

Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood?

Serena LECCE¹, Luca RONCHI¹, Paola DEL SETTE¹, Luca BISCHETTI², and Valentina BAMBINI^{2*}

¹Department of Brain and Behavioural Sciences, University of Pavia, Italy and

²Center for Neurocognition, Epistemology and theoretical Syntax (NETS), University School for Advanced Studies IUSS Pavia, Italy *Corresponding author: Valentina Bambini, Center for Neurocognition, Epistemology and theoretical Syntax (NETS), School of Advanced Studies IUSS Pavia, Piazza della Vittoria 15, 27100 Pavia, Italy;

fax +39-0382375899; tel +39-0382375841; E-mail: valentina.bambini@iusspavia.it

This article has been published in a revised form in Journal of Child Language at <https://doi.org/10.1017/S030500091800048X>. This version is published under a Creative Commons CC-BY-NC-ND. No commercial re-distribution or re-use allowed. Derivative works cannot be distributed. © Cambridge University Press.

Abstract

We investigated the association between individual differences in metaphor understanding and Theory of Mind (ToM) in typically developing children. We distinguished between two types of metaphors and created a Physical and Mental Metaphors task, echoing a similar distinction for ToM. Nine-year-olds scored lower than older age-groups in ToM as well as in the interpretation of mental, but not physical, metaphors. Moreover, nine-year-olds (but not older children) who are better in ToM are also better in interpreting mental, but not physical, metaphors. This suggests that the link between metaphor and ToM is stronger when metaphorical interpretation involves mental aspects, and it is more evident in early rather than later childhood.

Keywords: pragmatics; experimental pragmatics; figurative language

Introduction

Learning to understand metaphors is a complex task requiring the ability to adjust the literal meaning of words and infer the speaker's meaning based on the communicative context (Carston, 2010). As such, metaphor is a typical test ground in developmental pragmatics (Grigoroglou & Papafragou, 2017; Noveck, Bianco, & Castry, 2001). Classic literature is mostly based on verbal explanation tasks, requiring children to paraphrase metaphorical sentences, and indicates that full-fledged metaphor comprehension skills are not reached until adolescence. However, more recent literature pointed out that simpler tasks, such as elicited repetition or object substitution pretense, might reveal rudimentary metaphorical abilities already in preschool children (Pouscoulous, 2014). Whereas some authors emphasized the role of task demands, other authors reported differences related to metaphor types. For instance, in a seminal study, Winner and

colleagues (Winner, Rosenstiel, & Gardner, 1976) compared cross-sensory metaphors (communicating an experience in one sensory modality by referring to another sensory modality; e.g., Her perfume was bright sunshine) and psychological–physical metaphors (communicating a psychological experience by appealing to the physical domain; e.g., The prison guard was a hard rock) in a sample of children from six to fourteen years. Comprehension was not accurate until the age of ten, and in general psychological–physical metaphors were more difficult than cross-sensory ones. Other studies too suggested a facilitation for metaphors based on perceptual aspects (Vosniadou, Ortony, Reynolds, & Wilson, 1984; Wang & Dowker, 2010). Surprisingly, the distinction between the sensory and psychological aspects of metaphors has never been considered with respect to Theory of Mind (ToM), while it seems evident that the two types

differently capitalize on the ability to attribute mental states: one case requires an inference about physical aspects of the metaphor's topic, whereas the other involves an inference about mental aspects of the metaphor's topic. This observation – i.e., that not all metaphors are equal when it comes to attribution of mental states – constitutes the starting point of this work.

The role of ToM in metaphor processing is a largely debated topic in literature on Autism Spectrum Disorder (Vulchanova, Saldaña, Chahboun, & Vulchanov, 2015; Whyte & Nelson, 2015). Earlier studies claimed that (first-order) ToM is necessary to support metaphor comprehension (Happé, 1993), whereas more recent studies emphasize the role of vocabulary abilities (Kalandadze, Norbury, Naerland, & Naess, 2018; Norbury, 2005). In contrast, very little has been done on typical development. Considering other non-literal phenomena, Caillies and colleagues (Caillies & Le Sourn-Bissaoui, 2008) found that ToM is to some extent a prerequisite for idiom comprehension in five-, six-, and seven-year-olds, since idioms with plausible literal meanings (e.g., break the ice) require some appearance/reality distinction. A more sophisticated ToM is often assumed to be responsible for irony (Filippova & Astington, 2008; Massaro, Valle, & Marchetti, 2013), although results are conflicting (Bosco & Gabbatore, 2017). For metaphor, however, despite a sizable literature on developmental trends, few studies have considered individual differences in ToM. Moreover, none has distinguished between metaphor types, which seems to us a useful strategy to unpack the broad findings of association between ToM and pragmatics in development (Matthews, Biney, & Abbot-Smith, 2018).

