Available online at www.sciencedirect.com
ScienceDirect

Cognitive Development 22 (2007) 258–270

COGNITIVE DEVELOPMENT

A longitudinal study of child siblings and theory of mind development

Anna McAlister, Candida Peterson *

School of Psychology, University of Queensland, Brisbane, Queensland 4072, Australia

Abstract

This study tested a sample of 63 children twice in a longitudinal design over 14 months to examine their theory-of-mind (ToM) understanding in relation to their number of child-aged siblings (1–12 years). Age-appropriate batteries of ToM tests emphasising false belief were given at the start of the study, when children had a mean age of 4–2 (range: 3–3 to 5–6), and at the end, when mean age was 5–4 (range: 4–3 to 6–9). Irrespective of chronological age, children with 2 or more child siblings scored significantly higher on both the earlier and the later battery than those with no child-aged siblings. Hierarchical multiple regression analyses revealed that, over and above chronological age and verbal intelligence, having more child siblings predicted higher ToM scores at Times 1 and 2. Furthermore, at Time 2, the participant's number of child-aged siblings continued to predict higher ToM scores even after controlling for age, verbal intelligence, and Time 1 ToM scores. Results were considered in relation to the kinds of family-based social and conversational experiences that might foster ToM growth throughout the period from toddlerhood to the threshold of primary school.

Keywords: Theory of mind; Siblings; Longitudinal study; Preschoolers

When children develop a theory of mind (ToM) they become aware that human behavior is guided by mental states of belief, knowledge, memory and imagination that may conflict with overt reality. Empirically, the fundamental social understanding that is embodied in ToM is often assessed using inferential false belief tests requiring prediction or explanation of behavior by protagonists who have incorrect ideas that participants do not share. Wellman, Cross, and Watson (2001) conducted a meta-analysis of 178 studies of preschoolers' performance on these false belief tests. They reported a striking increase from below-chance success at age 3 to pervasive competency by age 6, indicative of "genuine conceptual change in the preschool years" (p. 655). Wide individual differences are also apparent during the preschool period, with some children

* Corresponding author. Tel.: +61 7 3365 6332; fax: +61 7 3365 4466.
E-mail address: candi@psy.uq.edu.au (C. Peterson).

0885-2014/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogdev.2006.10.009

mastering false belief many months, or even years, ahead of others. What might account for these differences?

In a search to explain variability in typically developing children's ages of ToM mastery, several cross-sectional studies have examined the sibling structure of preschoolers' households. Results have been mixed. Initially, Perner, Ruffman, and Leekam (1994) tested 76 British preschoolers and found a strong and linear pattern of association. Children with two or more siblings had almost twice the only-child's likelihood of passing false belief. Ruffman, Perner, Naito, Parkin, and Clements (1998) subsequently added a further 370 children to Perner et al.'s original sample. Higher false-belief scores were still linked with having more siblings, especially older ones. Yet there seemed to be a minimum age for the family size effect to be observed. Ruffman et al. found that 2-year olds and young 3-year olds failed to benefit from siblings, even when they had several all older than themselves.

With a Canadian sample, Jenkins and Astington (1996) confirmed the link between false belief scores and larger families while showing that it persisted even after children's language ability was statistically controlled. However language was relevant to the extent that young children with slow language development gained more than those with advanced language from the presence of siblings. Lewis, Freeman, Kryiakidou, Maridaki-Kassotaki, and Berridge (1996) studied Greek 3- and 4-year olds and, in addition to confirming false belief benefits of larger families found that only-children scored particularly low, prompting the view that: "having no siblings is particularly disadvantageous in the acquisition of false belief understanding" (p. 2937). Interestingly, when Lewis et al. compared families that were large because of having live-in kin (e.g., cousins, aunts, uncles, grandparents) versus those that were equally large owing to more siblings, they found each type of social variety in the family to be equally beneficial. Therefore they suggested that siblings were of no special benefit, opting instead for a "general apprenticeship" model in which ToM is enhanced by the number of family members of all ages that a child interacts with on a daily basis.

No significant links between and false belief scores and numbers of siblings (older, younger or both) were observed by Cutting and Dunn (1999) in a sample of 128 British children. The sample was highly diverse, with a large proportion of low income and single-parent families. This may have contributed to the null result. In fact, two other studies of economically disadvantaged families have likewise failed to reveal any significant links between sibling numbers and children's false belief understanding. In the first of these, Cole and Mitchell tested 119 British 3- to 5-year olds from a low income community in an initial study, and a further 71 children in a follow up. No correlation emerged in either sample between false belief scores and sibling numbers. Cole and Mitchell suggested a "potential contaminating influence of socio-economic status on the sibling advantage in theory of mind" (p. 279). Similarly, Hughes and Ensor (2005) failed to observe any association for lower-income British 2-year olds' between sibling numbers and scores on an age-appropriate battery of ToM tests. However, the affective quality of children's relationships with siblings did significantly predict ToM even after variables like age, verbal ability and quality of the parent–child relationship were statistically controlled. The authors suggested that 2-year olds' advanced ToM skills might be either the product of, or the impetus for, affectionate communication and play among siblings.

