Journal of Experimental Child Psychology 149 (2016) 41–61

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Conversations about mental states and theory of mind development during middle childhood: A training study

Federica Bianco [a,*], Serena Lecce [a], Robin Banerjee [b]

[a] Department of Brain and Behavioural Science, Università of Pavia, 27100 Pavia, Italy
[b] School of Psychology, University of Sussex, Falmer, Brighton BN1 9QH, UK

ARTICLE INFO

Article history:
Available online 23 December 2015

Keywords:
Theory of mind
Conversation
Intervention
Development
Mechanism
Middle childhood

ABSTRACT

Despite 30 years of productive research on theory of mind (ToM), we still know relatively little about variables that influence ToM development during middle childhood. Recent experimental studies have shown that conversations about the mind affect ToM abilities, but they have not explored the mechanisms underlying this developmental effect. In the current study, we examined two potential mechanisms through which conversations about mental states are likely to influence ToM: an increased frequency of references to mental states when explaining behavior and an increased accuracy of mental-state attributions. To this aim, we conducted a training study in which 101 children were assigned to either an intervention condition or a control condition. The conversation-based intervention was made up of four sessions scheduled over 2 weeks. Children completed a battery of assessments before and after the intervention as well as 2 months later. The groups were equivalent at Time 1 (T1) for age, family affluence, vocabulary, and executive functions. The ToM group showed an improvement in ToM skills (as evaluated on both the practiced tasks and a transfer task). Mediation analyses demonstrated that the accuracy of mental-state attributions, but not the mere frequency of mental-state references, mediated the positive effect of conversations about the mind on ToM development. Our results indicate that conversational experience can enhance mental-state

* Corresponding author. Fax: +39 0382986132.
E-mail address: federica.bianco@unipv.it (F. Bianco).

http://dx.doi.org/10.1016/j.jecp.2015.11.006

reasoning not by simply drawing children's attention to mental states but rather by scaffolding a mature understanding of social situations.

Introduction

Theory of mind (ToM) is defined as the ability to explain and predict social behavior on the basis of mental states (Wimmer & Perner, 1983). Three decades of fruitful research on ToM have clarified the existence of substantial individual differences between children of the same age during early childhood (e.g., Cutting & Dunn, 1999; Hughes et al., 2005) and, more recently, during middle childhood (Banerjee, Watling, & Caputi, 2011; Devine & Hughes, 2013; Lecce, Zocchi, Pagnin, Palladino, & Taumoepeau, 2010). Crucially, these individual differences play a key role in explaining the development of language (Tomasello, 2003), socially competent behaviors (Caputi, Lecce, Pagnin, & Banerjee, 2012), and metacognition (Lecce, Bianco, Demicheli, & Cavallini, 2014). Given this, authors have tried to understand the factors that are responsible for these individual differences.

Conversational approach to ToM development

Several strands of evidence have demonstrated that conversations about the mind are a key factor in determining individual differences in ToM (e.g., de Rosnay & Hughes, 2006; Hughes & Dunn, 1998). According to the *conversational approach*, the understanding of mental phenomena emerges from mental-state conversations because talking about the mind fosters the coordination of different points of view on the same event, stimulates the comparison between one's own and others' mental states, and promotes reflection about social experiences (Carpendale & Lewis, 2004; Nelson, 2005; Symons, 2004; Turnbull & Carpendale, 1999). Support for such a view comes from both longitudinal studies (Ensor & Hughes, 2008; Ruffman, Slade, & Crowe, 2002) and training studies (Ornaghi, Brockmeier, & Grazzani, 2011). This last group of studies has shown that involving children in conversations about mental states (beliefs, desires, and perceptions) improved their ToM (Appleton & Reddy, 1996; Slaughter & Gopnik, 1996). The use of sentential complement constructions (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003), feedback, and explanations in these conversations seems to be crucial (Clements, Rustin, & McCallum, 2000; Melot & Angeard, 2003).

It should be noted that the great majority of these studies have been conducted on preschoolers. Only recently is research shifting toward investigating the role of conversations about mental states in ToM development during the school years. Here it is important to note that during the school years children not only develop more complex understanding of mental states (for more comments on this issue, see Apperly, 2011, and Miller, 2009) but also become better at *using* their understanding of mental states in a more flexible and appropriate way (Apperly, 2012; Devine & Hughes, 2013) and at comprehending the subtle circumstances and conditions that influence the construction of representations when people interact.

In the current study, we build on these conceptual developments and consider *if* and *how* conversations about mental states affect older children's ToM. Research in this field is encouraging because it shows that taking part in conversations about mental states with peers and adults helps the development of ToM not only in preschoolers but also in primary school-aged children. For example, studies with deaf children have shown that the chance of taking part, at school, in such conversations with numerous and different partners enhances ToM skills (Meristo et al., 2007; Tomasuolo, Valeri, Di Renzo, Pasqualetti, & Volterra, 2013). In a similar vein, a longitudinal study of typically developing children has demonstrated that mothers' mental-state talk at child age 2 years predicts individual differences in children's ToM at ages 6 and 10 years (Ensor, Devine, Marks, & Hughes, 2014).

However, the strongest evidence for the role of mental-state conversations in the development of ToM during the primary school years comes from two recent intervention studies. The first is a study by Ornaghi and colleagues showing that mental-state conversations with peers are necessary for the development of emotion and false belief understanding in 7- and 8-year-olds (Ornaghi, Brockmeier, & Grazzani, 2014). The authors showed that children who took part in mental-state conversations about stories showed higher improvements in ToM than children who listened to the same stories but were asked to make a drawing after the stories instead of taking part in mental-state conversations. The second training study is an investigation by Lecce and colleagues on 9- and 10-year-olds (Lecce, Bianco, Devine, Hughes, & Banerjee, 2014). The authors designed an intervention using the Strange Stories task (Happé, 1994) as stimuli and distinguished between two conditions. This study included two training conditions: a ToM condition and a control condition. The ToM training condition was based on the ToM stories and consisted of group conversations about mental states and mental-state verbs involved in complex social situations. The control training condition was based on the physical stories and consisted of group conversations about physical states and verbs. Results showed that, after the intervention, children attending the experimental group were better at inferring mental states in the Strange Stories task than children attending the control group. Importantly, this effect was not due to differences between groups in verbal ability, executive functions, reading comprehension, and family affluence.

Overall, these two training studies are promising because they demonstrate that conversations about the mind *causally* impact ToM skills in children attending primary school. However, although these studies were interesting and innovative, they did not explore the mechanisms underpinning the training effects and, therefore, left open questions about *how* conversations about the mind impact ToM skills. Answering this question is theoretically compelling because it would allow us to better understand how modifications in ToM skills occur beyond the early acquisition of false belief understanding. To examine this issue, we adopted a training methodology because intervention studies of the kind described above allow us to evaluate the plausibility (and the power) of potential mechanisms of developmental change (Siegler & Crowley, 1991). Specifically, we focused on possible explanations of improvements that the ToM training program was expected to stimulate on a ToM "far-transfer" task. By transfer task, we mean a task that measures ToM in a very different way from the tasks practiced during the training program. More specifically, we selected a task that differed in structure (involving humans vs. geometric shapes) and modality (verbal static vs. visual dynamic) from those used in the training (see Method section below). The evaluation of transfer effects using a far-transfer task enabled us to examine potential developmental mechanisms in meaningful ToM improvements beyond the superficial learning strategies that might help children to simply become better with practice at a limited range of ToM stories.

Potential mediating factors in the effects of ToM training

Previous literature has pointed out that conversations about mental states are crucial for ToM development because they draw children's attention to inner states, explicitly connect the inner world to overt behavior, and shape children's expectations/experience of interpersonal events (Slaughter & Peterson, 2012). Interestingly, work by Thompson and colleagues attributes the positive effect of mental-state conversations on ToM to two mechanisms: increased awareness of the existence of mental states and improved capacity to perform accurate reasoning in this domain (Ontai & Thompson, 2008; Thompson, 2006). Notably, although a correlation between these two indexes (frequency of mental-state references and accuracy of mental-state attributions) is expected, it is also important to note that they can be dissociated, at least in older children (Hughes, 2011). It is, indeed, entirely possible that a child makes many references to mental states without being accurate and that another child makes relatively few references to mental states while being high in accuracy of mental-state inferences. Support for this view also comes from clinical psychology showing that high-functioning autistic children can refer to mental states with a frequency similar to that of typically developing children but show significantly less accuracy (Abell, Happé, & Frith, 2000; Happé, 1994).