This study focused on the relationships between ToM and metaphor understanding in development by adopting an individual differences approach and by also considering different types of metaphors. The study has two main aims. First, we aimed at investigating similarities and differences in developmental changes of the two focus constructs – metaphor understanding and ToM – in children from nine to twelve. This age range was selected because it is crucial for the development of both ToM (Lecce, Bianco, Devine, & Hughes, 2017) and metaphoric understanding (Winner et al., 1976). Importantly, we distinguished (both in ToM and in metaphors) between items that require inference on mental attributes and those requiring inferences on physical attributes. Specifically, we used a set of materials including metaphors with preferred physical (e.g., Dancers are butterflies) and preferred mental interpretation (Daddy is a volcano), echoing the distinctions between physical and mental inferences tested in a classic ToM task such as the Strange Stories. We called this task the 'Physical and Mental Metaphors task'. Both accuracy and interpretation were measured, allowing for a fine-grained analysis focusing on the

mechanisms of metaphor interpretation rather than simply on the ability to understand it. We expected to observe a change in performance at the age of ten, both for ToM and for metaphor understanding, in line with previous literature, and that such change would be more evident for the interpretation of mental metaphors. Second, we aimed at assessing the relationship between individual differences in ToM and interpretation of physical vs. mental metaphors in different age-groups. Here, we expected significant associations between ToM and the interpretation of mental, but not physical, metaphors. In doing so, we controlled for children's receptive language and working memory, which are known to be related to both metaphors (Carriedo, Corral, Montoro, Herrero, Ballestrino, & Sebastián, 2016; Norbury, 2005) and ToM (Happé, 1995; Lecce et al., 2017).

Method

Participants

We recruited 217 participants, ranging in age from nine to twelve years. The sample was made up of four age-groups: 62 nine-year-olds (33 F, $M_{age} = 9;6$, $SD = 3$ months, age range = 9;0–9;12), 48 ten-year-olds (21 F, $M_{age} = 10;5$, $SD = 3$ months, age range = 10;0–10;12), 51 eleven-year-olds (26 F, $M_{age} = 11;5$, $SD = 3$ months, age range = 11;0–11;12), and 56 twelve-year-olds (20 F, $M_{age} = 12;4$, $SD = 3$ months, age range = 12;1–12;10). Participants were recruited through local elementary and secondary schools. All of the children were fluent Italian speakers, and none had a history of developmental delay or learning disorder. Most of the children came from affluent families. Specifically, when the Family Affluence Scale (FAS; see below for details; Currie, Molcho, Boyce, Holstein, Torsheim, & Richter, 2008) was used to categorize the sample, 84.90% of the sample was classed as 'high affluence' (range = 6–9), 14.70% as 'middle affluence' (range = 3–5), and just one child (0.5%) as 'low affluence' (range = 0–2). In terms of family structure, 21.10% of the children were singletons, 56.90% had one sibling, 17% had two siblings, 3.70% had three siblings, and 1.40% had four or more siblings.

Procedure

Parental written consent was obtained for all participants. Data collection took place at the beginning of the school year. Children completed all the tasks individually at school during class time.

Measures

Socio-economic status

This was measured with the Family Affluence Scale (Currie et al., 2008), a short questionnaire with questions on family wealth:

family car ownership (range = 0–2), participants having/not having their own unshared room (range = 0–1), number of computers at home (range = 0–3), and number of times the participants went on a holiday during the past year (range = 0–3). Responses were summed into an overall index (range = 0–9).

Verbal ability

This was measured with the Italian version of the Vocabulary subtest of the Primary Mental Abilities (PMA) (Rubini & Rossi, 1982), requiring children to find the synonyms of 50 target words choosing among four alternatives. A time limit of 6 min was set. Total scores ranged from 0 to 50.

Working memory

This was measured with the Backward Digit Span task from the Italian version of the WISC-R (Orsini, 1997). Children listened to a series of digit sequences and were asked to recall them in reverse order. Children completed seven sequences of digits. The number of digits in each sequence increased from two to eight. Each sequence was marked as a pass (1) or fail (0), and total scores ranged from 0 to 7.

Theory of Mind

This was evaluated with the Strange Stories task (Happé, 1994), which consists of short stories depicting social situations followed by an open-ended question that requires participants to explain a character's behavior. We administered five mental stories, two involving double bluff, two involving misunderstanding, and one involving persuasion, together with three physical/control stories. 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, Hill, Happé, & Frith, 2009), we rated children's answers on a 3-point scale: 0 for an incorrect answer, 1 for a partially correct and implicit answer, and 2 for a full and explicit answer. Total scores ranged from 0 to 10 points for mental stories and from 0 to 6 for physical stories. A second rater independently coded 25% of the responses and inter-rater agreement was established using Cohen's kappa ($k = .85$).