Indeed, number of siblings is generally seen not as an influence in and of itself, but rather as a proxy for some social interaction variable, even in studies revealing a sibling-ToM link. A sibling's presence in the family affords the opportunity for sharing distinctively childish forms of conversation and social interaction that may relate to rapid ToM growth. However these opportunities are likely to vary not only with the emotional tone of the sibling bond (Hughes & Ensor, 2005) but also with siblings' chronological ages (Peterson, 2000; Wright-Cassidy, Shaw-Fineberg, Brown, & Perkins, 2005). Teenagers and young adults may not interact much with their preschooler siblings. Even if they do, their social input may not differ enough from a parent's to be influential. Very young infants may not yet be capable of ToM-stimulating interaction, and twins may not possess mental perspectives that are divergent enough to challenge and stimulate their co-twin's mind-reading (Wright-Cassidy et al., 2005).

In fact, with a sample of 265 Australian 3- to 5-year olds, Peterson (2000) found that those with no siblings apart from a young infant, a teenager, or an adult, scored no higher than only-children on false belief tests. But preschoolers who had at least one child-aged sibling (1–12 years) at home scored significantly higher than only-children, suggesting that siblings can benefit ToM development, but only when all of an age to interact together in distinctively childish ways. More recently, McAlister and Peterson (2006), with a different sample of 124 Australian preschoolers, and Wright-Cassidy et al. (2005), with a sample of 72 U.S. preschoolers, replicated the finding that siblings had to be in the childhood age range (1–12 years) to be associated with advanced false belief understanding. Wright-Cassidy et al. (2005) also found that twins with at least one child sibling either older or younger than themselves did better than twins who had no siblings apart from the co-twin. In other words, siblings' ages may be as relevant as sibling numbers to advanced ToM, prompting the conclusion that: "the sibling effect is associated not with mere exposure to another mind but specifically to a mind or minds different from one's own" (Wright-Cassidy et al., 2005, p. 103).

In sum, while suggesting some promising leads, the existing evidence is still quite mixed not only over the question of which features of the sibling constellation may be linked with advanced ToM, but even over the more basic question of whether or not siblings do benefit preschoolers' ToM development at all. As noted above, some studies with large samples have failed to observe sibling effects. It is conceivable that null results could have arisen in some cases owing to unknowns such as the proportions siblings too young or old to interact (Peterson, 2000), or of twins (Wright-Cassidy et al., 2005), or of siblings with a hostile relationship (Hughes & Ensor, 2005) in particular studies' samples. Equally, it is conceivable that the sibling benefit for ToM may be applicable only to children of certain ages (Ruffman et al., 1998) or social backgrounds (Cole & Mitchell, 2000). It is clear that further investigation of the influence of numbers of child siblings on preschoolers' ToM development is needed, hence motivating the present study.

The cross-sectional nature of most previous sibling-ToM research is a further limitation upon understanding of the developmental age trajectory of possible sibling influences. Most studies have examined children from a narrow age range (e.g. 3–6 to 4–6) and have used false belief as the only index of ToM. Thus our goal was to longitudinally investigate the possible link between ToM and the preschooler's number of child-aged siblings using age-appropriate batteries of ToM understanding including, but not limited to, false belief. We wanted to examine the possibilities either (1) that siblings might exert a continuing benefit upon the growth of social understanding from age 3 to age 6 versus (2) might only be advantageous early in this period (as indirectly suggested by Jenkins and Astington's finding that children with immature language gained disproportionately from the presence of a sibling) versus only later on (in line with Ruffman et al.'s finding of a minimum age before which siblings were of no benefit).

There are theoretical arguments in favour of each of these contrasting models. In support of the possibility of longitudinal continuity of the sibling ToM advantage throughout the preschool years, children are known to spend large amounts of time with their siblings over this period (Dunn & Kendrick, 1982), continuously sharing activities and experiences that could foster social-cognitive development. For example, a child who has one or more child-aged siblings will differ from an only-child throughout childhood in having more opportunities: (1) to eavesdrop upon, and intrude into, conversations between their parents and their siblings (Dunn & Shatz, 1989), (2) to listen to mental-state discourse, via the family language community, master the meanings of mind-related words (Jenkins et al., 2003), (3) to observe parents reasoning with siblings (Ruffman, Perner, & Parkin, 1999), (4) to negotiate conflicts (Dunn, 1996), (5) to pretend (Brown, Dunn, & Donellan-McCall, 1996) and (6) to cooperate and compete in games of strategy (Cole & Mitchell, 2000). Each of these continuing family-based social experiences could plausibly benefit the child's understanding of other minds steadily from age 3 to age 6.