In the current study, we examined the role of both these potential mechanisms in ToM development during middle childhood. According to the first explanation, mental-state conversations are likely to be effective for promoting ToM because they affect children's propensity to pay attention to the inner world of social agents; this, in turn, could trigger positive effects in performance on ToM tasks. This is plausible given that existing evidence already points to such an explanation for the normative developmental improvement in ToM during middle childhood. In particular, between 7 and 11 years of age, children intensify the frequency of mental states to describe their best friend (Meins, Fernyhough, Johnson, & Lidstone, 2006), to narrate stories (Peterson & Slaughter, 2006), and to invent and write the final part of a vignette (Longobardi, Piras, & Presaghi, 2008). This is especially important when we consider that (a) cognitive factors are not predictors of the tendency to refer to inner states (Camaioni, Longobardi, & Bellagamba, 1998) and that (b) children's frequency of use of mental-state terms predicts ToM scores during middle childhood (Peterson & Slaughter, 2006). Moreover, research has shown that children who engage in frequent use of mental-state terms in conversations are more likely than their peers to perform well on ToM tasks a year later (Hughes & Dunn, 1998). Remarkably, this relation was not strengthened when only "genuine/accurate" mental-state references were included. This indicates that—at least in this study—the overall tendency to focus on mental states can potentially be regarded as a meaningful predictor of later ToM performance. Following this line of interpretation, in the current study we should find that an increase in use of the mental-state lexicon to explain social behavior would account for improvements in the ToM far-transfer measure.

On the other hand, a second mechanism through which mental-state conversations could influence ToM development is the accuracy of mental-state attributions. Specifically, the exposure to mental-state conversations over time might not only encourage greater attention to mental states but also help children learn how to (accurately) detect the particular mental states that clarify what is happening in a given social situation (Apperly, 2011). During a social interaction, indeed, fast on-line inferences about mental states guide people toward the most likely interpretations of the words/actions of the actors (Grice, 1957; Sperber & Wilson, 2002). The improvement in the ability to infer context-sensitive mental states during school years is signaled by an increase in the understanding of the subtle circumstances and conditions that influence the construction of representations, as shown by improved performance in the resolution of faux pas tasks (i.e., the ability to understand the mental states involved in social gaffes; Banerjee et al., 2011) and interpretive ToM tasks (i.e., the ability to understand that beliefs are interpretations rather than reproductions of reality; Lalonde & Chandler, 2002). A link between understanding complex social situations and the ability to rely on social context-sensitive information is also suggested by studies of high-functioning autistic children. Work by Vivanti and colleagues (2011), for example, showed that high-functioning autistic children base their social reasoning on an irrelevant aspect of the scenarios (e.g., an object's characteristics) in order to anticipate and explain people's actions. If the program actually helps children learn how to successfully perform mental-state attributions, in the current study we should find that the positive effect of our ToM training program on ToM transfer gains is mediated by the improvements in the ability to make context-appropriate mental inferences over and above any changes in use of the mental-state lexicon in performing ToM tasks.

The current study

This study was designed to investigate how conversations about the mind may enhance ToM skills in primary school children. To do that, we applied the training program developed by Lecce and colleagues for 9- and 10-year-olds (Lecce, Bianco, Devine et al., 2014). Specifically, we compared two conditions that were very similar in structure (stories, lexicon exercises, and group discussion) and length (four sessions): an intervention condition that was focused on mental states and a control condition that was focused on physical states. Further details are given in the "Procedure" section of Method below. Children completed a battery of assessments before the intervention (Time 1, T1), after the intervention (Time 2, T2), and 2 months later (Time 3, T3).

We first checked whether the program genuinely affected theory of mind by administering not only the task that was practiced during the intervention but also a far-transfer measure not directly experienced during the program. We used the Frith–Happé animations task (Castelli, Happé, Frith, & Frith, 2000; White, Coniston, Rogers, & Frith, 2011) for this purpose. This is a ToM task that requires subjects to infer on-line mental states from the movement patterns of geometric shapes (i.e., triangles). We chose this task as the far-transfer task because it differs in structure and modality from the material used in the training program and, therefore, can be considered as a stringent test of the effects of the training program. Second, we examined the mechanisms (use of mental-state lexicon vs. accuracy of mental-state attribution) that may have accounted for this far-transfer effect. To do that, we started by comparing a direct effect model of training condition on children's gains in the ToM far-transfer task with an indirect effects model with performance on the specific practiced task as the mediator. We then tested our two main hypotheses by evaluating the extent to which children's gains in theory of mind could be explained by increased use of the mental-state lexicon (indicating a greater propensity to pay attention to mental states) and/or by an increased accuracy of mental-state attributions. In addressing these issues, we also controlled for individual differences in variables that are known to be associated with ToM: family affluence (Hughes et al., 2005; Shatz, Diesendruck, Martinez-Beck, & Akar, 2003), verbal ability (Milligan, Astington, & Dack, 2007), executive functions (Devine & Hughes, 2014), and reading comprehension (Lecce & Hughes, 2015).

Method

Participants

The study sample consisted of 101 children (53 boys) having a mean age of 9.62 years (SD = 0.33, range = 8.83–10.42). These children were randomly assigned to one of the two training conditions: ToM or control. The ToM group consisted of 53 children (27 boys, M_{age} = 9.61 years, SD = 0.33, range = 8.83–10.33), and the control group consisted of 48 children (26 boys, M_{age} = 9.64 years, SD = 0.33, range = 9.00–10.42). Children were recruited from Year 4[1] classrooms located in two primary schools in Northern Italy. Participants were not clinically referred for any cognitive or learning difficulties. A total of 131 Italian children were initially recruited for participation in this study. Children who did not complete all of the measures or had ceiling scores on Strange Stories measures (and therefore left no room for improvement over time) were removed from the main dataset (n = 18). With the remaining 113 children, we checked whether any participants unduly affected the regression paths in our main mediation model. This procedure is strongly recommended (e.g., Cohen, Cohen, West, & Aiken, 2003; Roth & Switzer, 2002) to make sure that results are reliable and not due to a small number of children who make the regression line change substantially (McClelland, 2000). As a result of this preliminary stringent analysis, 12 children (7 from the control group) were excluded from the data set (for more details, see Results).

Materials

Control variables

Family affluence. Information about children's *family affluence* was collected using the Family Affluence Scale (Currie et al., 2008). It is a short questionnaire on family wealth. There are four questions about the following: family car ownership (range = 0–2), the participants having/not having their own unshared room (range = 0–1), the number of computers at home (range = 0–3), and the number of times the participants went on a holiday during the past year (range = 0–3). Responses to the four items were summed into an overall index of family affluence (range = 0–9).

[1] In Italy, children begin school at 6 years of age (Year 1).

Verbal ability. Children's *verbal ability* was measured through the Italian version of the Vocabulary subtest of the Primary Mental Abilities (PMA; Rubini & Rossi, 1982; Thurstone & Thurstone, 1962). This test requires children to find the synonym of 30 target words choosing among four alternatives. A time limit of 7 min was set. Possible total scores could range from 0 to 30.

Executive functions. We administered a modified version of the Tower of London (Shallice, 1982) to evaluate *planning* and the Backward Digit Span test from the Italian version of the Wechsler Intelligence Scale for Children–Revised (WISC-R; Orsini, 1997) to index *working memory*. In the Tower of London task, children were given a series of cards with the starting and final configurations of three colored balls on three pegs of different heights. Children were told that the big peg could carry all three balls, the middle peg could carry two balls, and the little peg could carry just one ball and that they could move only one ball at a time. Children were asked to imagine the number of moves necessary to obtain the final configuration and write the total number of moves they would need to complete each trial. For each item, children were credited with success (1 point) if they wrote the correct number of moves. Total scores ranged from 0 to 7. In the Backward Digit Span task (Orsini, 1997), children were presented with seven sequences of two to eight digits and were asked to recall them in reverse order. For each sequence, children were credited with success only if they recalled all of the numbers in the right order. Total scores ranged from 0 to 7.

