non-ToM control group. Implications for (a) explanation's interconnection with conceptual development, (b) designing ToM interventions, and (c) teaching deaf and hearing children are discussed.

Explanation is fundamental to cognition and a natural part of everyday social interaction (Keil, 2006). When A asks "Why?" and B supplies an explanation, both parties are able to evaluate and learn about causal concepts. Why questions are evident in children's spontaneous speech from as early as age 2 years (Hickling & Wellman, 2001), and their focus, like that of children's explanatory answers, is most frequently on people's actions, desires, and beliefs (e.g., Callanan & Oakes, 1992; Hickling & Wellman, 2001). Children's own spontaneous explanations likewise draw heavily on themes of human behavior, emotion, and belief (e.g., Dunn & Brown, 1993). Schult and Wellman (1997) found young preschoolers explained people's mistaken actions (A wanted to do X but did Y) via psychological states over 85% of the time. Thus, children's causal questions and explanations reflect their developing theories of mind. Moreover, in both spontaneous conversation and in controlled laboratory tests, children's explanations are often surprisingly adept (Wellman, 2011). Hearing preschoolers often do better at explanation than prediction consistent with an "explanation advantage" hypothesis (Amsterlaw & Wellman, 2006; Rhodes & Wellman, 2013). ToM or the child's understanding of the mental world (Flavell, 2004) develops rapidly during childhood (Harris, 2006; Hughes, 2016; Wellman, 2014). One

key aspect, the recognition that human behavior is the product of mental states like beliefs, is prototypically assessed using inferential false-belief (FB) tasks (Wellman, Cross, & Watson, 2001).

There can be limitations to using tasks with a specific conceptual focus (e.g., FB) as a marker for a larger multifaceted conceptual domain (e.g., ToM), as has been noted by several critics (e.g., Astington, 2001; Wellman & Liu, 2004). For some purposes, however, a focus on FB remains useful because such tasks (a) have been used and validated worldwide with thousands of children (see Wellman et al., 2001, for a meta-analysis), (b) are consistently reliable in multiple variations (e.g., yielding the same results with dolls vs. real people as protagonists: Wellman et al., 2001), and (c) have proved to reveal developmental change in both typically developing (TD) children and those with ToM delays owing to autism or deafness (see Happé, 1995; Peterson, 2009, for reviews). Moreover, variation in FB competence reliably predicts various social-interactive competencies such as peer interactions, reciprocated friendship and popularity (see Slaughter, Imuta, Peterson, & Henry, 2015 for a meta-analysis), children's engagement in pretending and deception games (Lalonde & Chandler, 1995; Peskin & Ardino, 2003), or persuasion (Bartsch & London, 2000; Slaughter, Peterson, & Moore, 2013), as well as academic achievement and relationships

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

with teachers during the transition to school (Lecce, Caputi, & Hughes, 2011).

FB tasks, the prime measure of children's ToM in past research, usually require predictions of the actions or thoughts of naive protagonists with beliefs that clash with reality. Most TD children under age 5 fail by making reality-driven predictions that ignore the protagonist's state of mind. Another less common FB task, however, involves false-belief explanation (FBE). As in prediction tasks, a naive protagonist lacks information known to the child. Then, the protagonist acts misguidedly (e.g., searches for an object where she last saw it rather than where it is now), and the child has to explain why. Strikingly, in direct comparisons, young children who do not yet fully understand FBs often perform better on explanation than prediction versions of the test. Children's superior performance on FBE tasks is provocative because, on the surface, prediction tasks might seem to be easier. Responses simply require choosing between two presented options (e.g., former vs. current hiding places), whereas explanation tasks require mounting a more self-initiated and extended causal argument. Thus, prediction tasks have more the cognitive quality of recognition where explanation tasks demand production and recognition is often easier than production.

Yet, an explanation advantage has been documented and replicated in several ToM studies using standard false-belief prediction (FBP) and explanation tasks with hearing preschoolers (e.g., Amsterlaw & Wellman, 2006; Bartsch, Campbell, & Troseth, 2007; Bartsch & Wellman, 1989; Rhodes & Wellman, 2013). Importantly, these standard explanation tasks only require understanding of ToM-driven *actions* (e.g., Why did she *go there* for a bandaid?) in keeping with a standard FB task's focus on predicting action ("Where will she *go* for a bandaid?"). As such, they do not require more sophisticated metacognitive understanding of the causes for *cognition* (e.g., "Why does Ann *think* the bandaids are there?"). Wimmer and Mayringer (1998) found no explanation advantage using metacognitive tests requiring children to explain an actor's reasons for thinking what she thought, as opposed to an actor's reasons for acting as she acted. Difficulty with reasoning about causes for beliefs is consistent with other research showing that tests of higher order metacognitive "thinking-about-thinking" skills often lead to responses that are unclear and confused in young children, even including those who are already at ceiling on both prediction and explanation versions of standard FB

tests that carefully ask only about overt behavior (see Flavell, 2004; Miller, 2009 for reviews).

What could account for an explanation advantage on standard FB tests? There are roughly two classes of proposal. First are interactional-familiarity accounts. Because TD preschoolers from age 2 on frequently engage spontaneously in explanatory conversation, asking and answering why questions about human behavior many times a day (Callanan & Oakes, 1992; Hickling & Wellman, 2001), this interest and practice might promote surprising facility on ToM explanation tasks. Relatedly, educators (e.g., Chi, DeLeeuw, Chiu, & LaVancher, 1994) have demonstrated the power of students' explaining why an author or teacher has claimed something for the mathematical and scientific learning of children and adults, supporting a general claim that practice at explanation aids cognitive growth. Sociocognitive conflict paradigms involving peers exchanging explanations for Piagetian conservation problems likewise show cognitive benefits from the act of explaining. Collaborators' gains on posttests are typically greater than for control groups exposed to individual training (e.g., Perret-Clermont, 1980), and this is true even for dyads whose explanations during collaboration are consistently incorrect (Azmitia, 1996).

Other accounts propose in varied ways that explanation (including private "self-explanation" as well as the social exchange of explanations discussed earlier) is a fundamental human cognitive activity, a core part of the young child's attempts to make sense of the world. For example, from a theory-theory perspective, children are forming naive theories of the world and a key function of theories is to frame and promote causal-explanatory understanding (e.g., Gopnik & Wellman, 1994, 2012). A special cognitive status for explanatory understanding is part of many other perspectives within cognitive science, as reviewed by Lombrozo (2006).

More generally, such data and varied theoretical claims have helped fuel increased research into the nature, role, and development of children's explanatory processes (e.g., Frazier, Gelman, & Wellman, 2009, 2016; Keil, 2006; Legare, 2012; Legare, Gelman, & Wellman, 2010; Legare, Wellman, & Gelman, 2009; Rhodes & Wellman, 2013; Walker, Lombrozo, Legare, & Gopnik, 2014; Wellman, 2011). However, this burgeoning research has all been conducted with TD children, essentially those from advantaged backgrounds whose parents engage them in conversation, including frequent causal-explanatory exchanges. Better and broader understanding of children's putative explanatory

prowess would be achieved by examining children who may not have “normative” interactional familiarity with explanatory conversation, whose daily upbringing encompasses vastly reduced requests and opportunities for explanation. For ToM specifically, comparisons of children with differing daily levels of exposure to conversationally discussing and explaining mental states would be especially informative.

Children born severely or profoundly deaf into hearing families are likely to be one such group. In the current research, we focus on their explanatory competencies and, in particular, on whether an explanation advantage for ToM is present for them as for their hearing peers. In general, during early childhood, conversational access for many deaf children of hearing parents (DoH children) depends on their hearing family members’ levels of signing skills. Yet, with some notable exceptions, these are often poor so that DoH children gain only limited early exposure to family conversations, especially causal talk surrounding mental states (Kritzer, 2008). “Most hearing parents who do sign are limited to relatively concrete conversation with little or no capacity for extended discussions in sign of social phenomena, or intangibles like thoughts and feelings” (Vaccari & Marschark, 1997, p. 797). Yet, if at least one parent is a deaf native signer, such mentalistic family conversation can proceed normally, though in sign rather than speech: that is, in a language that is fluently accessible to both conversational parties (Moeller & Schick, 2006).

A review by Marschark, Spencer, Adams, and Sapere (2011) suggested that hearing adults more often take control of conversations when a child is deaf and, so as to avoid confusion, rarely mention their own doubts, mistakes, misunderstandings or faulty ideas. Similarly, for orally communicating deaf children (including those with successful cochlear implants), easy access to spoken conversation is often not achieved for years after acquiring the implant (Schorr, Roth, & Fox, 2008) and often remains a significant challenge especially when conversing with other children or in group situations (e.g., Preisler, Tvingstedt, & Ahlström, 2005). Consequently, even DoH children receiving their cochlear implants in toddlerhood are unlikely to have had as much access as hearing preschoolers to the informal causal-explanatory conversations about people that in hearing-only households are linked with early FB success (e.g., Dunn, Brown, Slomowski, Tesla, & Youngblade, 1991; Harris, 2006; Peterson & Slaugther, 2003).

Thus, studying these children’s ToM explanations and ToM predictions presents important opportunities for further insight into how these two processes develop and interconnect with one another. Possibly, DoH children may be poor at explanation in general, and at ToM explanation in particular, given their likely reduced conversational exposure to formal (e.g., with parents) and informal (e.g., with peers and siblings) causal-explanatory talk, especially about unobservable mental states. Thus, for them, the ToM explanation–prediction relation might be very different. In this vein, past studies have raised serious questions about deaf children’s causal-explanatory thinking and willingness to use question–answer conversations to gain explanatory information even outside the ToM domain. Pioneering studies asked DoH children to produce explanations for rules of school comportment (Rachford & Furth, 1986) or a novel game (Hoemann, 1972) and found substantially poorer performance than by younger hearing children. Further, Brice (1985) exposed 10-year-old DoH children to concrete demonstrations of puzzling phenomena and found that they asked fewer why questions and were significantly more tolerant of cognitive ambiguity than hearing 6-year-olds. Unlike the latter, they almost never offered their own explanations. Arnold, Palmer, and Lloyd (1999) used referential communication tests and found that, compared with hearing peers, deaf 8-year-olds rarely or never requested necessary explanations or queried blatant ambiguities and omissions. More recently, Calderon and Greenberg (2003) likewise found that DoH children often misunderstood or ignored why questions in conversation, thus concluding: “The impact of limited explanations and restricted experiences denies to many deaf children their rightful opportunity to learn to understand others” (p. 179).