On the other hand, discontinuity is also plausible. The period from the end of toddlerhood to the start of primary school heralds an increase in friendships with peers outside the family (Slomkowski & Dunn, 1996), and increased concern with peer acceptance and popularity (Peterson & Siegal, 2002). During their longer hours spent in peers' company, older 5- and 6-year olds may increasingly replace sibling interaction at home with extra-familial pretend play, rule games, negotiation, deception and teasing. If so, only-children might no longer be disadvantaged in the richness of their encounters with other children's points of view, and thus the sibling-ToM link might diminish late in the preschool years. Alternatively, a longitudinal increase in the sibling-ToM association might emerge if children must reach some minimum chronological age (Ruffman et al., 1998) or minimum level of cognitive, linguistic or social maturity in order to be able to benefit from play and conversation with siblings. Our study will evaluate these possibilities.

1. Method

1.1. Participants

The sample consisted of 63 Australian children,[1] 35 boys and 28 girls, who had a mean age of 4 years 2 months (range 3–3 to 5–6) when they were initially tested at Time 1, and a mean age of 5 years 4 months (range 4–3 to 6–9) when the second battery of tests was administered an average of 14 months later, at Time 2. In identifying children for retest at Time 2, we imposed a number of selection criteria, the most important of which were: (a) that they had passed all control questions (see below) on standard false belief tests at Time 1 (so as to ensure their language skills were sufficient to assure valid test question responses), and (b) that their total complement of child-aged siblings (aged 1–12 years) had not altered between Times 1 and 2 (we imposed this criterion for the sake of clarity, consistency and homogeneity in our key variable of interest over the period of the study).

The children were recruited from preschools and child-care centres that served predominantly middle-class neighbourhoods. All had English as a first language and the vast majority were Caucasian. In terms of sibling constellation, there were 10 children in the sample (16%) with no child-aged siblings. Thirty-three children (52% of the sample) had just one child sibling, 17 (27% of the sample) had 2 siblings in the childhood age range and the remaining 3 children (5% of the sample) had exactly 3 child siblings. As well as living in economically advantaged communities, the parents of the children in this sample were relatively well-educated. Seventy percent of them had continued their studies beyond high school graduation and 45% of them had university educations.

[1] A portion of the Time 1 data for 55 of these 63 children was included as part of separate cross-sectional study of a larger sample of 124 children (McAlister and Peterson, 2006) that had a different focus, rationale and set of experimental measures.

1.2. Testing procedure

Each child was tested individually. At Time 1, all testing took place in a quiet area of the preschool or child-care centre that the participant was attending. At Time 2, those who were still attending preschool or child-care were again tested there. Those who had moved on to primary school were tested after school or on a weekend either at home or at a university laboratory, depending on their parents' preference.

1.3. Tasks, materials and scoring

At Time 1 the children completed a battery of standard ToM tests. These tests were chosen to include widely used tests of representational understanding of ToM that varied sufficiently in difficulty level to be able to pick up early signs of ToM awareness while also including standard measures like false belief. To ensure against floor effects, given that most children at Time 1 were below the age when false belief is usually passed (Wellman et al., 2001), we followed Hughes and Ensor's (2005) strategy of including tests of the understanding of mental representation in pretend play together with false belief in the Time 1 battery. We also adopted Hughes and Ensor's one-third weighting on pretend-real tasks since they had obtained a "modest but acceptable" (p. 656) Cronbach's α of .61 with such a mix. Thus our Time 1 battery had two standard false belief tests and two pretend-real tests, all requiring understanding of "conflicting mental representations" (Welch-Ross, Diecidue & Miller, 1997, p. 43), or the fact that mental representations in belief or pretence may clash with physical reality.

At Time 2, we designed a package of standard ToM tasks that would be likely to prove more challenging than those used at Time 1, and hence appropriate to children who were an average of 14 months older. Consequently, at Time 2, we replaced the Time 1 pretend-real tests with the standard appearance-reality tasks that Flavell, Flavell, and Green (1987) had discovered were more difficult than their pretend-real analogues (it is worth noting that both pretend-real and appearance-reality tasks are frequently used in conjunction with false belief tests in ToM research (Flavell, 1999; Welch-Ross et al., 1997), though rarely in sibling studies).

To increase the difficulty of the Time 2 tasks, while retaining changed-location false belief, we replaced the Time 1 misleading-container false belief with a more challenging false belief measure requiring emotional inferences. According to Hughes et al. (2000), the ability to attribute an emotion that results from a false belief can be deemed to be more difficult than the attribution of behavior in a standard false belief test, owing to the requirement for integration of two different kinds of mental states (beliefs and emotions). Their data confirmed this for typically developing 5-year olds. Thus we included Hughes et al.'s emotion-false belief test as a more difficult false belief test at Time 2, to guard against ceiling effects.

At Time 1, verbal intelligence (VIQ) was also estimated using a standardised receptive vocabulary test, the Peabody Picture Vocabulary Test (PPVT-R: Dunn & Dunn, 1989) which was scored as specified in the test manual to provide a standard score estimate of verbal ability, normed as a deviation-type IQ score with a population mean of 100 and a population standard deviation of 15. While standardized only on a U.S. population, widespread Australian use of this test has confirmed its local suitability and large-sample Australian studies (e.g., Oddy et al., 2003) have suggested it correlates well with other standardized intelligence measures that are in wide use in Australia.