Reading comprehension. Children's *reading comprehension* was evaluated through the MT task (Cornoldi & Colpo, 1998). This standardized task requires children to answer 10 multiple-choice questions after having read a passage silently. To provide the correct answers, inferential processes are needed because the questions do not probe literal information. Possible total scores could range from 0 to 10.

Focus variables

Theory of mind. Children's theory of mind was tested, at each time point, using the Strange Stories task (Happé, 1994; White, Hill, Happé, & Frith, 2009) as the practiced ToM measure and the Frith–Happé animations (Castelli et al., 2000; White et al., 2011) as the far-transfer ToM measure. The Strange Stories task is an advanced test of ToM that requires participants to interpret nonliteral statements, understanding the actual intentions of the speaker. We administered two stories about double bluff, one story about misunderstanding, one story about white lie, and two stories about persuasion. After reading the stories, children were asked to explain a character's sentence in a written format. No time limit was imposed. In line with scoring guidelines (White et al., 2009), we rated children's answers on the basis of the correctness of explanations using a 3-point scale: 0 for an incorrect answer, 1 for a partially correct answer, and 2 for a full and explicit answer. A second rater independently coded 25% of the responses at each time point, and interrater agreement was established using Cohen's kappa (at T1, $\kappa = .88$; at T2, $\kappa = .80$; at T3, $\kappa = .85$). The Strange Stories task has been reported as having good internal consistency in confirmatory factor analysis adjusted for measurement error, with loadings on the same latent factor ranging from .44 to .71, $p < .01$ (Devine & Hughes, 2013). The Strange Stories task also showed convergent validity with traditional measures of ToM (i.e., Wellman and Liu's (2004) ToM scale; second-order false belief tasks), $r = .42$, $p = .001$ (White et al., 2009). Total Strange Stories scores ranged from 0 to 12.

Given that successful performance on the Strange Stories task requires both specific attention to the inner world of the characters and a mature skill in performing mental-state attributions (Happé, 1994), we used this task as the basis for evaluating our hypotheses about possible mechanisms in the effects of our ToM training program. Specifically, we coded children's answers to the Strange Stories according to two separate indexes that served as our proposed mediators: the mental-state lexicon and the accuracy of mental-state attributions. The *mental-state lexicon* index reflects the propensity to consider the inner world of characters. It was computed by summing the number of mental-state terms used in children's answers, independent of their suitability or relevance for answering the question, according to a list of mental-state terms previously selected in other studies (Hughes, Lecce, & Wilson, 2007). The *accuracy* *of mental-state attribution* score reflects the extent to which the mental states attributed to the characters of the stories are appropriate given the social scenario of each narrative. More specifically, each answer was scored as 1 if the participant accurately made at least one correct attribution that took into account the mental states of all characters in the scenario; otherwise, a score of 0 was attributed. This index does not necessarily equate to a right answer to the Strange Stories task, but it accounts for whether or not there is appropriate context-sensitive processing of mental states involved in the scenario. Total accuracy of mental-state attributions scores ranged from 0 to 6. A second rater independently coded 25% of the responses at each time point, and interrater agreement was established using Cohen's kappa (at T1, $\kappa = .84$; at T2, $\kappa = .82$; at T3, $\kappa = .84$). See Appendix A for some scoring examples on the three indexes from the Strange Stories task.

The Frith–Happé animations task evaluates children's ability to rapidly attribute mental states to geometric shapes on the basis of their movements. In each Frith–Happé animation, a big red triangle and a small blue triangle made some movements on the screen. We administered three animations, and children, after viewing each video clip, were asked to explain what happened. No feedback was provided. All of the sequences were scripted to imply complex mental states. Specifically, the small triangle in each clip appeared to be mocking the big one. The score for each clip reflected the degree to which participants attributed complex intentional mental states to the shapes, according to the original guidelines. The score for each item ranged from 0 (*no deliberate action*) to 5 (*deliberate action aimed at affecting another's mental state*). Two raters independently coded 25% of the responses at each time point, and interrater agreement was established using Cohen's kappa (at T1, $\kappa = .90$; at T2, $\kappa = .80$; at T3, $\kappa = .89$). The Frith–Happé animations task has been demonstrated to show both divergent validity with matched animations focused on physical events (Abell et al., 2000) and convergent validity with other tasks assessing mental-state reasoning (Campbell et al., 2006). Total Frith–Happé animations scores ranged from 0 to 15.

Procedure

Parental written consent was obtained at the beginning of the study. Before the intervention, children were pretested in their primary school (T1). The T1 phase was scheduled in two appointments given the number of tasks administered. In this phase, children completed the Family Affluence Scale (Currie et al., 2008), the Vocabulary subtest from the WISC-R (Italian version: Orsini, 1997), the Tower of London (Shallice, 1982), the Backward Digit Span from the WISC-R (Orsini, 1997), the MT task (Cornoldi & Colpo, 1998), the Strange Stories task (Happé, 1994), and the Frith–Happé animations task (Castelli et al., 2000). At the end of the training, all children were posttested twice (T2 and T3 assessments) to investigate the training effects, the generalization of effects to tasks not directly experienced during the intervention, and the maintenance of these benefits over 2 months. T2 took place 14 days after the end of the training ($M = 13.72$ days, $SD = 5.4$). T3 took place 2 months after the end of the intervention ($M = 58.68$ days, $SD = 8.0$). At T2 and T3, children were administered the Strange Stories task (White et al., 2009) and the Frith–Happé animations task (Castelli et al., 2000).

We applied the training conditions developed by Lecce, Bianco, Devine, and colleagues (2014). Both conditions were made up of four sessions. Each training session lasted approximately 50 min and consisted of two trials: a story narrative and a language exercise. In each trial, children were first presented with a story in a written form and were asked to individually answer the story questions. When all children had written their answers, each of the questions was discussed in groups led by the researcher, who made frequent use of positive and corrective feedback, expanding children's comments and explaining the reasons why their answers were wrong or right. At the end of the discussion, the experimenter made a final comment starting from the last question and highlighting the core dimension of each type of story. Children were then asked to imagine or recall an episode similar to the one presented in the story and to tell it to their classmates. Then a language exercise was administered. Children were individually presented with a sentence from the narrative (written on a piece of paper) and were encouraged to find a synonym of a chosen verb from this sentence, selecting one from four alternatives. In the ToM condition, the stories were about mental states, the language exercises involved mental-state verbs, and the experimenter made heavy use of the mental-state lexicon within sentential complement constructions. In the control condition, the stories were about physical events, the language exercises involved physical verbs, and the experimenter made no use of the mental-state lexicon. See Appendix B for further details on the contents of the training program.

Results

Preliminary analyses

As noted in the "Participants" section of Method above, preliminary analyses were carried out to detect the presence of potential influential cases with respect to the regressions of our hypothesized mediation model. To do that, we calculated DiffBeta standardized values that allowed us to measure the change in the regression coefficients that would result from the exclusion of a particular case. Following recommended guidelines (Tarling, 2008), we excluded cases presenting DiffBeta standardized values greater than $2/\sqrt{N}$ (in our analyses >.188). To be consistent, we applied this rule regardless of to which part of the model DiffBeta referred. As noted earlier, this stringent test led us to exclude 12 cases from the main dataset.

Preliminary separate one-way analyses of variance (ANOVAs) were run in order to establish the equivalence of the two condition groups before the intervention. As shown in Table 1, no significant differences were found in the control variables, $p \geqslant .096$. The only exception was reading comprehension, on which the ToM group scored significantly higher than the control group, $t(99) = 2.81$, $p = .006$, $d = 0.56$, 95% confidence interval (CI) [0.35, 2.03]. Given these results, in all subsequent analyses, we controlled for reading comprehension.

ToM training effects

To analyze the effect of the training program, a mixed analysis of covariance (ANCOVA) was conducted on each task, with time (T1, T2, or T3) as the within-participants factor, training condition as the between-participants factor, and all control variables as the covariates to make the analyses more stringent. To break down interactions, pairwise comparisons with Bonferroni's correction for multiple comparisons at $p < .05$ were performed. Lastly, if both groups showed improvements over time, group contrasts on the amount of gain were analyzed using univariate ANCOVA. Gain scores were computed by subtracting T1 scores from the corresponding scores at T2 ($\Delta 1$) and T3 ($\Delta 2$). To examine whether the baseline in ToM accounted for the training effect on gains, we controlled for ToM performance at T1. Given the significant difference between groups in this ability, in this analysis we also included reading comprehension as a covariate. See Tables 1 and 2 for descriptive analyses on the study variables.