Such findings suggest DoH children are relatively *unfamiliar* with the conversational exchange of “why” questions and answers, at least about ToM concepts. This could mean, according to interactional-familiarity accounts, that DoH children would not display an explanation advantage. Perhaps, they might find both types of ToM test equally difficult. Still more likely, the simple pointing responses required by FBP tasks could make them easier than FBE tasks that require generating extended spontaneous conversational responses to “why” questions. Yet, there are also reasons to predict that an explanation advantage similar to hearing preschoolers’ could arise for DoH children as well. As noted earlier, explanation may be

especially fundamental to human reasoning. If so, this could be true for deaf and hearing children alike. Children ponder the causes of things in private reflection (explaining to oneself), not just socially through the conversation or more formal instruction. In contrast to the social exchange of well-formulated “why” questions and answers, self-explanations are often merely “implicit and skeletal” (Keil, 2006) and hence potentially less dependent on conversational (or social-interactional) familiarity. Thus, complimenting theory-theory’s explanatory emphasis, Keil and Wilson (2000) summarized some of these arguments and data by saying that “explanations seem to be a large and natural part of our cognitive lives” (p. 1).

Beyond these arguments and evidence, which to reiterate, have exclusively considered hearing children and adults, there is tantalizing data from one prior study as to deaf children’s possibly accurate explanations for mental states. Peterson and Slaughter (2006) asked a group of late-signing DoH children (5–12 years) to tell stories (in sign) in response to socially complex, wordless pictures. Like hearing preschoolers, many deaf children who failed standard FBP tasks not only described the pictured protagonists using simple cognitive words (e.g., “think”) but also offered causally coherent accounts of mental states in the form of “elaborations” (e.g., “Her brother is pretending that he is a ghost”) or even “explanations” (e.g., “She is scared because she thinks the ghost is real”). Because these DoH children mostly failed FBP tasks, these findings support the hypothesis (albeit indirectly) that DoH children could also evidence an explanation advantage similar to hearing preschoolers.

#### *The Current Research*

In two studies, we compare FBE with FBP in DoH children. Because we are the first to do so, our study is novel and exploratory. Past evidence is too sparse and indirect to permit forecasting of precise directional hypotheses. Instead, any of the three possible outcomes (explanation advantage, prediction advantage, or no difference between the two) all appear plausible in advance of empirical evidence. Importantly, whichever of the three possible outcomes emerges, it will not only be informative with regard to deaf children but will also have theoretical relevance for a broader understanding of how concepts about FBs and their causes develop in all children irrespective of hearing status.

In this regard, it is important for theories of ToM to discover how explanations and predictions about

people interconnect over development in children generally, not merely hearing ones. Further, owing to their slower ToM development than for hearing children or deaf native signers (Harris, 2006), studies of DoH groups can arguably supply unique insight into the developmental course of ToM—in this case predictive and explanatory ToM competence. Comparisons between deaf and hearing groups can suggest testable hypotheses about the triggers for developmental change (e.g., early conversational experiences at home and at school) and are relevant to broad theories of the nature and development of ToM (e.g., experientially driven conceptual change vs. modular brain maturation: Slaughter & De Rosnay, 2017; Wellman, 2014).

### **Study 1**

Study 1 aimed to use well-matched standard tasks measuring FBP and explanation to test whether DoH children’s ToM reasoning is characterized by an explanation advantage or, alternatively, whether they find explanation comparatively difficult. Possibly, it might be as difficult as prediction, or even more so. Relatedly, we compare DoH children’s FBE and prediction skills with those of younger hearing preschoolers of matched (and imperfect) ToM understanding. Making such comparisons (between explanation and prediction and between hearing and DoH children) requires not only good tasks but also convincing ways to code children’s explanatory answers. Hence, we developed and used both a “standard” coding system (parallel to that used in prior research with TD children) and a more stringently conservative coding system for explanations that required explicit mention of cognition and, thus, was designed to reveal a convincing depth of DoH (and hearing) children’s explanatory competence.

#### *Method*

##### *Experimental Rationale*

We compared deaf elementary school children (DoH) with TD hearing preschoolers. Although ages are different, these groups are well matched for our focal questions. In order to see if X (explaining FB actions) precedes or is easier than Y (predicting FB actions) we need to test children in both deaf and hearing groups who have not yet reached ceiling on X or Y. Children who always pass both sorts of tests, or always fail them both, provide



little or no information for this comparison. This dictated our design. Past research shows that (a) hearing children have typically reached ceiling on FBP by about age 5 (Wellman et al., 2001) and (b) that many DoH children aged 6–11 years, although rarely at ceiling, will have at least begun to develop some skill at FBP (see Peterson, 2009, for a review).

### Participants

Seventy-five Australian children aged 3–12 years participated in two groups. Group 1 had 31 DoH children ( $M_{\text{age}} = 9.34$  years; range = 5.83–12.42; 18 boys) from specialist bilingual hearing-impairment units using Auslan (Australian Sign Language) and spoken English as instruction media. Having grown up in hearing homes that, like most, are “neither fluent nor proficient” in sign (Goldin-Meadow & Mayberry, 2001, p. 224), these children were “late signers” whose daily immersion in a natively fluent signing community (of peers and teachers) had typically coincided with their entry into formal schooling between age 4 and 5. Nevertheless, by the time we tested them, all Group 1 children had signing skills that were, according to teachers, at least “adequate for everyday communication.” They had either severe (71+ dB) or profound (91+ dB) prelingual hearing losses and 17 (55%) had a cochlear implant. Precise information on age at implantation was unavailable to us. However, school records confirmed that all had received the implant prior to school entry (i.e., before about age 4.5 years). All the DoH children in our sample were bilingual in sign and speech to the extent that they (a) used both modalities to some degree on a daily basis, and (b) all preferred bilingual testing to either sign or speech alone, as noted in the following.

Group 2 had 44 TD hearing preschoolers ( $M_{\text{age}} = 4.13$ ; range = 3.08–5.00; 25 boys). Gender balance did not differ between groups,  $\chi^2(1) < 1$ ,  $N = 75$ ,  $p = .914$ , but mean age did,  $t(73) = 17.33$ ,  $p < .001$ , as a necessary consequence of our experimental design that, as noted earlier, required children who were unlikely to score at ceiling on at least one of the two sorts of tasks. No Group 2 child was bilingual, according to teachers.

All in both groups had written parental informed consent and English as their family’s sole or primary language. Although precluded by the terms of our institutional ethical approval from seeking details of parents’ incomes or educational attainments, teachers reported no serious economic hardship in this sample. All schools were

government funded (rather than private fee paying) and drew from neighborhoods of predominantly middle socioeconomic status, each scoring above the 50th percentile (in the advantaged direction) on the Australian government’s “socioeconomic indexes for areas” (Australian Bureau of Statistics, 2011), a composite reflecting the neighborhood’s median levels of parental education, income, rentals (vs. home ownership), unemployment, and proportion of unskilled occupations.

### Tasks and Scoring

*False-belief prediction.* Three standard FBP tests were given. One was a misleading container task from Wellman and Liu (2004): After guessing the expected contents of a closed box (e.g., crayons or eggs), children were shown that it contained something unexpected (e.g., a toy car). A naive boy doll then saw the closed box, followed by a test question “What does he think is in the box?” and a control “Did he see inside the box?” A pass required both correct. Two tasks were standard changed-location items from Baron-Cohen, Leslie, and Frith (1985): A girl put her ball inside a covered basket and departed. A boy moved it to a box (Task 1) or the experimenter’s pocket (Task 2). The girl returned and the test question “Where will she look for her ball?” was followed by two control questions, “Where is the ball now?” and “Where did the girl put the ball in the beginning?” All had to be correct to pass. Summed, the total FBP (TFBP) score ranged from 0 to 3.

*False-belief explanation.* In a separate session, children received three FBE tasks. All were modeled closely on tasks used successfully in prior research with hearing preschoolers (e.g., Amsterlaw & Wellman, 2006; Bartsch et al., 2007), and all were likewise similar to the prediction tests (described earlier) except for the nature of the test question. For the misleading container explanation task, after guessing the contents of a closed pictorially labeled bandaid box, children saw it was empty whereas a plain blue box held bandaids. A naive girl doll with a bleeding knee wanted a bandaid and marched purposefully toward the labeled bandaid box. Children were asked “Why is she looking there?” along with a control question, “Where are the bandaids really?” The two changed-location explanation items involved pictorial stories about protagonists who became distracted (by a book, TV, or phone) after seeing two items at specific locations on a table. The protagonist, thus, missed a crucial transposition of the items’ positions (which the child

witnessed) and acted upon a FB (e.g., drank from a flower vase rather than a glass). Children's spontaneous signed or spoken replies to the ensuing test question (e.g., "Why is she drinking the flowers?") were recorded in full. If any child (rare in this sample) said "Don't know" or nothing, there was one repeat (e.g., "She's drinking the flower water, why?") but no further prompts. Thus, children's own spontaneous explanations were elicited without any modeling or corrective feedback from the experimenter.

*Scoring of explanations.* Table 1 shows a sampling of explanations taken from both groups' transcripts. We coded these using two different scoring schemes. First, we applied a more lenient, "broadly epistemic" (BE) accuracy criterion that has been used effectively in many past studies of hearing preschoolers. Then, to gain a more cautiously conservative picture of children's explanatory capacities (especially the DoH children's), we recoded their explanations with a novel "strictly cognitive" (SC) scheme. Both schemes are explained and illustrated in Table 1. Although our SC scheme was more stringent than previous studies' through its requirement that cognition be mentioned explicitly, we reasoned that the most conservatively compelling evidence for ToM arises when explanations accurately infer and label protagonists' mental states of ignorance or FB using specifically cognitive verbs ("think," "know," or synonyms: see Table 1). However, in order not to underestimate children or neglect other informative aspects of their reasoning, we separately applied the BE scoring of past research. Specifically, Amsterlaw and Wellman (2006) had included mistakes (unintentional action) along with cognition in their most sophisticated "Belief/Mistake" (p. 154) category, and Wimmer and Mayringer (1998) included "epistemic" references to perception. Thus, our lenient BE scoring allowed intention, perception, and/or cognition terms to qualify for a pass, but not desire terms because Wimmer and Mayringer (1998) argued these (e.g., "she wants a bandaid") merely explain going *somewhere* but not going to the *wrong* location. Similarly, following Amsterlaw and Wellman (2006), we did not count "situational" justifications as correct (e.g., references to the current or past physical positions of objects) although Wimmer and Mayringer had done so. Simply remembering that an object has moved does not guarantee that the child understands what is in the protagonist's mind. Invalid reasons (e.g., "because she thinks flowers are yummy") were also incorrect even if they included cognitive verbs.