Details of the ToM tasks comprising the Times 1 and 2 assessments were:

1.3.1. Time 1 ToM tests

Baron-Cohen, Leslie, and Frith's (1985) "Sally" changed location false belief test, presented and scored exactly as described by the original authors apart from replacement of the second girl doll with a boy (Peterson & Siegal, 1999). Two trials each entailed displacements of a marble that the girl did not see. When she returned wanting her marble, the same test questions on each trial ("Where will the girl look for her marble?") were followed by two control questions ("Where is the marble really?", "Where did the girl put the marble in the beginning?"). All children passed all four control questions. One point was given for each correct test question response to a total of 2 for the task. The other false belief test, from Gopnik and Astington (1988) involved a misleading container. After being made privy to its unexpected contents, children were asked: (1) "[Name of classmate] is coming next. He [she] has not seen inside this box before. When I show it to him [her] all closed up like this, what will he [she] say is in it?" and (2) "When I first showed you this box, before you looked inside, what did you think was in it?" These test questions were followed by a control question: "What is really in it?" that all children passed. Correct test question responses earned 1 point each, so the total could range from 0 to 2. This was summed with the changed location total to yield a total false belief score for Time 1 that could (and did) range from 0 to 4.

There were two pretend representation tests, both involving "junk" objects (a carrot and a potato). Pretence was elicited first with one of the objects (e.g., "Pretend this [carrot] is a phone"). After minimal prompting (e.g., "Your phone's ringing"), all children successfully engaged in actions revealing their understanding and acceptance of the pretend stipulation (e.g., holding the carrot to their ear). The tester said: "Okay, we have finished our game now. We have stopped pretending" and gave two test questions: (1) "What is this [carrot] really?" (2) "What did we pretend this was?". Displaying a real version of the pretend item (e.g., a real mobile phone), the junk item used and an irrelevant decoy (e.g., a tennis ball), the final test questions were: (3) "Which one did we pretend was a [phone]?" and (4) "Which one is really a [phone]?" The potato was used as soap in an identical procedure. Thus for each task, using different syntax, there were two questions about real identities and two about pretend identities of the objects. Since children could have answered any individual question correctly by chance, and since a child who got only some items right (for example, by naming the real identity of the pretend object throughout) could be seen to understand less about pretending than another who got all questions wrong by mixing up "pretend" and "real" labels, we imposed the conservative scoring criterion of requiring correct answers to all four questions to pass each test with a score of 1 point, to a maximum total of 2 for the two pretend tests combined.

1.3.2. Time 1 ToM composite score

In order to obtain as sensitive a ToM measure as possible, we assessed the possibility of combining children's scores on the false belief total (0–4) and the pretend total (0–2) into a reliable composite. For the present sample, Cronbach's alpha for the sum of these two scores was $\alpha = .59$. This was deemed to indicate modest, but adequate, internal consistency in order to use the combined total score (0–6) as our main index of Time 1 ToM understanding.

1.3.3. Time 2 ToM tests

The same 2-trial changed location false belief test from Time 1 was used at Time 2 with new objects, dolls and hiding places. As before, all children passed all control questions. Correct test question responses earned 1 point and were summed so that the task total ranged from 0 to 2. The more advanced false belief test, requiring the inference of emotions based on false beliefs was adapted from Hughes et al. (2000), who themselves had adapted a test from Harris, Johnson, Hutton, Andrews, and Cooke (1989). In our version, a puppet who liked animal cookies was hidden while a clearly labelled box of these was emptied and filled with rocks. The puppet returned, wanting a snack, and children were asked two test questions, one requiring a prediction of the puppet's emotion before, and the other after, he opened the closed cookie box. Most children named an emotion (e.g., "sad"). If not, or if their response that was ambiguous (e.g., "confused"), a picture scale with happy, neutral and sad cartoon faces was presented to allow for pointing responses. Children earned 1 point for stating that the puppet would be happy with the gift of the cookies before he opened the box and 1 point for a negative emotion ("sad", "angry") after he looked. Totals for the task ranged from 0 to 2 and were summed with the changed location total to yield a Time 2 false belief total score that could (and did) range from 0 to 4.

For the two appearance-reality (AR) tests we used a procedure based closely on Flavell, Green, and Flavell (1986). One test involved a fruit-shaped candle and the other a ballpoint pen that looked like a carrot, banana or flower. The tester demonstrated the object's real identity (e.g., lit the candle) and asked two test questions (counterbalanced): "What is this really and truly?" and "When you look at this with your eyes right now, what does it look like, [pause] does it look like a [candle] or like [an apple]?" Accurate understanding of the appearance-reality distinction requires that a child genuinely distinguish the two concepts. Thus one who gets only a single test question right (e.g., by saying "apple" to both questions) can be deemed to have less understanding than who distinguishes the two constructs but transposes their labels (e.g., says its really an apple but it looks like a candle). Consequently, to pass each task (scored as 1 point), our conservative scoring required correct answers both test questions. Thus, by answering all four questions correctly, children could earn a maximum total of two points for appearance-reality as a whole.