Table 1
Descriptive statistics on control and ToM measures.

	ToM group ($n = 53$)			Control group ($n = 48$)		
	T1	T2	T3	T1	T2	T3
Age (years)	9.61 (0.33)	–	–	9.64 (0.33)	–	–
Family affluence (0–9)	6.66 (1.4)	–	–	6.67 (1.6)	–	–
Verbal ability (0–64)	27.89 (2.2)	–	–	27.35 (3.6)	–	–
Reading comprehension (0–10)	6.77 (2.0)	–	–	5.58 (2.2)	–	–
Planning (0–7)	4.23 (2.3)	–	–	4.23 (2.3)	–	–
Working memory (0–7)	2.57 (0.8)	–	–	2.56 (1.0)	–	–
ToM: Strange Stories task (0–12)	7.77 (1.9)	10.36 (1.3)	10.57 (1.4)	7.10 (2.1)	8.98 (2.0)	9.13 (1.8)
ToM: Frith–Happé animations (0–15)	8.89 (1.7)	10.43 (1.7)	10.83 (1.6)	9.06 (2.0)	9.5 (1.6)	9.69 (2.1)
Accuracy of mental-state attributions	2.91 (1.2)	4.68 (1.1)	4.83 (1.2)	2.96 (1.3)	3.17 (1.2)	3.48 (1.5)
Mental-state lexicon	8.72 (3.00)	10.80 (4.15)	9.70 (3.64)	7.73 (3.00)	9.52 (3.50)	9.00 (3.48)

Note: Values are means (and standard deviations).

Table 2
Descriptive statistics on gains.

	ToM group	Control group
Δ1 Strange Stories task	2.58 (1.7)	1.88 (1.8)
Δ1 Frith–Happé animations	1.55 (2.0)	0.49 (2.0)
Δ1 Accuracy of mental-state attributions	1.77 (1.6)	0.21 (1.1)
Δ1 Mental-state lexicon	2.08 (4.4)	1.79 (3.2)
Δ2 Strange Stories task	2.79 (1.9)	2.02 (1.8)
Δ2 Frith–Happé animations	1.94 (2.0)	0.63 (2.2)
Δ2 Accuracy of mental-state attributions	1.92 (1.7)	0.50 (1.4)
Δ2 Mental-state lexicon	1.02 (4.1)	1.27 (3.5)

Note: Values are means (and standard deviations).

Strange Stories performance

Results for the Strange Stories task showed a significant main effect of training group, $F(1, 94) = 9.75$, $p = .002$, partial $\eta^2 = .09$, but not a significant main effect of time, $F(2, 188) = 0.81$, $p = .448$. There was also a significant time by group interaction, $F(2, 188) = 5.83$, $p = .004$, partial $\eta^2 = .06$. Pairwise contrasts revealed significant improvements in both the control condition, $p < .001$, $d = 0.92$, 95% CI [1.11, 2.31], and the experimental condition, $p < .001$, $d = 1.61$, 95% CI [2.17, 3.31], between T1 and T2. Significant improvements were also found between T1 and T3 in both the ToM group, $p < .001$, $d = 1.70$, 95% CI [2.31, 3.55], and the control group, $p < .001$, $d = 1.04$, 95% CI [1.22, 2.52]. No significant change in performance was found between T2 and T3 in any group, *p*s = 1.00. Crucially, pairwise contrasts also showed equivalent scores at T1, $p = .946$, 95% CI [−0.64, 0.68], but showed reliable differences at T2, $p = .001$, $d = 0.80$, 95% CI [0.45, 1.65], and T3, $p < .001$, $d = 0.89$, 95% CI [0.51, 1.66]. Moreover, our ANCOVA on the gain scores revealed that the ToM group exhibited higher gains than the control group at T2, Δ1 Strange Stories, $F(1, 97) = 11.27$, $p = .001$, partial $\eta^2 = .10$, and at T3, Δ2 Strange Stories, $F(1, 97) = 13.20$, $p < .001$, partial $\eta^2 = .12$.[2]

ToM far-transfer task

Results for the ToM far-transfer task showed a significant time by group interaction, $F(2, 188) = 4.91$, $p = .008$, partial $\eta^2 = .05$. The main effects of time and group were not significant, $F(2, 188) = 0.90$, $p = .408$, and $F(1, 94) = 2.46$, $p = .120$, respectively. Despite a lack of differences between groups at T1, $p = .435$, groups were significantly different at both T2, $p = .026$, $d = 0.56$, 95% CI [0.09, 1.46], and T3, $p = .026$, $d = 0.61$, 95% CI [0.10, 1.59]. Analyses also revealed that the ToM group had a significant improvement both between T1 and T2, $p < .001$, $d = 0.91$, 95% CI [0.88, 2.21], and between T1 and T3, $p < .001$, $d = 1.18$, 95% CI [1.15, 2.57]. Control group performance, in contrast, did not significantly change between T1 and T2, $p = .325$, and had only a marginally significant difference between T1 and T3, $p = .066$, $d = 0.25$, 95% CI [−0.03, 1.46].[3,4,5]

[2] The pattern of results on the Strange Stories measure does not change if participants excluded because of DiffBeta standardized values are included in the analyses.

[3] The pattern of results on the far-transfer measure does not change if participants excluded because of DiffBeta standardized values are included in the analyses.

[4] There are no significant differences between classrooms on either Δ Strange Stories or Δ Frith–Happé animations, $F(6, 46) \geqslant 1.07$, $p \geqslant .22$.

[5] Children excluded from the main analyses were not significantly different from the included children on any background variables (reading comprehension, verbal ability, executive functions, and family affluence) or ToM gains (on both the Strange Stories and triangle tasks). The only exception was found on gain scores between T1 and T2 in the Strange Stories in the control condition. Here excluded children had a significantly lower score than included children, $t(60) = 2.35$, $p = .02$.

Mediation analyses

The second main question of the current study concerned the mechanisms that may explain the positive effects of the ToM training on the ToM far-transfer task. Given that we were particularly interested in developmental processes that may account for the hypothesized positive effects of our training program, we examined whether gains in the ToM far-transfer measure might be mediated over time by improvements (Δ) in specific areas of mental-state reasoning—the mere propensity to refer to mental states when explaining social behavior and/or the accurate attribution of mental states when explaining social situations. We decided to work with gain scores ($\Delta 1$ = T2 minus T1 and $\Delta 2$ = T3 minus T1), instead of time-point scores, in order to predict actual changes in task performance.

First, we checked if and how training condition influenced the far-transfer measure. We started by comparing a direct effect model (in which the effect of the training condition on the far-transfer measure was not mediated by any variable) with an indirect effect model (in which the effect of the training condition on the far-transfer measure was mediated by performance on Strange Stories, which was the practiced task). We used the Process macro for SPSS (Hayes, 2013) for these (and subsequent) mediation analyses. Specifically, we ran our mediation analyses using training condition as the independent variable, $\Delta 2$ Frith–Happé animations score as the dependent variable, $\Delta 1$ Strange Stories as the potential mediating variable, and reading comprehension as the covariate. We report results using bootstrap tests, with a resample procedure of 1000 bootstrap samples (bias corrected and 95% CI, 95% BCa CI). If the CI does not include 0, the mediated pathway is significant. This test has greater power and more appropriate Type 1 error rates than Baron and Kenny's (1986) approach (MacKinnon, 2008).

Results showed that training condition predicted $\Delta 1$ Strange Stories, $B = 0.97$, CI [0.27, 1.68], and $\Delta 1$ Strange Stories predicted $\Delta 2$ Frith–Happé animations, $B = 0.29$, CI [0.05, 0.53]. Results showed that training condition had a significant total effect on $\Delta 2$ Frith–Happé animations, $B = 1.11$, CI [0.26, 1.96]. Crucially, training condition had only a marginally significant direct effect on $\Delta 2$ Frith–Happé animations, $B = 0.83$, CI [−0.03, 1.69], but had a significant indirect effect via $\Delta 1$ Strange Stories, $b = .28$, 95% BCa CI [0.07, 0.65]. This represent a relatively small effect, $k^2 = 0.04$, 95% BCa CI [0.003, 0.10].[6]

Overall, these results support the indirect effect model and suggest that the positive effect of the ToM program on the far-transfer task was obtained via the improved performance in the Strange Stories task. Therefore, as a second step, we tested which specific mediator(s) could account for the positive effects of the training program by reconfiguring the Strange Stories performance as two scores: one concerning the mere frequency of the mental-state lexicon and the other concerning the accuracy of the mental-state attributions. Notably, at each time point, both the proposed mediators (mental-state lexicon and accuracy of mental-state attribution) independently predicted at least 44% of the variance of the overall standard measure of Strange Stories performance, $\Delta R^2 \geqslant .44$, $F(7, 93) \geqslant 10.36$, $p < .001$, after controlling for baseline scores. This result allowed us to go on to consider each mediator independently of the other.