Using these criteria, children either passed (1) or failed (0) each FBE task with no extra credit given for using multiple correct terms. Total scores (summed over the three tasks) could range from 0 to 3 for BE and (separately) from 0 to 3 for SC. The first author coded all transcripts. Then, to establish interrater reliability, a second coder (blind to children's ages, hearing status and the study's hypotheses) used just the rules and examples in Table 1 to independently code a randomly chosen 86 responses representing both groups. Agreement between coders was almost perfect, 98%. The few disagreements were resolved by discussion.

*Language ability test.* To assess general language skill, we used the 22-item syntax ("Sentence Structure") subscale of the CELF (Clinical Evaluation of Language Fundamentals, 1st ed.—preschool [CELF/P]: Wiig, Secord, & Semel, 1992). This test has been used effectively in prior ToM research with TD children (e.g., Ruffman, Slade, & Crowe, 2002) and with deaf children in the same age range as the present sample (e.g., Paatsch & Toe, 2014; Peterson, Wellman, & Slaughter, 2012). It assesses a broad range of developmentally sequenced lexical, morphological, and syntactic concepts via picture-pointing responses and is uniquely suitable for validly assessing linguistic maturity among Auslan users (see Wellman & Peterson, 2013). Because no Auslan norms were available, we used raw total scores (out of the full 22 items) as the language measure for both groups (these are also better for comparing groups differing in age). All DoH children had complete language data. Six hearing preschoolers had missing data owing to school absence or other scheduling difficulties.

### Testing Procedure

Children were tested individually at school by a male experimenter assisted, for all deaf children, by a professionally qualified sign-language interpreter who translated the main experimenter's speech into Auslan and the child's Auslan into spoken words that the experimenter recorded. Interpreters were professionally accredited at the top interpreter level (Level 3) of Auslan signing skill by the peak national accreditation body in Australia, National Accreditation Authority for Translators and Interpreters (NAATI, 2011), and were already familiar to the deaf children from prior school interactions. The interpreter sat beside the main experimenter (both facing the child), translating the experimenter's speech into Auslan and repeating the child's Auslan orally for the experimenter to record.

Table 1

Study 1's Scoring of Explanations Along With Examples From the Children's Transcripts

Explanation type	Description	Examples from transcripts	Strictly cognitive score	Broadly epistemic score
Strictly cognitive	<i>Cognitive</i> verb(s) used to relevantly and causally explain the false-belief-based action	"She <i>assumed</i> it had bandaids because of the picture"; "She <i>thought</i> it was her drink"; "She <i>forgot</i> where she put her cup"; "She did not <i>realize</i> it was the flowers"; "She was not <i>thinking</i> and grabbed the closest one";	"Correct"	"Correct"
Noncognitive epistemic	<i>Perception</i> or <i>intention</i> term(s) used to plausibly and relevantly explain the target action	"She <i>saw</i> her glass there before"; "She <i>sees</i> the bandaid picture"; "She <i>reads</i> the label"; "She's on the phone with her <i>eyes shut</i> "; "She <i>meant to</i> take the other one"; "She grabbed the <i>wrong</i> one"; "She made a <i>mistake</i> "	"Incorrect"	"Correct"
Empty or incomplete (noncredited)	Explanations that fail to clearly identify a necessary causal precondition for the mistaken action, including <i>desire</i> (D), <i>situational</i> (S), <i>noncausal</i> (NC), <i>irrelevant</i> (I), and <i>empty</i> (E) explanations	"She wants chocolate sauce" (D); "She's thirsty" (D); "She needs a bandaid" (D); "Her glass was there before" (S); "Someone swapped the bottles" (S); "She opens the bandaids and <i>sees</i> it's empty" (NC); "She's a bad girl" (I); "Flowers will die without water" (I); "Don't know" (E)	"Incorrect"	"Incorrect"

This bilingual mode of task presentation, a familiar part of everyday school routines, was preferable to sign alone because—despite all being signers—many children in the sample also used some oral communication on a regular basis. Hearing preschoolers took the tasks individually in the oral mode only. Data collection for Study 1 took place between March 2009 and February 2014.

### Results

Table 2 shows children's scores on key measures. There was no significant difference between the groups in language skill,  $t(67) = 0.96$ ,  $p = .340$ . That DoH children averaged roughly 75% items correct (a clear pass) helps confirm reports by teachers that these DoH children had "sound" general language skills. Both groups likewise performed equivalently (and poorly) on TFBP. Only 16% of deaf children and 25% of hearing preschoolers (who were chosen to be younger than the normative FB mastery age: Wellman et al., 2001) passed TFBP by getting at least 2 of the 3 prediction items correct, a group difference that was not statistically significant,  $\chi^2(1) < 1$ ,  $N = 75$ ,  $p = .356$ . Thus, as planned, there was room in each group for an explanation advantage to

manifest itself. Prior to computing statistical analyses, we checked score distributions for normality. Shapiro–Wilk tests showed that TFBP, SC, and BE scores were not normally distributed. Nonparametric tests were, therefore, employed for all analyses. To compare groups, we used Mann–Whitney  $U$  tests. Monte Carlo confidence intervals (CI) around the  $p$  values are also reported. For total SC explanations, both groups scored equivalently, Mann–Whitney  $U = 536.50$ ,  $z = 1.67$ ,  $p = .096$ , 95% CI [.091, .102]. However, 14 deaf children (45%) gave an SC explanation on at least two of three problems compared with only 23% of the TD group, a significant difference:  $\chi^2(1) = 4.21$ ,  $N = 75$ ,  $p = .040$ . Similarly, DoH children outperformed hearing preschoolers on total BE explanations, Mann–Whitney  $U = 453.50$ ,  $z = 2.57$ ,  $p = .010$ , 95% CI [.007, .011], and 23 deaf children (74%) but only 21 TD preschoolers (48%) gave a BE response to at least two of the three explanation problems,  $\chi^2(1) = 5.25$ ,  $N = 75$ ,  $p = .022$ . At the same time, a large majority of deaf (87%) and of hearing (64%) children managed to explain at least one FB scenario using a correct BE reason, and this was a cognitive (SC) reason for 58% of the deaf and 43% of the hearing children. Thus, both groups displayed skill with ToM explanation.

Comparisons between school-aged deaf children and hearing preschoolers are, of course, only suggestive because, despite equivalent ToM prediction scores, the two groups differed in many age-related ways. Therefore, we tested for any “explanation advantage” (explanation superior to prediction) by analyzing Groups 1 and 2 separately. First, we compared total BE explanation scores with total prediction (TFBP) scores. For the DoH children, the BE explanation total was significantly higher than TFBP total, Wilcoxon  $z = 4.21$ ,  $p < .001$ , 95% CI [ $< .001$ ,  $< .001$ ], and the same was true for hearing preschoolers, Wilcoxon  $z = 2.99$ ,  $p = .003$ , 95% CI [ $.001$ ,  $.002$ ]. Thus, both groups demonstrated an explanation advantage in FB understanding using this traditional BE scoring of past research. Next, we made the same comparisons using our stricter cognitive (SC) scoring. For deaf children, the difference favoring SC explanation over prediction was significant, Wilcoxon  $z = 2.74$ ,  $p = .006$ , 95% CI [ $.004$ ,  $.007$ ]. But for hearing preschoolers, SC scoring yielded only a nonsignificant difference, Wilcoxon  $z < 1$ ,  $p = .738$ . Spearman correlations revealed that language ability was significantly linked with all three ToM measures for deaf children, all  $\rho$ s  $> \text{or} = .51$ , all  $p$ s  $< \text{or} = .003$ . For them, TFBP was also significantly correlated with age ( $p = .034$ ). For the hearing preschoolers, SC and BE were significantly correlated with language ability:  $\rho = .39$ ,  $p = .017$  and  $\rho = .46$ ,  $p = .004$ , respectively, but TFBP was not ( $p = .253$ ).

Table 2  
Mean Scores on Key Study 1 Variables for Children in Each Group

Variable	Group 1: deaf ( $n = 31$ )	Group 2: hearing ( $n = 44$ )
Age (in years)		
$M$	9.34	4.13
$SD$	1.92	0.46
Range	5.83–12.42	3.08–5.00
Total false-belief prediction		
$M$	0.68	0.80
$SD$	1.04	0.98
Range	0–3	0–3
Total broadly epistemic explanation		
$M$	2.13	1.39
$SD$	1.09	1.24
Range	0–3	0–3
Total strictly cognitive explanation		
$M$	1.26	0.75
$SD$	1.24	0.81
Range	0–3	0–2
Language ability (out of 22)		
$M$	17.68	17.03
$SD$	2.55	2.99
Range	15–21	10–22

## Discussion

Overall, a clear explanation advantage was demonstrated for signing deaf children aged 5–12 years, including when using a narrowly conservative scoring scheme that demanded explicit causal mention of cognition. Hearing preschoolers replicated past studies (e.g., Amsterlaw & Wellman, 2006) in showing a clear explanation advantage when using traditional (BE) scoring that, like earlier studies, credited accurate intention- and perception-based explanations together with cognitive ones. But the advantage was not statistically significant using the stricter SC scoring. These results reinforce the view that the process of explaining is fundamental to early cognitive development (Keil, 2006), especially in the naive psychological domain (Wellman, 2011). More specifically, they not only continue to confirm that hearing preschoolers are better at explaining the consequences of others' FBs than at predicting them but also, for the first time, highlight how relatively proficient DoH children are at using ToM concepts to explain human behavior. For both deaf and hearing children alike, generating valid explanations for FB-driven behavior was significantly easier than making simple behavioral predictions about it and these school-age deaf children stood out also for often including cognitive verbs like “think” in their explanations.

Although differing baselines of chance success complicate precise comparisons between ToM explanations and predictions, we addressed this issue in two ways. First, we directly compared scores on closely matched explanation and prediction tasks, regardless of baseline possibilities for spurious chance success. Ignoring chance baselines arguably works *against* finding better (or even equal) performance on ToM explanation because children could (and did) provide many possible kinds of explanations for the events presented in the stories, whereas the binary choice offered by a standard prediction task (e.g., predicting search of the basket vs. the box) assures 50% odds of chance accuracy. This partly explains why the notion of an explanation advantage can initially seem counterintuitive. Second, beyond the types of coding used in prior research, we assessed children's explanations via our more stringent cognition-only scoring scheme. Notably, DoH children's explanations outstripped their predictions even using this stringently conservative scoring.