1.3.4. Total ToM score for Time 2

The changed-location false belief, emotion-false-belief, and appearance-reality tasks displayed their predicted ordering of difficulty for this sample. The AR test proved easiest, with 62% of children scoring perfectly. The standard false belief test produced 43% perfect scores while, in line with Hughes et al. (2000), only 24% of the children in this sample scored perfectly on the emotion-false-belief task. In order to obtain as sensitive a measure of Time 2 ToM as possible, we assessed the possibility of summing totals on changed-location false belief (0–2), emotion-false-belief (0–2) and appearance-reality (0–2) tasks. Cronbach's alpha for this total was $\alpha = .61$. This was deemed to indicate adequate internal consistency for use of the composite total as our main index of Time 2 ToM understanding. Total scores ranged from 1 to 6 in this sample.

2. Results

At Time 1, the children earned a mean ToM total of 3.63 (S.D. = 1.90). No children scored zero for ToM at Time 2, and only four children (6%) did so at Time 1, revealing a low rate of floor effects. Conversely, only 15 children (24%) scored at ceiling at Time 1 by making no errors at all on any task. Only 13 (21%) were at ceiling at Time 2. Given that the Time 2 measures were deliberately chosen as to be normatively more difficult than those for Time 1 (and hence appropriately challenging to children who were 14 months older), we did not necessarily expect overall gains from Times 1 to 2. In fact, the sample's total ToM mean at Time 2 was 3.97 (S.D. = 1.60), a difference from Time 1 that was not statistically significant, $t\,(62) = 1.25$, $p > .20$. Increased difficulty of the Time 2 battery, especially the advanced emotion-false-belief tests (Hughes et al., 2000), may have compensated for any overall improvement in ToM understanding over the 14 months. However this was not problematic for our main research goals. Quite the contrary, the fact that the age-appropriate sets of ToM tests avoided both floor and ceiling effects at both times meant there was substantial room for Times 1–2 gains, and enough individual variability at each time point to enable sensitive testing of our main hypotheses.

2.1. ToM at Time 1

There was a significant correlation between the children's Time 1 ToM total scores and the number of child-aged siblings in their families, $r\ (61) = .39$, $p < .01$. To more closely examine the basis for this effect, a one-way ANOVA was conducted to compare the 10 children who had no child siblings ($M = 2.60$) with the 33 who had exactly 1 child-aged sibling ($M = 3.36$) and the 20 who had either 2 or 3 child-aged siblings ($M = 4.60$). A statistically significant difference in Time 1 ToM scores emerged, $F(2,60) = 4.99$, $p < .01$. A post hoc Newman–Keuls test showed that the only-children scored significantly below those with two or three child-aged siblings, whereas the group with a single child-aged sibling did not differ significantly either from the only-children or from those with two siblings or more. The results for only-children are line with previous suggestions (Lewis et al., 1996; Perner et al., 1994; Peterson, 2000) that the absence of any child siblings at all in the household is particularly disadvantageous for ToM performance.

Since the ages of the sample at Time 1 ranged rather broadly, we were able to cross-sectionally test whether the sibling benefit was age-specific within this early period in development when the children were predominantly aged 3 (30%) or 4 (57%). Subdividing children's ages at the median of 4 years 2 months, a 2 (age group) × 3 (sibling group: 0, 1 or 2+) ANOVA revealed no significant main effect for age, $F < 1$, and no significant interaction, $F < 1$. Only the sibling effect was significant, $F(2,57) = 4.80$, $p = .01$, echoing results of the one-way ANOVA in showing higher ToM scores for those children with two siblings or more.

To examine the extent to which child siblings contributed to individual variability in children's levels of ToM understanding independently of age and verbal intelligence (VIQ), a hierarchical multiple regression analysis was conducted. With total ToM scores at Time 1 as the dependent variable, the predictors age and VIQ were entered first, as control variables, at Step 1. The regression equation differed significantly from zero at the end of this step, *Mult.* $R = .47$, $R^2 = .22$, *Adj.* $R^2 = .20$, $F(2,60) = 8.50$, $p < .01$. Next, with the entry of participants' numbers of child-aged siblings at Step 2, a significant increment in the prediction arose, R^2 (change) = .09; F (change) = 7.80, $p < .01$. The full model also remained statistically significant at the end of this final step, *Mult.* $R = .56$, $R^2 = .31$, *Adj.* $R^2 = .28$, $F(3,59) = 8.91$, $p < .001$. Table 1 shows the unstandardized regression coefficients (B), their standard errors (S.E.(B)), the standardized regression coefficients (beta) and the significance levels for all variables in the full model at the end of Step 2.