Table 3 shows correlations between improvements from T1 to T2 ($\Delta 1$) and from T1 to T3 ($\Delta 2$) in the following variables: Frith–Happé animations scores, mental-state lexicon index, and accuracy of mental-state attributions index. As can be seen, both of the proposed mediators ($\Delta 1$ mental-state lexicon and $\Delta 1$ accuracy of mental-state attributions) correlate with $\Delta 2$ Frith–Happé animations scores.

The path diagram describing our multiple mediator model is shown in Fig. 1. This mediation model tests both of the proposed mediators at the same time. The inclusion of the two mediators simultaneously allows us to assess the effect of each mediator after controlling for the other mediator. Moreover, effect ratios are calculated to express the amount of the total effect that is explained by each significant indirect effect. Specifically, we ran our mediation analyses using training condition as the independent variable, $\Delta 2$ Frith–Happé animation score as the dependent variable, $\Delta 1$ accuracy

[6] To exclude the possibility that what children first learned through training was something general about mental states (rather than something task specific conferring an advantage in terms of Strange Stories performance), we tested a model in which $\Delta 1$ Frith–Happé animations was a mediator of training effects on $\Delta 2$ Strange Stories. However, this model did not show good fit to the data, and there was no evidence for this indirect pathway.

Table 3
Partial correlations between T1 versus T2 gain scores (Δ1) and T1 versus T3 gain scores (Δ2) in Frith–Happé animations, mental-state lexicon, and accuracy of mental-state attributions when controlling for reading comprehension scores

	Δ1 Accuracy MS attributions	Δ1 Strange Stories	Δ1 Frith–Happé animations	Δ2 MS lexicon	Δ2 Accuracy MS attributions	Δ2 Strange Stories	Δ2 Frith–Happé animations
Δ1 MS lexicon	.47***	.41***	.23*	.69***	.34***	.27**	.17+
Δ1 Accuracy MS attributions	–	.63***	.30**	.30**	.74***	.50***	.38***
Δ1 Strange Stories		–	.17+	.25*	.46***	.61***	.29**
Δ1 Frith–Happé animations			–	.12	.35***	.15	.59***
Δ2 MS lexicon				–	.34***	.26**	.12
Δ2 Accuracy MS attributions					–	.68***	.38***
Δ2 Strange Stories						–	.27***
Δ2 Frith–Happé animations							–

Note: MS, mental state.
+ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

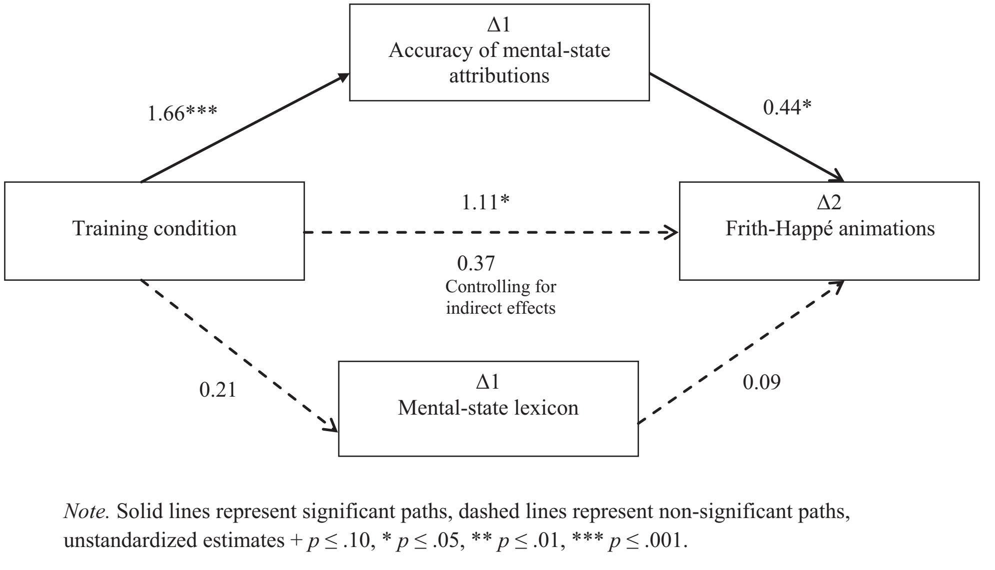

Fig. 1. Δ1 (T2 minus T1) accuracy of mental-state attributions and Δ1 mental-state lexicon as potential mediators of the association between training condition and Δ2 (T3 minus T1) Frith–Happé animations. Solid lines represent significant paths, and dashed lines represent nonsignificant paths (unstandardized estimates), * $p \leqslant .05$; *** $p \leqslant .001$.

of mental-state attributions and Δ1 mental-state lexicon as the potential mediating variables, and reading comprehension as the covariate.

Results showed that training condition predicted Δ1 accuracy of mental-state attributions, $B = 1.66$, CI [1.09, 2.20], but not Δ1 mental-state lexicon, $B = 0.21$, CI [−1.40, 1.81]. In turn, accuracy of mental-state attributions predicted Δ2 Frith–Happé animations, $B = 0.44$, CI [0.11, 0.78]. In contrast, Δ2 Frith–Happé animations was not independently predicted by Δ1 mental-state lexicon, $B = 0.009$, CI [−0.11, 0.13]. Lastly, results demonstrated that training condition had a significant indirect effect on Δ2 Frith–Happé animations via Δ1 accuracy of mental-state attributions, $b = .74$, 95% BCa CI [0.29, 1.31]. The proportion of the total effect explained by the significant mediator Δ1 accuracy of mental-state attributions was 0.46, CI [0.14, 1.36]. This represents a medium effect size, $\kappa^2 = .14$, 95% BCa CI [0.06, 0.23]. The resulting pattern of mediation is illustrated in Fig. 1.

Discussion

The current study builds on the efforts of recent research in developmental psychology to understand the nature of ToM development (Benson, Sabbagh, Carlson, & Zelazo, 2013; Wellman & Peterson, 2013). In a shift from the traditional interest in the preschool years, our study examined processes of ToM improvement in 9- and 10-year-olds. Our main aim was to examine the mechanisms that account for ToM developments using a conversational approach. In doing so, we compared the effects of two potential mechanisms through which mental-state conversations were expected to influence ToM during middle childhood: an increase in the tendency to take into account inner states and the accuracy in making appropriate context-sensitive mental-state attributions. Crucially, in this study, we adopted a training methodology because it offers a rigorous way of assessing changes while they are occurring. We selected Lecce and colleagues' ToM training program because it is based on mental-state conversations and has been developed for children attending primary school (Lecce, Bianco, Devine et al., 2014). Overall, our results shed light on the role of mental-state conversations in the development of children's ToM and on the way in which this development occurs in 9- and 10-year-olds. We first address the overall impact of the ToM training program before turning to our key finding regarding the role played by improved accuracy in making attributions of mental states.

Training of ToM during primary school years

Our first important result is that group conversations about mental states promote a genuine change in children's ability to explain social behaviors on the basis of inner states. Four kinds of results support such a conclusion. First, our data showed that, compared with the control group (matched for family affluence, age, verbal ability, planning, working memory, and T1 ToM performance), children in the experimental group performed significantly better on ToM tasks. Second, the positive effect of our intervention was evident not only in the specific ToM task practiced during the intervention (i.e., the Strange Stories) but also in the far-transfer ToM task (i.e., the Frith–Happé animations) for which children received no direct training. Indeed, our analyses revealed that being exposed to mental-state conversations help children to attribute complex mental states to moving shapes. Third, the advantage of children who took part in mental-state conversations was evident at both T2 and T3, showing that the positive effects of our training program last over a reasonable length of time. Finally, the effect of mental-state conversations generalized to the whole group given that it was independent of individual differences on all T1 measures (age, planning, working memory, reading comprehension, family affluence, and verbal ability).