It is worth considering whether (a) our prediction tasks might somehow have been spuriously difficult or (b) our explanation tasks might not have



required genuine ToM understanding. Both possibilities seem unlikely. Our explanation tasks were taken from past studies of hearing children and were carefully modeled on standard changed-location and misleading-container prediction tests. Stories and overall procedures for both types of task were essentially the same apart from the test question (e.g., “Why does she look here?” vs. “Where will she look?”). Furthermore, when children supply SC explanations like “Because she thinks the bandaids are in the box with the bandaids picture” it is hard to argue that they fail to understand FB.

By the same token, our prediction tasks were exactly those used in prior research with deaf and hearing children (see Peterson, 2009; Wellman Fang, & Peterson, 2011, for reviews), and these deaf children scored similarly to their counterparts in past studies (e.g., ours were 23% correct as against 17% for Italian DoH children in Meristo et al., 2007 and 32% [at mean age 7] for Schick, deVilliers, deVilliers, & Hoffmeister’s, 2007 U.S. sample). Thus, it is hard to argue that our ToM prediction measures were somehow unusually or spuriously difficult compared with past research.

In this way, our deaf sample’s poor FBP performance was no surprise. What was notable was how comparatively well these same children did on FBE. This is a novel and important finding. Of course, it is also important not to overstate DoH children’s competence at ToM explanation or relatedly at ToM understanding. Even using the more lenient BE scoring of past research only 52% scored perfectly. More focally, even when deaf children gave sensible FBEs, these did not translate into equally correct FBPs. Though averaging 5 years older, the DoH children performed slightly worse than the TD preschoolers ( $M_{TFBP} = .68$  vs.  $.80$ ) on the latter, although not significantly so. Conversely, even our youngest DoH children (aged 5 and 6 years) showed some explanatory competence (e.g., 67% of those aged 7.00 years and under gave at least one correct BE explanation). Nevertheless, further investigation of DoH children’s explanatory competence in future research is clearly desirable, including testing younger DoH children (provided they have sufficient language to comprehend the tasks).

As noted earlier, the difference in chronological age between our deaf and hearing groups was inevitable, given our research design and questions (i.e., to avoid uniform ceiling or floor effects in either group). Conceivably, this age difference may have contributed to the deaf-hearing difference in frequency of SC explanations in some unknown way. This warrants exploration in future research.

Meanwhile, however, it is important to note that our other (BE) scoring scheme is the one used by all past research on this topic and that using this standard scoring deaf and hearing children did not differ. Both groups displayed a clearcut, statistically significant FBE advantage using BE scoring. Furthermore, the advantage was of equivalent magnitude in both groups despite their age difference. SC scores have the methodological strength of being cautiously conservative, but they are unique to our study. The evidence of the deaf children’s explanation advantage with this additional more stringent measure is important, but further research using SC scores with other deaf and hearing groups is obviously needed before definitive conclusions or wider implications can validly be drawn.

## Study 2

A crucial theoretical question concerning explanation’s advantages over prediction for deaf and hearing children alike concerns whether and how explanation assists further learning. Conceptually, being required to explain a now-apparent but unexpected fact or occurrence often makes the limited nature of one’s understanding blatantly obvious (Keil, 2006). Empirically, mounting evidence from studies of TD hearing children in numerous paradigms suggests that generating causal explanations may supply a mechanism for causal learning (Keil, 2006; Lombrozo, 2006). When asked repeatedly to explain events, TD children’s learning often exceeds that of nonintervention control groups. The quality of subsequent explanations, predictions, and/or information retention have all been seen to improve over time for both preschoolers (Amsterlaw & Wellman, 2006; Frazier et al., 2016) and school-age children (Chi et al., 1994). These effects do not require any feedback as to the adequacy of one’s explanation; explanatory effort is sufficient.

But, of course, things could be different for deaf children. If so (or if not), this would be theoretically interesting. Returning to the theoretical alternatives outlined in our introduction, but now focusing specifically on learning, it is interesting to speculate on possible mechanisms that could contribute, within the ToM domain, both to children’s learning how to explain and to their learning from explaining. Perhaps, explanatory attempts operate at a fundamental level to facilitate conceptual development. They might, as Keil argued, work by emphasizing one’s unexpected ignorance. Equally, or alternatively, because explanations for misguided actions

are initially more sophisticated than corresponding predictions (as demonstrated for both deaf and hearing children in Study 1), the act of explaining may well motivate children to muster their most advanced ToM reasoning. This could conceivably provide a helpful platform for further insightful learning, ultimately assisting prediction as well. Indeed, for hearing children, being asked for ToM explanations in extended microgenetic sessions has been shown to boost subsequent performance on FBP tests (Amsterlaw & Wellman, 2006; Rhodes & Wellman, 2013).

An initial study, such as ours, cannot directly or deeply adjudicate between such theoretical alternatives. However, it could take a needed step by establishing whether DoH children improve after being required to engage in explanatory effort in the context of ToM problems. Because deaf children are arguably exposed less often to explanatory conversational exchanges in general (see Marschark et al., 2011; Wood & Wood, 1997 as outlined in the introduction), then if DoH children's conceptual gains from being asked to explain people's puzzling actions turn out to resemble hearing children's, this could prove suggestive. So too would the alternative possibility—namely that extended practice in generating explanations fails to benefit ToM understanding for DoH children, contrary to some past evidence for hearing children. If such practice leads to subsequent gains for the deaf, this could generate further hypotheses. For example, the impact of explanation may operate via more fundamental cognitive mechanisms such as a universal human drive for ever greater understanding (Keil, 2006). To examine this, we designed an intervention study exposing DoH children intensively to the challenge of explaining acts of paradoxical ToM-driven behavior.

One prior study, in particular, sets the stage for our research. Using a microgenetic training approach, Amsterlaw and Wellman (2006) gave hearing children ( $N = 12$ ) practice at generating explanations for a protagonist's FB-based behavior over multiple tasks spread over 6 weeks. Half the problems involved unseen displacements and half, misleading containers. On every trial, a full FBP test was initially given in standard format with a prediction test question (Where will Sally look for her ball?). Then, the test story continued with the protagonist acting on the basis of a FB, providing implicit feedback about whether the child's prediction was correct. A FBE test question followed (e.g., "Why is she looking in the drawer?"). If no explanation was spontaneously forthcoming, the protagonist's cognition was queried (e.g., "What does

Marcia think?"). The 12 children in the focal microgenetic training group, who explained 12 or 24 problems in their multiple sessions over 6 weeks, evidenced significant improvements on standard FBP tests at posttest relative not only to pretest but also to two control groups, one with microgenetic sessions of FBP and another with no intervention. As well as improving on FBP, the focal training group displayed generalization to another type of ToM task (knowledge access [KA]), although not to an appearance-reality test assessing awareness that an object can look like one thing (e.g., an apple) but really be something else (e.g., a candle). These findings (also see Rhodes & Wellman, 2013) provide evidence that giving hearing preschoolers extended practice with explaining and predicting can result in higher scores on standard FBP posttests.

We followed a modified version of this template in Study 2 to explore FBE training in a sample of 25 school-aged DoH children. Our main change was to omit the FBP training that each of these past studies had included along with FBE practice on every training trial. Thus, as well as focusing on deaf children, our training procedure was novel in testing whether FBE practice can be given *alone*, versus whether it must be *combined* with FBP practice, in order to result in posttest gains. This in itself was an important methodological refinement that could help to narrow down the focal conditions responsible for past effects of FBE training. We also had a non-ToM control group. It was an "active nonintervention" or "placebo" control (Boot, Simons, Stothart, & Stutts, 2013) that, similar to the training group, engaged in interesting individualized activities with an experimenter and an Auslan interpreter, albeit with no ToM content. This type of control has the advantage over a pure nonintervention control (where nothing except time intervenes between pretests and posttests) of ruling out the hypothesis that merely growing older and/or participating in interesting activities could fully explain any posttest gains made by the focal training group.

We predicted that children in our focal ToM-training group (who were exposed, like Rhodes and Wellman's (2013) and Amsterlaw and Wellman's (2006) hearing children to 12 microgenetic sessions with a story protagonist acting out FB-driven behaviors followed by requests to explain her actions) would display higher posttest FBP scores than at pretest. Moreover, we predicted they would outperform the appropriately matched children in the non-ToM control group who, we predicted, would make little or no ToM progress.

Method

Experimental Design

The procedure involved four phases: a pretest, an intervention, an immediate posttest, and a delayed posttest. At pretest, children received a battery of four standard FBP tests. Eligible children were then selected as those who consistently failed these at pretest. We defined this stringently via the requirement that the child must have failed at least 75% (3 of 4) of pretest FB-prediction items. This ensured that all children (both intervention and controls) had room to improve prior to the intervention.

For the intervention phase, the ToM-training group was given explanation practice with 12 FBE scenarios (modeled closely on Amsterlaw & Wellman, 2006 study: see Table 3), presented in three blocks of four problems per session. Each session was separated from the next by 3–7 days. Thus, training extended over a 3-week period, followed

by the immediate posttest and, roughly 12 weeks later, a delayed posttest to address whether ToM gains, if any, persisted over time.

The control group took the same pre- and posttests as the ToM-training group and the gap between these was the same (see details in the following). Rather than ToM explanation, however, they engaged in a 45-min artistic (visual representation) problem-solving activity that had no ToM content. It was closely modeled on the procedures of Wellman and Peterson (2013).