These regression results indicate that numbers of child-aged siblings made a significant independent contribution to the prediction of Time 1 ToM over and above age and VIQ, and that

Table 1
Final step of regression model predicting Time 1 ToM

Variable	*B*	S.E.(*B*)	*β*	*t*
Child age	.27	.37	.08	.74
Verbal IQ	.04	.01	.40	3.62**
Number of child siblings	.77	.28	.31	2.79**

** $p < .01$, two-tailed.

number of child-aged siblings and verbal ability each accounted for significant variability in Time 1 ToM scores in the final model.

2.2. Time 2 ToM

As at Time 1, the children's total ToM scores at Time 2 were correlated with their total child-aged siblings, r (61) = .38, $p < .01$. To locate the basis for this effect, a one-way ANOVA was conducted to compare the total Time 2 ToM scores of the 10 children who had no child siblings ($M = 2.90$) with the 33 who had exactly 1 child sibling ($M = 3.82$) and the 20 who had either 2 or 3 child siblings ($M = 4.75$). A statistically significant difference emerged, $F(2,60) = 5.48$, $p < .01$. A post hoc Newman–Keuls test showed that the only-children scored significantly below those with 2 or 3 child-aged siblings, whereas the group with a single child-aged sibling did not differ significantly from either of the other groups.

To see whether the child's age at the time of completing the second ToM assessment interacted with his or her total number of child-aged siblings in the prediction of Time 2 ToM performance, we computed a 2 (age group) × 3 (sibling group: 0, 1 or 2+) ANOVA on Time 2 ToM scores. The younger age group for this analysis ranged from 4.25 to 5.25 years ($n = 31$) and the older, from 5.33 to 6.75 years ($n = 32$). No significant effects emerged for age group, $F(1,57) = 1.75$, $p > .15$, or the interaction, $F < 1$. Only the sibling main effect was significant, $F(2,57) = 4.29$, $p < .05$, echoing results of the one-way ANOVA in showing higher ToM scores for those children with 2 or more child siblings than those with none, and also showing that this sibling influence was equally apparent in both younger and older subgroups at Time 2, just as it had been at Time 1.

To more closely examine the effects of having child siblings on individual variability in children's levels of ToM understanding at Time 2, a hierarchical multiple regression analysis was conducted. With total ToM scores at Time 2 as the dependent variable, the predictors age (at Time 2) and VIQ were entered as control variables at Step 1. The regression equation did not differ from zero at the end of this step, *Mult.* $R = .26$, $R^2 = .07$, *Adj.* $R^2 = .03$, $F(2,60) = 2.08$, $p > .10$. With only age and VIQ in the equation, age was separately significant as a predictor ($\beta = .25$, $t = 2.01$, $p < .05$), but verbal intelligence was not, $t < 1$. Though unpredicted, it is conceivable that the inclusion of emotion understanding as a component of the Time 2 false belief assessment may have contributed to its weaker link with language (Astington, 2001).

At Step 2, with the entry of the Time 1 ToM score as a predictor, no statistically significant increment arose, R^2 (change) = .05; F (change) = 3.51, $p > .05$, nor was the full model significant at a conventional level at the end of this second step, *Mult.* $R = .34$, $R^2 = .12$, *Adj.* $R^2 = .07$, $F(3,59) = 2.62$, $p = .06$. However, at Step 3, with the entry of number of child siblings, a significant increment in the prediction of Time 2 ToM arose, R^2 (change) = .07; F (change) = 4.59, $p < .05$. The full regression equation was likewise significant at the end of this final step, *Mult.* $R = .43$, $R^2 = .18$, *Adj.* $R^2 = .13$, $F(4,58) = 3.23$, $p < .05$. Table 2 shows the unstandardized regression coefficients (B), their standard errors (S.E.(B)), the standardized coefficients (beta) and the significance levels for all variables in the full model at the end of Step 3.

In other words at Time 2, when most participants were aged 5 and 6, the number of child-aged siblings made a significant independent contribution to the prediction of variability in ToM performance over and above the influences of Time 2 age, verbal IQ and Time 1 ToM scores. In the final model at Step 3, only number of child siblings emerged as a statistically significantly predictor of variability in Time 2 ToM.

In line with the regression for Time 1, and with the ANOVA results at both time points, these results suggest, that throughout the typical ToM acquisition period from the early 3 s through age 6

Table 2
Final step of regression model predicting Time 2 ToM

Variable	*B*	S.E.(*B*)	*β*	*t*
Time 2 age	.38	.34	.14	1.14
Verbal IQ	.01	.01	−.09	.69
Time 1 ToM total score	.14	.12	.16	1.14
Number of child siblings	.59	.28	.28	2.14*

* $p < .05$, two-tailed.

years, having a larger number of child-aged siblings continuously enhanced the ToM performance of the children in this particular sample. Furthermore, the advantage conferred by having more child-aged siblings at home, as well as remaining relatively constant across fairly broad cross-sectional and longitudinal age spans, and across the two different age-appropriate ToM composites we used, was relatively independent of verbal intelligence, especially at the older testing point.