Altogether, these results corroborate those reported by Lecce and colleagues with the same training program (Lecce, Bianco, Devine et al., 2014) and provide further evidence of the efficacy of the conversation-based training program. In addition to confirming preexisting findings, the current study also expands them by addressing unexplored issues. As far as we know, this is the first study that has examined in typically developing children the efficacy of a group-based ToM training on a far-transfer task. Previous studies have, indeed, showed a transfer effect between first-order false belief understanding and the appearance–reality distinction task (Melot & Angeard, 2003; Slaughter & Gopnik, 1996) or between cognitive and affective ToM tasks (Ornaghi et al., 2014). These results are certainly interesting; however, the tasks used in these studies are close, in both modality and materials, to the training activities, and this similarity may have had a role to play in yielding the positive results. Notably, in our study, we used a transfer task that was very different from the training material and can be considered as a stringent test of the efficacy of the conversation-based training program. The transfer effect on the Frith–Happé animations is also relevant, we believe, because research demonstrated that in order to understand what is happening when actions are performed under novel and unusual situations (just like the ones presented in the Frith–Happé task), children appear to activate processes similar to the ones implied in the comprehension of real-life interactions (Brass, Schmitt, Spengler, & Gergely, 2007).

Mental-state conversations and ToM development: Key mechanisms

The main aim of the current study was to examine the mechanisms through which mental-state conversations may affect children's ToM during the school years. Our study was designed in such a way that it enabled us to disentangle the effect of two plausible mediators. The first was the propensity to pay attention to the mind as indexed by the frequency of the mental-state lexicon. The second was the ability to make context-sensitive inferences about mental states, as indexed by our score for the accuracy of mental-state attributions. Results of our analyses clearly showed that it is was the latter that accounted for the training effects. In other words, our training did not simply teach children that mental states are relevant and, thus, deserve attention; rather, it also helped children to put mental states in context and to make accurate judgments about them on the basis of contextual information.

We believe that these results may be explained by considering the features of the mental-state conversations that we used in our training program. Two facets seem to be particularly relevant to us: the presence of feedback with explanations and the group nature of these conversations. Previous research on ToM training has indeed shown that informing children about whether their answers were right or not and explaining why is necessary in order to produce a change in ToM competence (Clements et al., 2000; Melot & Angeard, 2003). From a broader perspective, research on family conversations has highlighted the importance of the quality versus quantity of mothers' use of internal state talk for children's ToM development. In a seminal study, Dunn and colleagues revealed that children from families whose conversations were characterized by high frequencies of causal sentences were more likely than their peers to succeed on false belief and emotion understanding tasks 7 months later (Dunn, Brown, & Beardsall, 1991). Garner and colleagues extended these results to a low-income sample and showed that mothers' explanations about the causes and consequences of emotions, but not unelaborated comments about emotions, correlate with children's emotion knowledge and emotion understanding (Garner, Carlson Jones, Gaddy, & Rennie, 1997). Crucially, mothers' explanations about emotions seem to become increasingly important for children's emotion understanding as children grow up. Indeed, Cervantes and Callanan (1998) showed that at 4 years of age, but not at 2 or 3 years of age, it was mothers' explanations about emotions and not simply the frequency of emotion state terms that correlated with children's total emotion utterances. Generally speaking, giving feedback and explanations during conversations makes explicit the connections among the world, the inner states of characters, and overt behavior (Slaughter & Peterson, 2012).

It should also be noted that the conversations examined in the current study occurred within the context of the classroom and not between two partners such as the ones typically examined in the literature on mothers' mental-state talk. This is, we believe, an important feature of our training program that is likely to have maximized the positive effects on children's ToM. The fact that all class members discussed the same story/topic in a structured way may have made very salient the similarities or differences between children's points of view, and this is likely to have enhanced children's capacity to reason in psychological terms (Harris, 1999). When combined with the feedback and explanations discussed above, these conversations constitute a powerful tool for helping children to link words and social stimuli into coherent schemata that in turn can promote the development of social cognition.

Of note is not only the result that the accuracy of mental-state attributions mediates the effect of mental-state conversations on ToM development but also the finding that the frequency of using the mental-state lexicon did not. This result at first sight appears unexpected given that previous literature on preschoolers consistently found associations between child mental-state talk and ToM scores both concurrently (e.g., Ensor & Hughes, 2008; Hughes & Dunn, 1997) and over time (Hughes & Dunn, 1998). The lack of a significant indirect effect via the mental-state lexicon is also surprising if we consider that the effect of adults' (mainly mothers') mental-state conversations on preschooler's ToM development is traditionally explained by an increase in both the capacity of children to reason in the mental domain and children's attention toward mental phenomena (Slaughter & Peterson, 2012; Thompson, 2006). However, these are theoretical explanations and not the results of empirical research. Indeed, our study is, to our knowledge, the first that directly tests the effect of these two mediator mechanisms in children.

Generally speaking, we believe that our results are interesting in that they speak to the changes in mechanisms of ToM development. Considering existing findings, it seems that simply gaining a greater awareness of inner states has a crucial role to play during the early steps of ToM acquisition, when children are in preschool, whereas its importance may decrease as children proceed through middle childhood. Children's attention to mental states is likely to be a driving factor during early development when the challenge facing children is to become able to separate the mental and physical worlds and to learn the rules that govern links between these domains (Wellman & Liu, 2004). Older children need to face different developmental goals. The social environment of children during the school years becomes increasingly complex, ambiguous, and relevant. Therefore, it requires the use of ToM skills in a flexible and context-sensitive way in order to make sense of what is happening and to be successful while interacting with others (Banerjee et al., 2011). In other words, even if we have just started to investigate the nature of more advanced ToM skills, the current study, together with recent literature (e.g., Apperly, 2012), suggests that what makes the difference for older children is the ability not simply to pay attention to people's mental states but also to put this awareness to use when considering the complexity of social scenarios.

Implications, limitations, and future research directions

Our main result is in line with data showing that the accuracy in ToM performance is far from perfect in the adult population even when people are willing to infer the thoughts and feelings of another person (Klein & Hodges, 2001). Our findings have a practical implication, suggesting that successful interactions are likely to depend on being able to make correct judgments about internal states; recognizing people as intentional agents is not a sufficient condition (Mitchell, 2006). Remarkably, it has been shown that although many high-functioning individuals with autism make use of mental states in conversations (Nuyts & Roeck, 1997) and are able to use an adequate number of mental-state terms when responding to the Strange Stories (Happé, 1994), they have difficulties in making accurate context-sensitive mental inferences (Abell et al., 2000; Mitchell, 2006) and in using the appropriate mental-state terms to describe behavior (Happé, 1994). Our results make sense of this discrepancy and lead us to hypothesize that social–pragmatic difficulties during on-line interactions could be explained by deficits in mental-state inference processes, as proposed for individuals with autism (Dennis, Lazenby, & Lockyer, 2001; Roeyers, Buysse, Ponnet, & Pichal, 2001). It will be important for future research to further examine this hypothesis with typically developing children and to deeply understand the effect of individual differences in the ability to select and integrate relevant social information in on-line interactions.

In the interpretation of our results, caution is also warranted because we focused on (only) two potential mediators. Thus, we cannot exclude that other developmental processes occurred. For example, it could be that conversations about the mind are useful also because they communicate common norms and rules about social circumstances, so that the transmission of shared social scripts within the same culture becomes possible (Nelson, 2005). Clearly, a great deal of future research is needed in order to unfold the specific processes that occur during developmental maturation. The inclusion of microgenetic assessments in future studies, for example, would be particularly helpful to understand the rate and breadth of changes as well as the presence of individual differences in the developmental trajectories (Siegler, 1995). Similarly, the current study did not explore the role of potential moderating factors that might strengthen or reduce the impact of mental-state conversational input on mature ToM development. One obvious candidate here is the identity/nature of the conversational partners (e.g., peer vs. adult, dyad vs. group) and the quality of the relationship with them. Finally, it will be important for future research to extend our knowledge about how and to what extent other variables play a role in the impact of ToM training programs. For example, although our study design included controls for several cognitive variables known to be associated with ToM development, a variety of social–contextual variables beyond family affluence (e.g., parental education, mental-state talk in the family context, peer relationships) are also likely to be relevant and, therefore, need close attention.