Participants

The full sample of 25 DoH children in Study 2 (17 boys) had a mean age of 8.96 years (range = 6.25–13.08). None had taken part in Study 1. Just as in Study 1, children were preselected as having prelingual hearing losses that were either severe (71–90 dB) or profound (91 dB or greater)

Table 3  
Examples of Study 2 Explanation Training Tasks With Test Questions and Sample Responses

Explanation training tasks	Test question	Sample responses from transcripts
1. A kitten chases ball of wool from under the bed to under a chest of drawers. A boy who saw the ball under the bed was out of the room when the kitten moved it ( <i>pictured, changed location</i> )	“Why is he looking under the bed?”	“He thinks his cat is there” (C) “He did not see the kitten hiding the ball” (BE) “He wants the ball” (W) “Because he can’t find his cat” (W)
2. A gift is hidden under the big blue bed by A and, in A’s absence, it is moved by B who hides it under the small pink bed ( <i>props, changed location</i> )	“Why is A going to the big blue bed?”	“He was not watching her when she put it there” (BE) “He wants to wrap the present” (W)
3. Boy pours a glass of desired drink (juice) and puts it just behind him on desk; Mum silently moves the juice far side of the table and puts an undesired drink (water) where the juice was ( <i>pictured, changed location</i> )	“Why is boy drinking the water?”	“He doesn’t know Mum moved the water there” (C) “Because that’s not orange juice” (W) “Because Mum swapped the drinks around” (W)
4. Puzzle pieces are moved from their commercial box (pictorially labeled) to a plain box. Then, a naive doll arrives with a partly assembled puzzle wanting a missing piece ( <i>props, misleading container</i> )	“Why does she look here [pointing at pictorially labeled] box for her dinosaur puzzle?”	“She sees the dinosaur [label] so she thinks it’s in there” (C) “She does not know the puzzle is in the other box” (C) “She loves doing puzzles” (W)
5. Doll with a bleeding knee marches toward a pictorially labeled bandaid box. (Only the child knows it is empty and that bandaids are in blue box across the desk; <i>props, misleading container</i> )	“Why is she going to this box [pointing]?”	“Cos it looks like bandaids should be in that kind of box” (BE) “She needs a bandaid” (W) “She is bleeding” (W)

Note. BE = broadly epistemic; SC = strictly cognitive; C = correct via SC (and BE) scoring; BE = correct only via BE scoring; W = wrong (zero score).

and were recruited from two specialist units for hearing impairment located within government-funded primary schools where a bilingual (sign-plus-speech) mode of communication was employed. All the children had hearing parents, and, despite access to early intervention and efforts by some parents to learn some signs, their teachers reported that no child in this sample had any family member at home who used any form of signing as fluently as a native speaker. Further eligibility requirements included: (a) freedom from disabilities apart from hearing loss, (b) English as parents' sole or primary language, (c) written parental informed consent, and (d) at least one full year's attendance at the bilingual school coupled with Auslan skills rated at least "adequate for everyday communication" by teachers.

We allocated children into two groups—11 to the focal ToM-training group and 14 to the non-ToM control group—on a semi-random basis with the constraints that (a) the two participating schools were equally represented in training and control groups, and (b) younger and older children were distributed across groups as equally as possible. Subsequent statistical comparisons confirmed that these precautions were successful and that groups were well matched in several respects: There were no group differences for (a) school attended,  $\chi^2(1) < 1.00$ ,  $N = 25$ ,  $p = .622$ , (b) mean age ( $M = 9.75$  and  $8.34$  years, respectively),  $t(23) = 1.96$ ,  $p = .062$ , (c) gender balance,  $\chi^2(1) < 1.00$ ,  $N = 25$ ,  $p = .653$ , (d) proportions with cochlear implant(s), 54% and 29%,  $\chi^2(1) = 1.73$ ,  $N = 25$ ,  $p = .188$ , or (e) mean language ability,  $t(23) < 1.00$ ,  $p = .887$ .

### General Procedures

Each child in the ToM-training group was individually tested on all tasks in a bilingual modality (sign + speech: see Study 1) in a quiet school area. For the control group's pretests and posttests, the same was true. But their "placebo" intervention involved not only periods of individual (bilingual) interaction (with the same experimenter and interpreter as for the training group) but also periods of working alongside 1–4 other children in a small group on an art activity. For all children, task presentation and interpretation procedures were exactly as in Study 1, with the interpreter translating experimenter's speech into Auslan for the child, and vice versa. All interpreters were fully professionally accredited Auslan interpreters (NAATI, 2011). Data collection for Study 2 took place between March 2014 and October 2014.

*ToM pretests.* Before their intervention began, all children in both groups were individually pretested on a battery of standard FBP tasks, two involving changed locations (both trials of the Baron-Cohen et al., 1985 Sally-Ann task) and two involving misleading containers (a toy car in a crayon box and a candy box containing pencils). Procedures, questions, and scoring matched Study 1, including the requirement for perfect accuracy on a task's control as well as test questions for a pass.

*Language ability pretest.* The Sentence Structure (Syntax) subscale of the CELF-P test (Wiig et al., 1992) was presented and scored just as described for Study 1.

### Intervention Phase and Procedures

*ToM intervention.* The ToM explanation intervention was modeled closely on Amsterlaw and Wellman (2006) with one major modification. Rather than including any FBP questions or corrective feedback, our training omitted both of these. Each training task began with a protagonist acting on the basis of a FB (see Table 3) followed by the FBE question ("Why . . .?"). Our rationale for excluding Amsterlaw and Wellman's (2006) use of a prediction test question on each trial prior to the focal explanation test question was twofold. First, the language for asking FBP questions (e.g., "What does James think is in the lunchbox?") could cue children to use verbs like "think" in their explanations. Second, with Amsterlaw and Wellman's method, the implicit corrective feedback given to children who fail the FBP test questions during training could add to (or even substitute for) any benefit directly attributable to the act of explaining. Thus (as noted earlier), our training more conservatively tested possible benefits from FBE practice alone, without concomitant FBP practice. Half our training tasks (Table 3) involved changed locations and half misleading containers. Within each half, dolls and props were used half the time and picture-book stories about real children for the other half. No differences were observed as a function of these variations in task format. Our sample, like Amsterlaw and Wellman's focal group, received a total of 12 FBE problems. For us, these were presented individually in three separate sessions of four problems each. Sessions were spaced 3–7 days apart across a total span of roughly 3 weeks.

As Table 3 shows, the ToM-training group's scenarios closely resembled standard FBP tasks in all respects except the wording of the test question (e.g., "Why will she look" rather than "Where will



she look"). Children's answers to test questions were recorded verbatim in full. If a control question (e.g., "Where are the bandaids really?") was answered incorrectly (rare in this sample), the story and questions were repeated once. No more than one repetition was ever necessary. Finally, if the child's explanation did not include a cognitive term, the supplementary question "What is he [she] thinking?" was asked, just as in Amsterlaw and Wellman (2006). The next task began immediately so that no corrective feedback or any other implicit or explicit instruction was ever given. Thus, strictly speaking, children were not "trained" either about FBs or about how best to explain them. Nor were they informed as to the accuracy of their own spontaneous explanations. Instead, they were merely given extended practice with "why" questions that encouraged them to consider the causal basis for protagonists' unexpected reality-discrepant behaviors, together with the subtle suggestion to consider the protagonist's mental state ("What does she think?") if they failed to spontaneously mention cognition in their explanation.

*Control intervention.* The control group's intervention occupied the same interval from pre- to posttests as the ToM-training group's ( $M_s = 26.17$  and  $25.45$  days, respectively:  $t < 1$ ,  $p = .711$ ). As noted earlier, it consisted of a visual-representation exercise offering general practice in following task instructions, deploying attention between teacher and interpreter and dealing with representational materials, albeit in a manner that was visual and pictorial rather than mentalistic. Briefly, a visual stimulus (e.g., a photo of elephants, a vase of real flowers) was shown and, as in ToM training, a short introductory narrative framed the exercise (e.g., "Look at this photo: What are these? Yes, elephants. Look: there is a big one, a small one, and [point] what size is this? Yes, middle sized. Now here is your paper. You draw three elephants. Do them so they fill the whole page. Use lots of color. Try to do yours just like the picture. Make one big, one small, and one middle sized"). After children began to work, the experimenter and interpreter visited each child individually in turn. Using the same bilingual mode as for ToM testing, individualized comments included specific suggestions and general encouragement while taking care never to mention thoughts, intentions, beliefs, or other mental states.

#### *Posttest Procedures and Scoring*

*Immediate posttests.* The immediate posttest had four standard FBP tests (two involving

changed locations and two involving misleading containers). These mirrored those used at pretest with different scenarios, pictures, and props. In addition, there were two appearance-reality FBP tests (Lohmann & Tomasello, 2003). There was also a KA task at immediate and delayed posttest (see Peterson, Wellman, and Liu's [2005] Appendix for exact wording) to assess generalization to a novel (untrained) ToM concept other than FB. All 11 children in the ToM-training group took the immediate posttest but two in the control group missed it owing to protracted school absence. (Both were available for the delayed posttest and so were retained for all analyses not involving the immediate posttest.)

*Delayed posttest.* For the delayed posttest (assessing retention over 3 months), we used new dolls and materials for the four core FBP items (i.e., two new Sally-Ann-type tasks and two misleading container tasks) plus KA.

*Total posttest scores.* The core measure of FB-prediction skill for each of the testing phases (pretest, immediate posttest, and delayed posttest) was a four-item total FBP (TFBP4) composite summing scores on the two misleading container and the two changed-location items. At immediate posttest, there was a six-item TFBP6 score summing TFBP4 with the two appearance-reality FB tasks requiring prediction of a naive other's belief about the true identity of a deceptive item. The single KA generalization item was scored pass (1) or fail (0).

#### *Results and Discussion*

Results are considered first in terms of changes in children's performance on standard FBP tests as a result of our focal ToM training (involving explanation but not prediction). Next, we examine possible differences between children in the training group versus the control group on the immediate and delayed posttests. Finally, we explore the specific explanations generated by children in the training group during their training sessions and whether these relate to increases on their posttest FBP scores relative to pretest.

#### *Changes in FBP Skills From Pretest to Posttests*

Figure 1 shows the primary measure, mean total FBP (TFBP4) scores at pretest, immediate posttest, and delayed posttest. As the figure illustrates, children in the ToM-training group gained significantly from the pretest to the immediate posttest in scores on standard misleading container and

changed-location FBP tasks. As in Study 1, non-parametric Wilcoxon tests were used (owing to the nonnormal distribution of TFBP4 scores), and we also report Monte Carlo confidence intervals around the  $p$  values for significant effects. For the ToM-training group, there was a significant gain from the pretest to the immediate posttest in TFBP4 scores (see Figure 1 and Table 4), Wilcoxon  $z = 2.63$ ,  $p = .007$ , 95% CI [.006, .009]. Furthermore, at delayed posttest, their TFBP4 scores were also significantly higher than at pretest, Wilcoxon  $z = 2.72$ ,  $p = .006$ , 95% CI [.004, .007], indicating maintenance of their significant gains 12 weeks later. By contrast, the non-ToM control group did not improve significantly on TFBP4 from pretest to either the immediate posttest, Wilcoxon  $z = .45$ ,  $p = .655$ , or the delayed posttest, Wilcoxon  $z = 1.55$ ,  $p = .121$ .