3. Discussion

The present longitudinal findings were consistent with those of several previous cross-sectional studies in revealing significant associations between children's advanced performance on standard ToM tests and their access at home to larger numbers of child-aged siblings with whom to play and converse. In addition, these findings extended upon previous research by suggesting that access to child siblings continues to be of benefit for children's ToM performance throughout the preschool period, from age 3 to age 6. Of course, the nature of our sample must be considered in interpreting these findings in the context of previous research. The parents of our sample were mostly middle-class and well-educated. The fact of their longitudinal involvement testified to their residential stability and high motivation for research. Thus our results fail to address the possibility (Cole & Mitchell, 2000) that children in families with high mobility, low incomes and poor education may not gain ToM benefits from siblings. This possibility could account for several large-scale failures to replicate the sibling-ToM link (e.g., Cutting & Dunn, 1999; Hughes & Ensor, 2005). Furthermore, the ToM tasks we used, though including related concepts like pretence-reality and perception-reality contrasts, had a strong emphasis on false belief. It is conceivable that false belief may be one ToM concept that benefits particularly strongly from child–sibling interaction, and only-children may be selectively deprived of exposure to false beliefs. Their conversational partners (parents) are generally well-informed and truthful communicators, whereas young siblings' conversations may expose preschoolers to many instances of false beliefs via ignorance, misinformation, fantasies, disputes, innocent mistakes, and deception. As Wright-Cassidy et al. (2005) suggested, parents with young siblings of mixed ages may also discuss false beliefs more in order "to explain one child's viewpoint to the other" (p. 104). But only-children may not miss out as much in exposure to other ToM concepts, like desire or intention, even when interacting exclusively with adults at home.

If access to interaction with child-aged siblings at home does have special value for advancing the understanding of false belief, as opposed to other aspects of social understanding, it would be useful in future research to further broaden the assessment of children's ToM beyond the disproportionate emphasis on false belief that has prevailed in both the present, and previous, sibling-ToM studies. In the context of recommending such a broadening of empirical approach, Astington (2001) noted, "In looking beyond false belief understanding to differences in desire and intention understanding one may find that . . . there may be consequences for social sensitivity that are independent of language" (p. 687).

The contemporaneous and longitudinal links that we observed between child–sibling numbers and higher scores on standard ToM tests are consistent with a direction of influence from child siblings to advanced mental-state understanding, rather than vice versa. Indeed, it is hard to imagine how the reverse sequence (in which children's own ToM scores could cause them to have more or fewer siblings) could operate, short of parents altering their childbearing decisions in response to the levels of social understanding displayed by their older offspring. Given the closely spaced ages of the child siblings in our sample, this seems unlikely. Nevertheless, a simple path from child siblings to enhanced ToM understanding is not the only possibility. Some third variable could conceivably explain these longitudinal and cross-sectional links. For example, parental attitudes and parenting behaviors could possibly be underlying factors, along with parental conversational styles (Wright-Cassidy et al., 2005). It is known that families differ from one another in the extent to which parents are "mind-minded" (Meins & Fernyhough, 1999). Perhaps highly mind-minded parents, who ascribe beliefs and other cognitive states to their offspring from an early age, may interact with their children in ways that foster rapid ToM growth. At the same time, mind-minded parents might plan large, closely spaced families so as to give each of their offspring the benefit of social interaction with like-minded siblings. Many other possibilities are also conceivable, including mediating links between child siblings and ToM via factors such as executive functioning (Cole & Mitchell, 2000; Hughes & Ensor, 2005; McAlister & Peterson, 2006), language (Jenkins & Astington, 1996) or pretend play (Harris, 2005).

Though causality cannot be decided by data like these, the present study's support for a link between ToM and numbers of child siblings does assist theoretical understanding by suggesting a role for nurturant social experiences, rather than purely innate neurobiological maturation, in ToM development. From a strict nativist perspective, it is hard to see how family access to siblings could influence the pace of maturation of a preordained mental-state module. Of course, many nativist theories accord a role to the social environment in triggering the onset of modular maturation (e.g., Scholl & Leslie, 2001), so the link between siblings and ToM is not entirely incompatible with nativist positions.

From a "nurture" perspective, in which culture and accumulating experiences drive the pace of ToM acquisition, it is intriguing to speculate on what kinds of specific experiences arising in households with child-aged siblings, but not so often in only-children's families, could conceivably contribute to the rapid growth of false belief and other ToM concepts. In this context, it seems clear that sibling numbers are not direct causal factors, in and of themselves. Instead, like a child's chronological age, the sibling complement merely testifies to the operation for some underlying, not yet fully specified, direct cause. To view the analogy more clearly, in terms of age, a child in a coma who simply gained years of age without having conscious social experiences would not be expected to have mastered ToM if he or she suddenly awoke from the coma at age 5 and could somehow be tested for it. Similarly, in the unlikely case of a child having siblings at home with whom he or she never interacted, no sibling advantage for ToM would be expected. Thus number of child-aged siblings can, like chronological age, be deemed a visible marker for some as-yet-not-fully visible causal influences. Though possibilities must remain speculative, it is interesting to consider what these more directly formative experiences might be.