Conclusions

Our findings indicate, through experimental manipulation, that conversations about the mind help in the transition toward a more accurate use of (rather than simply a propensity to use) ToM skills, as measured by the accuracy of mental-state attributions index. In showing this, the current study provides critical insights into the mechanisms of later ToM development and, we hope, will open a new wave of research aimed at revealing the nuances and details of changes in sociocognitive understanding through the school years.

Acknowledgments

This research was supported by a scholarship from the University of Pavia to Federica Bianco. We thank the children who took part in this study, their parents, and school staff members for their interest in and support of this project.

Appendix A

Scoring examples for the Strange Stories task (White et al., 2009)

Double Bluff Story: During the war, the Red army captures a member of the Blue army. They (Red army members) want him to tell them where his army's tanks are; they know they are either by the sea or in the mountains. They know that the prisoner will not want to tell them; he will want to save his army, and so he will certainly lie to them. The prisoner is very brave and very clever; he will not let them find his army's tanks. The tanks are really in the mountains. Now when the other side asks him where his tanks are, he says, "They are in the mountains."

Question: "Why did the prisoner say that?"

Sample Answer 1: "Because the Red army probably thinks the prisoner is scared."

Score of Strange Stories index: 0 (from the scoring guidelines, reference to motivation that misses the point of double bluff).

Score of accuracy of mental-state attribution index: 1 (accurate context-sensitive attribution of mental states, taking into account the mental states of all the characters).

Score of mental-state lexicon index: 2.

Sample Answer 2: "Because the prisoner is scared."

Score of Strange Stories index: 0 (from the scoring guidelines, reference to motivation that misses the point of double bluff).

Score of accuracy of mental-state attribution index: 0 (the mental state of just one character is taken into account).

Score of mental-state lexicon index: 1.

Appendix B

Script of training procedures for each condition. Researchers' comments are given in Italics.

(A) ToM condition: Example of misunderstanding				
Story	Questions	Feedback example for each question	Conversation onset provided by experimenter	Language exercise
It is evening time, and Robin is taking a bag of rubbish outside to put in the bin. Suddenly, he sees his neighbor's cat running away. He thinks to run after it in order to return it to his old neighbor. Luckily, he manages to catch the cat. At that moment, his neighbor opens the door and glimpses her cat struggling in the arms of a boy. She has left her glasses in the dining room, so she can't see well. She starts to shout, "Help me! Stop the cat thief!"	–What was Robin intending to do?	–*Right!/No, actually Robin didn't want to steal the cat. He wanted to do a good deed. He wanted to return it to the neighbor*	*Right! If Robin did what you said, the woman would stop shouting. She would recognize Robin and understand his good intentions about returning her cat. So she would change her point of view. Indeed, people's beliefs can change, for example, when people understand that their ideas are wrong or that they have not got enough information in order to understand well. People can act or say things in order to change other people's wrong beliefs. In this way, they can solve misunderstandings, just as Robin would do if he made his neighbor recognize him and he explained to her that he wanted to return the cat*	What is in your opinion the meaning of this sentence in the story? he *thinks* to run after it":
	–Why does the neighbor start to shout, "Help me! Stop the cat thief!"?	–*Well done!/No, actually she shouts because she thinks he wants to steal her cat. She has misunderstood the situation. She has not understood Robin's good intentions*	*Imagine a misunderstanding episode similar to Robin's story. Describe it, explaining what you would do in that situation in order to solve it*	–he *imagines* to run after it

Appendix B (*continued*)

(A) ToM condition: Example of misunderstanding Story	Questions	Feedback example for each question	Conversation onset provided by experimenter	Language exercise
	–What does Robin think about the neighbor's behavior?	*–Right!/No, actually Robin imagines that the old woman didn't understand what had happened because she forgot her glasses*		–he *decides* to run after it*
	–Can Robin say or do something in order to stop his neighbor's shouts? What? If he does this, why would the old woman stop shouting?	*–You're right!/No, actually he could make his neighbor recognize him and explain that the cat was escaping and that he thought to run after it in order to return it. By doing this, the elderly woman would modify her point of view. She would realize her misunderstanding, and she would understand what really happened*		–he *believes* to run after it –he *understands* to run after it
(B) Control condition Story	**Questions**	**Feedback example for each question**	**Conversation onset provided by experimenter**	**Language exercise**
Today is Friday. Teachers are giving some homework to their students for the weekend. Sarah always writes all her homework in her diary. During English class, the teacher gives homework to the children,	–What does Sarah do when teachers give homework?	*–Okay/No, actually she writes the homework down in her diary*	*Sometimes people are very busy, and they can accidentally leave an object somewhere. Later, when they need that object, they can't find it. So, they can look for it starting in the last place they used it*	What is in your opinion the meaning of this sentence in the story? "teachers *give* homework":

(*continued on next page*)

(B) Control condition

Story	Questions	Feedback example for each question	Conversation onset provided by experimenter	Language exercise
and Sarah writes it in her diary. Then, students go to the Art room. There, the teacher says, “Write down that you must draw a picture of your family with crayons by next Monday.”	–Are Sarah and her classmates always in the same place? If no, where do they go? (List the places in the right order.)	–*Right answer!/No, actually they went to different places. They were in the English classroom first, then they went to the Art room, then to the Gym, and finally to the Computer Lab*	*Have you ever lost an object? How did you go about finding it? Tell your classmates about it*	–teachers *do* homework
After Art class, they go to the Gym and then to the Computer Lab. When the	–In which places did Sarah use the diary?	–*Well done!/No, actually Sarah used the diary in the English classroom and in the Art room*		–teachers *set* homework*
computer class teacher gives homework, Sarah can’t write it down because she can’t find her diary. So, she tells the teacher. The teacher looks for the diary in the Art room and in the Gym but not in the English classroom	–Why does the teacher look for the diary in the Art room and in the Gym but not in the English classroom?	–*You’re right! The teacher doesn’t look for the diary in the English classroom because Sarah didn’t leave her diary there. She used it in the next place she went, that is, the Art room. She left her diary in the Art room or in the Gym*		–teachers *check* homework –teachers *have* homework

Source: “Promoting theory of mind in middle childhood: A training program” (Lecce, Bianco, Devine et al., 2014, *Journal of Experimental Child Psychology*, *126*, p. 63).
Note: An asterisk (*) indicates the correct choice.

References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*, 1–16.
Apperly, I. A. (2011). *Mindreaders: The cognitive basis of "theory of mind"*. New York: Psychology Press.
Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes, and individual differences. *Quarterly Journal of Experimental Psychology, 65*, 825–839.
Appleton, M., & Reddy, V. (1996). Teaching three-year-olds to pass false belief tests: A conversational approach. *Social Development, 5*, 275–291.
Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development, 82*, 1887–1905.
Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology, 49*, 1615–1627.
Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17*, 2117–2121.
Camaioni, L., Longobardi, E., & Bellagamba, F. (1998). Evoluzione dei termini di stati mentali nelle storie di fantasia scritte da bambini in età scolare [The development of terms of mental states in fantasy stories written by school-aged children]. *Età Evolutiva, 60*, 20–29.
Campbell, R., Lawrence, K., Mandy, W., Mitra, C., Jeyakuma, L., & Skuse, D. (2006). Meanings in motion and faces: Developmental associations between the processing of intention from geometrical animations and gaze detection accuracy. *Development and Psychopathology, 18*, 99–118.
Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology, 48*, 257–270.
Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*, 79–151.
Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage, 12*, 314–325.
Cervantes, C. A., & Callanan, M. A. (1998). Labels and explanations in mother–child emotion talk: Age and gender differentiation. *Developmental Psychology, 34*, 88–98.
Clements, W., Rustin, C. L., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: A training study of false belief. *Developmental Science, 3*, 81–92.
Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum.
Cornoldi, C., & Colpo, G. (1998). *Prove di lettura MT per la scuola elementare 2 [MT tests of reading for Primary School 2]*. Firenze, Italy: Organizzazioni Speciali.
Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behavior in School-Aged Children (HBSC) family affluence scale. *Social Science and Medicine, 66*, 1429–1436.
Cutting, A., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development, 70*, 853–865.
de Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? *British Journal of Developmental Psychology, 24*, 7–37.
Dennis, M., Lazenby, A. L., & Lockyer, L. (2001). Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders, 31*, 47–54.
Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development, 84*, 989–1003.
Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development, 85*, 1777–1794.
Dunn, J., Brown, J., & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology, 27*, 448–455.
Ensor, R., Devine, R. T., Marks, A., & Hughes, C. (2014). Mothers' cognitive references to two-year-olds predict theory of mind at ages 6 and 10. *Child Development, 85*, 1222–1235.
Ensor, R., & Hughes, C. (2008). Content or connectedness? Mother–child talk and early social understanding. *Child Development, 79*, 201–216.
Garner, P. W., Carlson Jones, D., Gaddy, G., & Rennie, K. M. (1997). Low-income mothers' conversations about emotions and their children's emotional competence. *Social Development, 6*, 37–52.
Grice, H. P. (1957). Meaning. *Philosophical Review, 66*, 377–388. Retrieved from http://www.jstor.org/stable/2182440.
Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*, 346–359.
Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*, 129–154.
Harris, P. (1999). Acquiring the art of conversation: Children's developing conception of their conversation partner. In M. Bennett (Ed.), *Developmental psychology: Achievements and prospects* (pp. 89–105). London: Psychology Press.
Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford.
Hughes, C. (2011). *Social understanding and social lives: From toddlerhood through to the transition to school*. London: Psychology Press.