#### *Posttest Comparisons Between Training and Control Groups*

As Figure 1 illustrates, both groups scored equally on TFBP4 at pretest ( $p = .572$ ) but there were significant group differences on TFBP4 at the immediate posttest, Mann-Whitney  $U = 22.50$ ,  $z = 2.78$ ,  $p = .004$ , 95% CI [.003, .005] and at the delayed posttest, Mann-Whitney  $U = 33.50$ ,  $z = 2.49$ ,  $p = .013$ , 95% CI [.011, .015]. The ToM-training group significantly outperformed the control group at both times. On the single KA item testing for generalization to a ToM concept not used in training, there were no statistically significant group differences at immediate posttest or delayed posttest (see Table 4 for means). However, there was a trend for those in the ToM-training group to outperform the control group at immediate posttest,  $\chi^2(1) = 3.49$ ,  $N = 23$ ,  $p = .062$ .

#### *Explanations During Training*

Notably, each child in the ToM-training group gave a coherent explanation for each of the 12 training problems (there were no "don't knows"). This indicated consistent interest and involvement with the task and highlighted Study 1's findings that a drive to explain paradoxical human action is clearly evident in DoH children. Their consistent responsiveness indicated, furthermore, that all children could account plausibly (to their own satisfaction if not to ours) for the mistaken behavior in each of the training stories. We used the same strict scoring scheme as in Study 1 to tally both BE and SC explanations. As in Study 1, SC explanations had to

explicitly mention a cognitive verb (e.g., "think") in an appropriate (causally antecedent) context. Similarly, BE explanations had to appropriately invoke an epistemic state of perception, intention, or cognition (Tables 1 and 3 give examples). Desire (e.g., "want") explanations, although common, were not credited for reasons detailed in Study 1. The same was true of situational explanations (devoid of mental state terms) and noncausal mentions of an appropriate verb. Explanation performance of the training group during training partially replicated Study 1. On the first 4 training trials, correct BE explanations (normally distributed) were significantly more frequent than correct FBPs on the first four pretest items,  $t(10) = 2.80$ ,  $p = .019$ . Moreover, 9 of the 11 ToM-training group children (82%) used a correct BE explanation at least once during training, constituting 27% of total training explanations. Correct SC explanations were used by 6 of the 11 children (55%) at least once during training.

There were also wide individual differences within the ToM-training group in children's frequencies of using BE and SC explanations. To see if these were connected with training's effectiveness, we computed correlations with posttest FBP scores (we used nonparametric Spearman correlations for these analyses owing to the nonnormality of TFBP score distributions). Children's total SC explanations during training correlated significantly with their TFBP prediction scores on both the immediate posttest,  $\rho = .62$ ,  $p = .041$ , and the delayed posttest,  $\rho = .77$ ,  $p = .006$ . But there was no correlation of training SC scores with pretest TFBP, implicating explanation performance during the training phase, rather than preexisting FBP skill, in the posttest associations. Correlations for BE explanations during training trended in a similar direction, but were nonsignificant (all  $\rho$ s  $\leq .55$ , all  $p$ s  $\geq .079$ ) possibly due to the small sample size and the fact that all but two children gave at least one BE explanation during training, resulting in little individual variability.

#### *Study 2 Conclusions*

After repeated practice devising their own explanations for story protagonists' reality-discrepant actions during training, the children in our ToM-training group showed significant gains in FBP relative to pretest, and these were still apparent 3 months later. By contrast, the control group made little or no ToM progress, ruling out general factors like passage of time or opportunities for special activities as the factors accounting for progress by

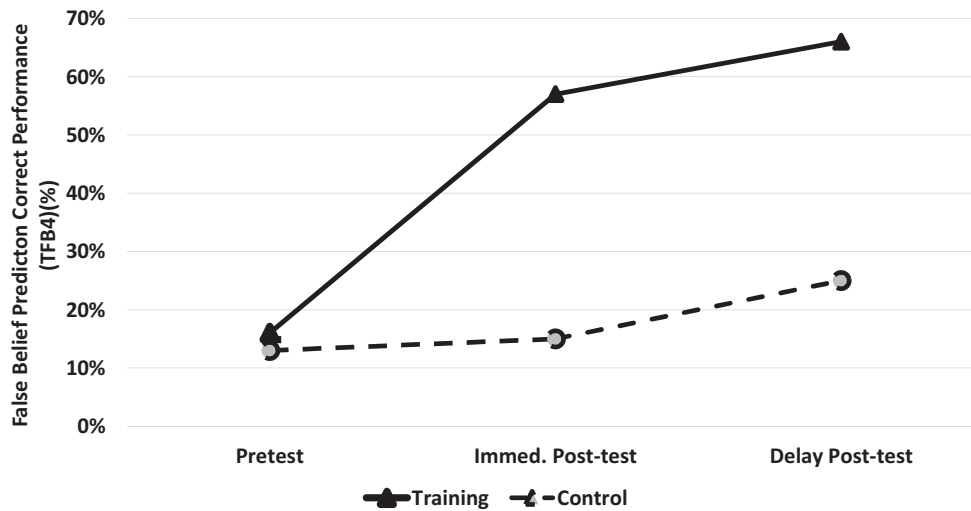


Figure 1. False-belief prediction (percent correct) by ToM-training and non-ToM-control groups at pretest, immediate posttest, and delayed posttest. ToM = theory of mind.

Table 4  
Performance on Study 2 Posttest Measures and Language Ability

	Group	N	M (SD)
Immediate posttest TFBP4 (max. 4)	Training	11	2.27 (1.42)
	Control	12	0.58 (0.793)
Immediate posttest KA (max. 1)	Training	11	0.64 (0.50)
	Control	12	0.25 (0.45)
Immediate posttest TFBP6 (max. 6)	Training	11	3.45 (1.92)
	Control	12	0.92 (1.23)
Delay posttest TFBP4 (max. 4)	Training	11	2.64 (1.47)
	Control	14	1.00 (0.84)
Delay posttest KA (max. 1)	Training	11	0.73 (0.47)
	Control	14	0.50 (0.52)
Language ability (max. 22)	Training	11	14.82 (3.00)
	Control	14	15.00 (3.26)

Note. TFBP = total false-belief prediction; KA = knowledge access.

the focal group. Our study is an advance in this regard over some prior ToM-training studies of hearing preschoolers that have often overlooked the need for any sort of active non-ToM control or indeed (sometimes) for any control condition at all. Without this, simply observing progress in the intervention group does not validly permit the conclusion that the intervention itself made any difference.

Arguably, our focus on DoH children for this demonstration is advantageous. It shows that experiences with ToM explanation can result in posttest gains even for a group whose ToM development would, without intervention, proceed slowly over a

very prolonged time period. This unusually protracted timetable for spontaneous ToM mastery by DoH children is instructive. In contrast to TD children, who ordinarily master FB so rapidly that spontaneous gains might arise without intervention between pretest and delayed posttest, this was much less likely for our DoH children (Peterson & Wellman, in press). Also, unlike some past studies of hearing preschoolers (e.g., Amsterlaw & Wellman, 2006), our focal training intervention did not include either practice with FBP or feedback on explanation or prediction accuracy. Nor did it include any exposure to the well-crafted model explanations created by adults that have been used in some past ToM-training studies of hearing children (e.g., Lecce, Bianco, Demicheli, & Cavallini, 2014). Instead, in our study, DoH children were simply faced with the challenge of explaining story protagonists' unexpected reality-discrepant actions. All in our sample rose to the challenge and managed to devise their own accounts of why the story protagonists were behaving so oddly, even if these were not always adequate by adult standards. That is, after being asked to explain, and finding a way to do so, even children whose explanations did not explicitly refer to epistemic mental states displayed improvement on FBP posttests. Our study has the limitation, however, of only having a single control group. Thus, although supportive of the inference that *something* about the ToM intervention we used was more beneficial than a matched control intervention devoid of ToM content, we cannot identify with certainty what this "something" was.

Our preferred hypothesis is that the key ingredient is DoH children's greatly enhanced practice with ToM explanation during our intervention. Yet because of the limitations of our control group, it is admittedly unclear whether improvement resulted from engaging in explanation or merely from being exposed to ToM materials and questions. Training research with TD children has shown that, for them, mere exposure to ToM materials is insufficient to produce training improvement (e.g., Amsterlaw & Wellman, 2006; Lohmann & Tomasello, 2003) suggesting the same could be true for deaf children, especially given their preexisting ToM delays. However, this is an empirical question that needs to be examined in future research. In particular, future research could examine this issue directly by including an additional control group that would be exposed to the same ToM stimulus materials but without any requests to generate explanations.

In broad terms, our findings have the general theoretical implications outlined at the start and supported by findings of Study 1. Because our participants were DoH children who arguably experience a paucity of explanation requests (especially ToM-relevant ones) in their homes and classrooms, and are known to be reluctant to ask why, our data argue against the hypothesis that training such as ours can only assist children with long and varied practice in seeking and receiving mental-state explanations during social interaction and conversation in everyday life from toddlerhood (Keil, 2006) onwards. Instead, our findings appear more consistent with the view that general cognitive processes recruited by explanation attempts may underpin the gains observed. Of course, further evaluation and support for such hypotheses depend on additional research, including future studies using multiple control groups.

### *General Discussion*

Results of both studies highlight the importance of FBE in relation to two broad interpretations outlined in our introduction. Study 1's demonstration of an FBE advantage over FBP for DoH children, just as for the hearing, is in line with suggestions that the drive to understand and explain represents a fundamental outcome of, and contributor to, cognitive development and learning across many different cultural, individual, and family situations (e.g., Keil, 2006; Lombrozo, 2006; Wellman, 2011). As Keil (2006) noted, "Explanations in their own right can be immensely rewarding things and may

be sought out as such, even by the youngest of children" (p. 234). Perhaps, this helps to explain superiority of deaf and hearing children's ToM explanations over their ToM predictions in Study 1, as well as the posttest ToM gains the DoH children displayed after completing our training intervention in Study 2.