In line with the longitudinal pattern we observed, previous research has identified a number of potentially valuable sibling-based family social experiences that are likely to be more frequent in households with larger numbers of young siblings throughout the whole of childhood (Dunn, 1996). As noted earlier, children with siblings have rich opportunities not only to converse with their siblings directly, using cognitive verbs and syntax, but also to eavesdrop upon conversations, negotiations, reminiscences, pretence, or disciplinary encounters between siblings and parents. Overhearing conversations could plausibly enable greater reflection than active participation, as for an only-child. Throughout childhood, sibling rivalries and conflicts are also frequent in large families, especially when siblings are closely spaced in age. Disputes between siblings provide many opportunities for exposure to others' discrepant mental perspectives (Herrera & Dunn, 1997). Conflict may give rise to negative emotions that call forth instructive mentalistic explanations (Lagatutta & Wellman, 2002) and sibling conflicts, together with team and competitive games, might also foster executive functioning skills like planning or inhibition of impulsiveness (Cole & Mitchell, 2000). In sharing in pretend play with siblings, or while talking about it (Brown et al., 1996), children are likely to gain exposure to imaginary ideas, skills for negotiating the sharing of fantasies, and rich role-taking experience (Harris, 2005). All of these could conceivably assist ToM development in a manner not so readily available to an only-child. Further research, longitudinal as well as cross-sectional, would be valuable in order to further explore these promising possibilities.

References

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

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

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology*, *18*, 279–295.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development*, *70*, 853–865.

Dunn, J. (1996). The Emanuel Miller Memorial Lecture: 1995. *Journal of Child Psychology & Psychiatry*, *37*, 507–518.

Dunn, J., & Kendrick, C. (1982). The speech of two- and three-year-olds to infant siblings: "Baby talk" and the context of communication. *Journal of Child Language*, *9*, 579–595.

Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child Development*, *60*, 399–410.

Dunn, L. M., & Dunn, L. M. (1989). *Peabody picture vocabulary test—revised*. Minnesota: American Guidance Service.

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

Flavell, J. H., Flavell, E. R., & Green, F. L. (1987). The effects of question clarification and memory aids on young children's performance on appearance-reality tasks. *Cognitive Development*, *2*, 127–144.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development*, *51*, 1–68.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearanceeality distinction. *Child Development*, *59*, 3–26.

Harris, P. L. (2005). Conversation, pretense and theory of mind. In J. W. Astington, & J. Baird (Eds.), *Why language matters for theory of mind* (pp. 70–83). New York: Oxford University Press.

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, A. (1989). Young children's theory of mind and emotion. *Cognition & Emotion*, *3*, 379–400 [Special Issue: Development of emotion-cognition relations].

Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry*, *41*, 483–490.

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: a family affair? *Developmental Neuropsychology*, *28*, 645–668.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology*, *32*, 70–78.

Lagatutta, K., & Wellman, H. M. (2002). Differences in early parent-child conversations about negative versus positive emotions. *Developmental Psychology*, *38*, 564–580.

Lewis, C., Freeman, N. H., Kyriakidou, C., Maridaki Kassotaki, K., et al. (1996). Social influences on false belief access: Specific sibling influences or general apprenticeship? *Child Development*, *67*, 2930–2947.

McAlister, A. R., & Peterson, C. C. (2006). Mental playmates: Siblings, executive functioning, and theory of mind. *British Journal of Developmental Psychology*, *24*, 733–751.
Meins, E., & Fernyhough, C. (1999). Linguistic acquisitional style and mentalizing development. *Cognitive Development*, *14*, 363–380.
Oddy, W., Kendall, G., Blair, E., de Klerk, N., Stanley, F., Landau, L., et al. (2003). Breast feeding and cognitive development in childhood. *Paediatric & Perinatal Epidemiology*, *17*, 81–90.
Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development*, *65*, 1228–1238.
Peterson, C. C. (2000). Influence of siblings' perspectives on theory of mind. *Cognitive Development*, *15*, 435–455.
Peterson, C. C., & Siegal, M. (1999). Representing inner worlds: Theory of mind in autistic, deaf, and normal hearing children. *Psychological Science*, *10*, 126–129.
Peterson, C. C., & Siegal, M. (2002). Mindreading and moral awareness in popular and rejected preschoolers. *British Journal of Developmental Psychology*, *2*, 205–224.
Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, *34*, 161–174.
Slomkowski, C., & Dunn, J. (1996). Young children's understanding of other people's beliefs and feelings and their connected communication with friends. *Developmental Psychology*, *32*, 442–447.
Welch-Ross, M. K., Diecidue, K., & Miller, S. A. (1997). Young children's understanding of conflicting mental representation predicts suggestibility. *Developmental Psychology*, *33*, 43–53.
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.