Hughes, C., & Dunn, J. (1997). "Pretend you didn't know": Preschoolers' talk about mental states in pretend play. *Cognitive Development, 12*, 381–403.
Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology, 34*, 1026–1037.
Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture? *Child Development, 76*, 356–370.
Hughes, C., Lecce, S., & Wilson, C. (2007). "Do you know what I want?" Preschoolers talk about desires, thoughts, and feelings in their conversations with sibs and friends. *Cognition and Emotion, 21*, 330–350.
Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 6*, 720–730.
Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas In Psychology, 20*, 163–198.
Lecce, S., & Hughes, C. (2015). Reading minds and stories: Theory of mind predicts reading in middle childhood. Manuscript in preparation.
Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014a). Training preschoolers on 1st-order false-belief understanding: Transfer on advanced ToM skills and metamemory. *Child Development, 85*, 2404–2418.
Lecce, S., Bianco, F., Devine, R., Hughes, C., & Banerjee, R. (2014b). Promoting theory of mind in middle childhood: A training program. *Journal of Experimental Child Psychology, 126*, 52–67.
Lecce, S., Zocchi, S., Pagnin, A., Palladino, P., & Taumoepeau, M. (2010). Reading minds: The relation between children's mental state knowledge and their metaknowledge about reading. *Child Development, 81*, 1876–1893.
Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74*, 1130–1144.
Longobardi, E., Piras, R., & Presaghi, F. (2008). Lessico psicologico nelle narrazioni dei bambini della scuola primaria [Mental-state lexicon in narrations of primary school children]. *Rivista di Psicolinguistica Applicata, 8*, 55–72. http://dx.doi.org/10.1400/94558.
MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. London: Routledge.
McClelland, G. H. (2000). Nasty data: Unruly, ill-mannered observations can ruin your analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 393–411). New York: Cambridge University Press.
Meins, E., Fernyhough, C., Johnson, F., & Lidstone, J. (2006). Mind-mindedness in children: Individual differences in internal-state talk in middle childhood. *British Journal of Developmental Psychology, 24*, 181–196.
Melot, A. N., & Angeard, N. (2003). Theory of mind: Is training contagious? *Developmental Science, 6*, 178–184.
Meristo, M., Falkman, K. W., Hjelmquist, E., Tedoldi, M., Surian, L., & Siegal, M. (2007). Language access and theory of mind reasoning: Evidence from deaf children in bilingual and oralist environments. *Developmental Psychology, 43*, 1156–1169.
Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin, 135*, 749–773.
Milligan, K. A., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*, 622–646.
Mitchell, J. P. (2006). Mentalizing and Marr: An information processing approach to the study of social cognition. *Brain Research, 1079*, 66–75.
Nelson, K. (2005). Language pathways into the community of minds. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 26–49). Oxford, UK: Oxford University Press.
Nuyts, J., & Roeck, A. D. (1997). Autism and meta-representation: The case of epistemic modality. *International Journal of Language & Communication Disorders, 2*, 113–137.
Ontai, L. A., & Thompson, R. A. (2008). Attachment, parent–child discourse and theory-of-mind development. *Social Development, 17*, 47–60.
Ornaghi, V., Brockmeier, J., & Grazzani, I. (2011). The role of language games in children's understanding of mental states: A training study. *Journal of Cognition and Development, 12*, 239–259.
Ornaghi, V., Brockmeier, J., & Grazzani, I. (2014). Enhancing social cognition by training children in emotion understanding: A primary school study. *Journal of Experimental Child Psychology, 119*, 26–39.
Orsini, A. (1997). *WISC-R: Contributo alla taratura italiana (WISC-R: Contribution to Italian norms)*. Firenze, Italy: Organizzazioni Speciali.
Peterson, C. C., & Slaughter, V. (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.
Roeyers, H., Buysse, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *Journal of Child Psychology and Psychiatry, 42*, 271–278.
Roth, P. L., & Switzer, F. I. (2002). Outliers and influential cases: Handling those discordant contaminated maverick rogues. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 297–309). Malden, MA: Blackwell.
Rubini, V., & Rossi, M. A. (1982). Analysis of the psycho-metric characteristics of the Primary Mental Abilities Tests (Level 4-6, 1962 revision) applied to a group of Italian subjects. Bollettino di Psicologia Applicata, 161–164, 87–105. (Available from PsycINFO, Ipswich, MA).
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.
Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 298*, 199–209.
Shatz, M., Diesendruck, G., Martinez-Beck, I., & Akar, D. (2003). The influence of language and socioeconomic status on children's understanding of false belief. *Developmental Psychology, 39*, 717–729.
Siegler, R. S. (1995). How does change occur? A microgenetic study of number conservation. *Cognitive Psychology, 28*, 225–273.
Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46*, 606–620.
Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67*, 2967–2988.

Slaughter, V., & Peterson, C. C. (2012). How conversational input shapes theory of mind development in infancy and early childhood. In M. Siegal & L. Surian (Eds.), *Access to language and cognitive development* (pp. 3–22). New York: Oxford University Press.
Sperber, D., & Wilson, D. (2002). Pragmatics, modularity, and mind-reading. *Mind & Language, 17*, 3–23.
Symons, D. K. (2004). Mental state discourse, theory of mind, and the internalization of self–other understanding. *Developmental Review, 24*, 159–188.
Tarling, R. (2008). *Statistical modelling for social researchers: Principles and practice*. New York: Routledge/Taylor & Francis Group.
Thompson, R. A. (2006). Conversation and developing understanding: Introduction to the special issue. *Merrill-Palmer Quarterly, 52*, 1–16.
Thurstone, L. L., & Thurstone, T. G. (1962). *Primary mental abilities*. Chicago: Science Research.
Tomasello, M. (2003). *Constructing a language: A use-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tomasuolo, E., Valeri, G., Di Renzo, A., Pasqualetti, P., & Volterra, V. (2013). Deaf children attending different school environments: Sign language abilities and theory of mind. *Journal of Deaf Studies and Deaf Education, 18*, 12–29.
Turnbull, W., & Carpendale, J. M. (1999). A social pragmatic model of talk: Implications for research on the development of children's social understanding. *Human Development, 42*, 328–355.
Vivanti, G., McCormick, C., Young, G. S., Abucayan, F., Hatt, N., Nadig, A., ... Rogers, S. J. (2011). Intact and impaired mechanisms of action understanding in autism. *Developmental Psychology, 47*, 841–856.
Wellman, H. D., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523–541.
Wellman, H. M., & Peterson, C. C. (2013). Deafness, thought bubbles, and theory-of-mind development. *Developmental Psychology, 49*, 2357–2367.
White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of theory of mind for adults with autism. *Autism Research, 4*, 149–154.
White, S., Hill, E., Happé, F. G. E., & Frith, U. (2009). Revisiting the Strange Stories: Revealing mentalizing impairments in autism. *Child Development, 80*, 1097–1117.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.