Of course, with only cross-sectional and training data to go on, current research (our own and that of other past studies) cannot yet demonstrate that explanation is fundamental for children in the sense of being a naturally occurring chronological precursor to prediction. Future longitudinal studies would be useful to examine the spontaneous emergence of ToM-relevant explanations and predictions. Longitudinal research with DoH children would be an especially welcome addition to such a project. Given DoH children's presumed reduced exposure to the everyday conversational exchange of "why" questions and explanations, our data suggest that the search to explain perplexing human behavior may be cognitively fundamental for all children in the sense of a readily available and important platform for learning. Explanations have the potential to scaffold ToM growth and, as we noted earlier, this possibility is supported by Study 2's findings. Arguably the act of explaining coupled with an explanation advantage evokes some of a child's most advanced cognitions (in line with results of Study 1). Relatedly, Legare et al. (2009) demonstrated an explanation advantage (over prediction) for preschoolers' naive biological reasoning about contamination and infection. This too suggests that explanatory reasoning is arguably a cognitive fundamental. That deaf children, in particular, along with hearing children, become so easily engaged with the task of explanation accords with Keil's (2006) view that, "all of us throughout the world share the same drive for explanation" (p. 44).

If so, what sort of mechanisms might account not just for children's interest in explanations but their relative competence with explanation relative to prediction? Wellman and Liu (2007) have argued that much everyday explanation may be cognitively easier than prediction. Prediction often entails open-ended uncertainty among a great many potential outcomes. Even in cut-and-dried location-change FB tasks, the protagonist might search for his desired candy in location A, or B, or perhaps nowhere (e.g., if he is no longer hungry or assumes someone else has already eaten it) or even in both places at once (because he thinks equally desirable identical candies are actually in both locations), and



so on. But often explaining merely requires postdiction of known outcomes that can be accounted for after the fact. The protagonist is in fact searching for his candy in location X. That is a given. So all the child has to do is come up with a reason why. Hence, explanation very often (though not always) requires postdiction. Not just in children's thinking but also in scientific theorizing, predicting future outcomes in situations of uncertainty is often deemed to be a particularly challenging undertaking whereas post hoc postdiction is often easier.

Explanation's advantage for FB learning could additionally involve the mechanism described earlier: Explanations evoke some of a child's most advanced cognitions, and this, in turn, could scaffold ToM growth by making such reasoning more practiced and consciously available. Possibly, many factors are at play in past demonstrations of any explanation advantage for ToM tasks. For one thing, adult-child conversations could drive early propensities for explanatory sense making. When children ask and answer "why" questions their everyday cognition is often directed at understanding recent events, especially those involving human acts and motives. Also, the ease of postdiction might contribute to early, formative explanatory successes. Thus, satisfaction at having resolved an explanatory puzzle could motivate further efforts and successes, and these could ultimately benefit prediction as well.

Besides their theoretical value, data from both studies suggest that there could be practical value to be gained from eliciting and encouraging explanations from DoH children. In particular, our findings suggest that this could benefit ToM development specifically, although in theory, this could also apply more broadly to other domains of knowledge. Of course, with only 11 DoH children in Study 2's ToM-training group, and only a single control group, our results are clearly tentative and require replication and extension before such practical applications are attempted. Further research both in the ToM domain and more broadly could address this. Meanwhile, our data demonstrate the promising potential of an intervention as simple as being asked "Why?" as a possible stimulus to cognitive growth.

Indeed, could greater use of "why" questions in classrooms benefit deaf children's education? Our Study 2 data do not directly address educational benefits of explanatory efforts for DoH children. Yet, they do enhance the plausibility of testable hypotheses such as that the successful sharing of "why" questions and explanations through

conversation (especially with fluently signing peers or teachers) could be particularly helpful for these children. This is further suggested indirectly by evidence of DoH children's superior performance on cognitive tasks (Piagetian conservation and justice reasoning) after having explained their views to signing deaf peers in sociocognitive conflict paradigms (e.g., Peterson & Peterson, 1990). Children who debated with a signing deaf partner and attempted to explain their conflicting views made significant gains on individual conservation posttests irrespective of whether or not the dyad had achieved a correct solution while interacting. Thus, hypothetically, training interventions incorporating explanation could conceivably prove helpful for boosting deaf children's understanding not just of ToM but also in many other cognitive domains.

Other aspects of our findings also warrant continued research. For example, it would be useful to investigate explanatory ToM reasoning in still younger DoH children. Admittedly this would be difficult because their delayed language competence means younger DoH children would often lack the linguistic resources necessary for understanding and generating the kinds of extended verbal explanations that our sample produced. Unlike FBP, FBE questions cannot adequately be answered via simple monosyllables or finger pointing. However, it remains possible that merely being asked "why" questions can engage young (and deaf) children's drive for explanation even in the absence of the overt production of a satisfactory explanation (e.g., Azmitia, 1996; Peterson & Peterson, 1990) or perhaps any explanation at all. Relatedly, tasks where children can reveal their explanatory curiosity via exploratory manipulation of objects (e.g., Legare, Zhu, & Wellman, 2013; Walker et al., 2014) could be feasible for this group.

Whether the explanation advantage is especially pronounced for deaf and/or hearing children when explaining intentional human actions is intriguing. Explanations appear very early and broadly for TD children (e.g., Keil, 2006), and human behavior has been found to be a particularly compelling magnet for young children's interest and explanatory prowess (e.g., Callanan & Oakes, 1992; Hickling & Wellman, 2001). Deaf children too might be more motivated to try to explain socially relevant acts than purely academic material. However, without direct evidence from DoH children, this is only one of several possibilities that further research could helpfully address. Naturalistic studies of DoH children's spontaneous conversations with hearing-speaking family members and signing peers could

also prove revealing. It would be interesting to actually know how often “why” questions and answers, especially about people’s mental states, are spontaneously exchanged and with whom. Further, even if causal-explanatory discourse turns out to be infrequent, DoH children might engage in it sometimes and in revealing ways. A possible first step would be to have caregivers compile diary records of deaf children’s spontaneous “why” questions and answers, as in Callanan and Oakes’ (1992) study of hearing children.

Meanwhile, the current studies make several important contributions. First, they contribute to burgeoning literature suggesting that the simple act of explaining can promote cognitive development in children generally. Lombrozo’s (2006) proposal that “Explaining novel information to oneself can facilitate learning . . . and foster generalization” (p. 471) aligns with our Study 2 finding that our focal explanation intervention resulted in superior posttest performance over pretest. This arose not only on FBP tasks that used scenarios similar to training but also on completely new types of posttest FBP problems. Furthermore, the specific explanations that the children in the training group generated, while satisfying to themselves, did not always meet adult standards for a convincing ToM-based argument.

The exact mechanisms responsible are uncertain and clearly require further study. Nonetheless, our findings reinforce the value of pursuing explanation research as an avenue toward greater scientific understanding of the development of ToM and cognition generally. They also initiate new methods and new populations for the pursuit of these research directions. Third, our findings enhance the understanding of the processes of ToM development, both typically and amid delay. Fourth, they highlight the need for further investigation of a novel and straightforward intervention (asking why) that merits applied evaluation in practical settings as a possible means to assist not only DoH children’s overcoming of ToM delays but also the formal and informal education of children generally.

## References

- Amsterlaw, J., & Wellman, H. M. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition & Development, 7*, 139–172. [https://doi.org/10.1207/s15327647jcd0702\\_1](https://doi.org/10.1207/s15327647jcd0702_1)
- Arnold, P., Palmer, C., & Lloyd, J. (1999). Hearing-impaired children’s listening skills in a referential communication task: An exploratory study. *Deafness & Education International, 1*, 47–55. <https://doi.org/10.1179/146431599790561451>
- Astington, J. W. (2001). The future of theory-of-mind research: Understanding motivational states, the role of language, and real-world consequences. Commentary on “meta-analysis of theory-of-mind development: The truth about false belief.” *Child Development, 72*, 685–687. <https://doi.org/10.1111/1467-8624.00305>
- Australian Bureau of Statistics. (2011). *Census of Population and Housing Socioeconomic Indexes for Areas (SEIFA) 2033.0.55.001*. Canberra, ACT: Author. Retrieved from [www.abs.gov.au](http://www.abs.gov.au)
- Azmitia, M. (1996). Peer interactive minds. In P. Baltes & U. Staudinger (Eds.), *Interactive minds* (pp. 133–162). Cambridge, UK: Cambridge University Press.
- Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a “theory of mind?” *Cognition, 21*, 37–46. [https://doi.org/10.1016/0010-277\(85\)900228](https://doi.org/10.1016/0010-277(85)900228)
- Bartsch, K., Campbell, M. D., & Troseth, G. L. (2007). Why else does Jenny run? Young children’s extended psychological explanations. *Journal of Cognition & Development, 8*, 33–61. [https://doi.org/10.1207/s15327647jcd0801\\_2](https://doi.org/10.1207/s15327647jcd0801_2)
- Bartsch, K., & London, K. (2000). Children’s use of mental state information in selecting persuasive arguments. *Developmental Psychology, 36*, 352–365. <https://doi.org/10.1037/0012-1649.36.3.352>
- Bartsch, K., & Wellman, H. (1989). Young children’s attribution of action to beliefs and desires. *Child Development, 60*, 946–964. <https://doi.org/10.2307/1131035>
- Boot, W., Simons, D., Stothart, C., & Stutts, C. (2013). The pervasive problem of placebos in psychology. *Perspectives in Psychological Science, 7*, 582–584. <https://doi.org/10.1177/1745691613491271>
- Brice, P. (1985). A comparison of levels of tolerance for ambiguity in deaf and hearing children. *American Annals of the Deaf, 130*, 226–230.
- Calderon, R., & Greenberg, M. (2003). Social and emotional development of deaf children. In M. Marschark & P. Spencer (Eds.), *Oxford handbook of deaf studies, language and education* (pp. 177–187). Oxford, UK: Oxford University Press.
- Callanan, M., & Oakes, L. (1992). Preschoolers’ questions and parents’ explanations. *Cognitive Development, 7*, 213–233. [https://doi.org/10.1016/0885-2014\(92\)90012-G](https://doi.org/10.1016/0885-2014(92)90012-G)
- Chi, M., DeLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477. [https://doi.org/10.1207/s15516709cog1803\\_3](https://doi.org/10.1207/s15516709cog1803_3)
- Dunn, J., & Brown, J. R. (1993). Early conversations about causality: Content, pragmatics and developmental change. *British Journal of Developmental Psychology, 11*, 107–123. <https://doi.org/10.1111/j.2044-835X.1993.tb00591.x>

- Dunn, J., Brown, J., Slomsowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs. *Child Development*, 62, 1352–1366. <https://doi.org/10.2307/1130811>
- Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, 50, 274–290. <https://doi.org/10.1353/mpq.2004.0018>
- Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2009). Preschoolers' search for explanatory information within adult-child conversation. *Child Development*, 80, 1592–1611. <https://doi.org/10.1111/j.1467-8624.2009.01356.x>
- Frazier, B. N., Gelman, S. A., & Wellman, H. M. (2016). Young children prefer and remember satisfying explanations. *Journal of Cognition & Development*, 17, 718–736. <https://doi.org/10.1080/15248372.2015.1098649>
- Goldin-Meadow, S., & Mayberry, R. I. (2001). How do profoundly deaf children learn to read? *Learning Disabilities Research & Practice*, 16, 222–229. <https://doi.org/10.1111/0938-8982.00022>
- Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfeld & S. Gelman (Eds.), *Domain specificity in cognition and culture* (pp. 257–293). New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9780511752902>
- Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138, 1085–1108. <https://doi.org/10.1037/a0028044>
- Harris, P. L. (2006). Social cognition. In W. Damon (Ed.), *Handbook of child psychology* (pp. 811–857). New York, NY: Wiley.
- Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology*, 37, 668–683. <https://doi.org/10.1037/0012-1649.37.5.668>
- Hoemann, H. W. (1972). The development of communication skills in deaf and hearing children. *Child Development*, 43, 990–1003. <https://doi.org/10.2307/1127649>
- Hughes, C. (2016). Theory of mind grows up. *Journal of Experimental Child Psychology*, 149, 1–5. <https://doi.org/10.1016/j.jecp.2016.01.017>
- Keil, F. (2006). Explanation and understanding. *Annual Review of Psychology*, 57, 227–254. <https://doi.org/10.1146/annurev.psych.57.102904.190100>
- Keil, F., & Wilson, R. (2000). Explaining explanations. In F. Keil (Ed.), *Explanation and cognition* (pp. 1–18). Boston, MA: MIT Press.
- Kritzer, K. L. (2008). Family mediation of mathematically based concepts while engaged in a problem-solving activity with their young deaf children. *Journal of Deaf Studies and Deaf Education*, 13, 503–517. <https://doi.org/10.1093/deafed/enn007>
- Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. *Cognition and Emotion*, 9, 167–185. <https://doi.org/10.1080/02699939508409007>
- Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014). Training preschoolers on first-order false belief understanding: Transfer on advanced ToM skills and metamemory. *Child Development*, 85, 2404–2418. <https://doi.org/10.1111/cdev.12267>
- Lecce, S., Caputi, M., & Hughes, C. (2011). Does sensitivity to criticism mediate the relationship between theory of mind and academic achievement? *Journal of Experimental Child Psychology*, 110, 313–331. <https://doi.org/10.1016/j.jecp.2011.04.011>
- Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*, 83, 173–185. <https://doi.org/10.1111/j.1467-8624.2011.01691.x>
- Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development*, 81, 929–944. <https://doi.org/10.1111/j.1467-8624.2010.01443.x>
- Legare, C. H., Wellman, H. M., & Gelman, S. A. (2009). Evidence for an explanation advantage in naive biological reasoning. *Cognitive Psychology*, 58, 177–194. <https://doi.org/10.1016/j.cogpsych.2008.06.002>
- Legare, C. H., Zhu, L., & Wellman, H. M. (2013). Examining biological explanations in Chinese preschool children: A cross-cultural comparison. *Journal of Cognition and Culture*, 13, 67–93. <https://doi.org/10.1163/15685373-12342085>
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, 74, 1130–1144. <https://doi.org/10.1111/1467-8624.00597>
- Lombrozo, T. (2006). The structure and function of explanation. *Trends in Cognitive Sciences*, 10, 464–470. <https://doi.org/10.1016/j.tics.2006.08.004>
- Marschark, M., Spencer, P. E., Adams, J., & Sapere, P. (2011). Evidence-based practice in educating deaf and hard-of-hearing children: Teaching to their cognitive strengths and needs. *European Journal of Special Needs Education*, 26, 3–16. <https://doi.org/10.1080/08856257.2011.543540>
- Meristo, M., Falkman, K., Hjelmquist, E., Tedoldi, M., Surian, L., & Siegal, M. (2007). Language access and theory of mind reasoning: Evidence from deaf children in bilingual and oral environments. *Developmental Psychology*, 43, 1156–1169. <https://doi.org/10.1037/0012-1649.43.5.1156>
- Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, 135, 749–773. <https://doi.org/10.1037/a0016854>
- Moeller, M. P., & Schick, B. (2006). Relations between maternal input and theory of mind understanding in deaf children. *Child Development*, 73, 751–766. <https://doi.org/10.1111/j.1467-8624.20067703-0016>



- NAATI (National Accreditation Authority for Translators and Interpreters Ltd.). (2011). *About NAATI*. Retrieved from <http://www.naati.com.au>
- Paatsch, L., & Toe, D. (2014). A comparison of pragmatic language abilities of children who are deaf or hard of hearing and their hearing peers. *Journal of Deaf Studies and Deaf Education*, 19, 1–19. <https://doi.org/10.1093/deafed/ent03>
- Perret-Clermont, A.-N. (1980). *Social interaction and cognitive development in children*. London, UK: Academic Press.
- Peskin, J., & Ardino, V. (2003). Representing the mental world in children's social behavior: Playing hide-and-seek and keeping a secret. *Social Development*, 12, 496–512. <https://doi.org/10.1111/1467-9507.00245>
- Peterson, C. C. (2009). Development of social-cognitive and communication skills in children born deaf. *Scandinavian Journal of Psychology*, 50, 475–483. <https://doi.org/10.1111/j.1467-9450.2009.00750.x>
- Peterson, C. C., & Peterson, J. L. (1990). Socio-cognitive conflict and spatial perspective-taking in deaf children. *Journal of Applied Developmental Psychology*, 11, 267–281. [https://doi.org/10.1016/0193-3973\(90\)90010-H](https://doi.org/10.1016/0193-3973(90)90010-H)
- Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, 18, 399–429. [https://doi.org/10.1016/s0885-2014\(03\)00041-8](https://doi.org/10.1016/s0885-2014(03)00041-8)
- Peterson, C. C., & Slaughter, V. P. (2006). Telling the story of theory of mind: Deaf and hearing children's narratives and mental state understanding. *British Journal of Developmental Psychology*, 24, 151–179. <https://doi.org/10.1348/026151005X60022>
- Peterson, C. C., & Wellman, H. M. (in press). Longitudinal ToM development in children with and without ToM delay. *Child Development*. Advance online publication. <https://doi.org/10.1111/cdev.13064>
- Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory of mind development among children with autism, deafness or typical development. *Child Development*, 76, 502–517. <https://doi.org/10.1111/j.1467-8624.2005.00859.x>
- Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message. *Child Development*, 83, 469–485. <https://doi.org/10.1111/j.1467-8624.2011.01728x>
- Preisler, G., Tvingstedt, A.-L., & Ahlström, M. (2005). Interviews with deaf children about their experiences using cochlear implants. *American Annals of the Deaf*, 150, 260–267. <https://doi.org/10.1353/aad.2005.0034>
- Rachford, D., & Furth, H. (1986). Understanding friendship and social rules by deaf and hearing children. *Journal of Applied Developmental Psychology*, 7, 391–402. [https://doi.org/10.1016/0193-3973\(86\)90008-0](https://doi.org/10.1016/0193-3973(86)90008-0)
- Rhodes, M., & Wellman, H. M. (2013). Constructing a new theory from old ideas and new evidence. *Cognitive Science*, 37, 592–604. <https://doi.org/10.1111/cogs.12031>
- Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development*, 73, 734–751. <https://doi.org/10.1111/1467-8624.00435>
- Schick, B., deVilliers, P., deVilliers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development*, 78, 376–396. <https://doi.org/10.1111/j.148624.2007.01004>
- Schorr, E. A., Roth, F. P., & Fox, N. A. (2008). A comparison of the speech and language skills of children with cochlear implants and children with normal hearing. *Communication Disorders Quarterly*, 29, 195–210. <https://doi.org/10.1177/1525740108321217>
- Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: Children's understanding of the limits of psychological explanation. *Cognition*, 62, 291–324. [https://doi.org/10.1016/S0010-0277\(96\)00786-X](https://doi.org/10.1016/S0010-0277(96)00786-X)
- Slaughter, V., & De Rosnay, M. (2017). *Theory of mind in context*. Oxford, UK: Routledge.
- Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Development*, 86, 1159–1174. <https://doi.org/10.1111/cdev.12372>
- Slaughter, V., Peterson, C. C., & Moore, C. (2013). I can talk you into it: Theory of mind and persuasion behavior in young children. *Developmental Psychology*, 49, 227–231. <https://doi.org/10.1037/a0028280>
- Vaccari, C., & Marschark, M. (1997). Communication between parents and deaf children: Implications for social-emotional development. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 38, 793–801. <https://doi.org/10.1111/j.1469-7610.1997.tb01597.x>
- Walker, C. M., Lombrozo, T., Legare, C., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, 133, 343–357. <https://doi.org/10.1016/j.cognition.2014.07.008>
- Wellman, H. M. (2011). Reinvigorating explanations for the study of cognitive development. *Child Development Perspectives*, 5, 33–38. <https://doi.org/10.1111/j.1750-8606.2010.00154.x>
- Wellman, H. M. (2014). *Making minds: How theory of mind develops*. New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199334919.001.0001>
- 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. <https://doi.org/10.1111/1467-8624.00304>
- Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory of mind scale: Longitudinal perspectives. *Child Development*, 82, 780–792. <https://doi.org/10.1111/j.1467-8624.2011.01583.x>



- Wellman, H. M., & Liu, D. (2004). Scaling theory-of-mind tasks. *Child Development*, 75, 523–541. <https://doi.org/10.1111/j.1467-8624.2004.00691.x>
- Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik & L. E. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 261–279). New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195176803.001.0001>
- Wellman, H. M., & Peterson, C. C. (2013). Deafness, thought bubbles and ToM development. *Developmental Psychology*, 49, 2357–2367. <https://doi.org/10.1037/a0032419>
- Wiig, E. H., Secord, W. A., & Semel, E. (1992). *CELF-P: Clinical evaluation of language fundamentals–preschool*. San Antonio, TX: Psychological Corp., Harcourt Brace Jovanovich.
- Wimmer, H., & Mayringer, H. (1998). False belief understanding in young children: Explanations do not develop before predictions. *International Journal of Behavioral Development*, 22, 403–422. <https://doi.org/10.1080/016502598384441>
- Wood, D., & Wood, H. (1997). Communicating with children who are deaf: Pitfalls and possibilities. *Language, Speech, and Hearing Services in Schools*, 28, 348–354. <https://doi.org/10.1044/0161-1461.2804.348>