Physical and Mental Metaphors task

We constructed a set of metaphors in Italian, following the prototypical 'X is Y' structure, where X is the topic and Y the vehicle of the metaphor, and distinguishing between two types of metaphor: physical and mental. In the physical metaphors, persons (X) are associated with non-human entities (Y) based on physical features; in the mental metaphors, persons (X) are associated with non-human entities (Y) based on psychological features. Thus, in order to understand physical metaphors,

inferences on appearance features or behaviors of the topic are needed, whereas, in order to understand mental metaphors, inferences on mental states of the topic are needed. The physical metaphors were: (1) Dancers are butterflies, (2) Climbers are squirrels, and (3) Players are elephants. The mental metaphors were: (4) Soldiers are lions, (5) Daddy is a volcano, and (6) The teacher is an icicle. Metaphors 1, 2, 3, and 4 were taken from Bambini, Ghio, Moro, and Schumacher (2013); Metaphor 5 was adapted from Pinto, Melogno, and Iliceto (2006).

We selected the metaphors in such a way that they were easy to understand for children, and balanced for lexical frequency and familiarity, differing only in the preferred interpretation (either physical or mental). To verify this, a number of measures were collected, following the standards for the construction of metaphor datasets (Bambini, Resta, & Grimaldi, 2014). See Table 1 for collected values and statistics. All target words were included in a frequency dictionary of child language (Marconi, Ott, Pesenti, Ratti, & Tavella, 1994), with no differences in lexical frequency between the target words in the physical and in the mental sets. Familiarity was checked with an online rating task, by means of a 7-point Likert scale, administered to 52 young adults. Participants were required to answer the question “How familiar is this metaphor?”, selecting a value from 1 (“not familiar at

Table 1. Measures Collected on the Materials Included in the Physical and Mental Metaphors Task

		All materials	Physical Set	Mental Set	Statistics
Frequency		74.38 (54.77)	85.83 (10.42)	62.92 (83.64)	$t(2.021) = 1.15, p = .37$
Familiarity (39 F, Mage = 23.88, SD = 2.35)		3.79 (0.87)	3.60 (1.30)	3.99 (0.34)	$t(4) = -0.50, p = .64$
Categorical task (23 F, Mage = 23, SD = 1.39)	Agreement on physical interpretation	–	97.85% (93.55–100)	2.15% (0–6.45)	$\chi^2 = 85.17, df = 1, p < .001$
	Agreement on mental interpretation	–	6.45% (0–12.9)	93.55% (87.1–100)	$\chi^2 = 70.55, df = 1, p < .001$
Parametric task (39 F, Mage = 23.88, SD = 2.35)	Physical attributes	4.47 (1.27)	5.43 (0.89)	3.51 (0.68)	$t(4) = 2.96, p < .05$
	Mental attributes	3.31 (1.92)	1.57 (0.14)	5.04 (0.33)	$t(4) = -16.76, p < .001$
	Difference score	–	3.86 (0.99)	1.53 (0.81)	$t(4) = 3.15, p < .05$

Notes. Frequency measures were collected using the Lessico Elementare (Marconi et al., 1994), a frequency dictionary of school-age children's language including 6095 lemmas, whereas familiarity scores (on a 7-point scale), scores in the categorical task (physical vs. mental interpretation), and scores in the parametric task (physical and mental attributes, on a 7-point scale) were collected from samples of young adults. For frequency, familiarity, and the parametric task scores, the Table reports mean and SD in parenthesis. For the categorical task scores, the Table reports mean agreement and range in parenthesis. For the statistics on frequency, we employed the Welch Two Sample t-test on log-transformed values. The chi-squared analysis on categorical scores was also run on each metaphor, being highly significant in all cases (metaphor 1, $\chi^2 = 31, df = 1, p < .001$; metaphor 2, $\chi^2 = 31, df = 1, p < .001$; metaphor 3, $\chi^2 = 23.52, df = 1, p < .001$; metaphor 4, $\chi^2 = 17.065, df = 1, p < .001$; metaphor 5, $\chi^2 = 23.5, df = 1, p < .001$; metaphor 6 $\chi^2 = 31, df = 1, p < .001$)

all”) to 7 (“totally familiar”), and trying to use all values in the scale. Familiarity was explained to participants in the following way: “Familiarity judgment expresses how much the metaphor sounds familiar to you, that is how much it seems to you that you have heard or used the metaphor in the presented form or in a similar form.” Obtained scores were around medium values (i.e., around the middle of the scale), indicating that the selected metaphors are not lexicalized, with no differences across sets. The distinction between physical and mental metaphors was checked through two tasks: (i) categorical and (ii) parametric. (i) Categorical task: 31 young adults blind to the hypothesis of the study were asked to classify each metaphor as referring either to physical or to mental attributes. For metaphors in the physical set, agreement on physical interpretation was >95% and for metaphors in the mental set, agreement on mental interpretation was >90%, with highly significant chi-square for each set and for each metaphor. (ii) Parametric task: 52 young adults blind to the hypothesis of the study were asked to rate each metaphor for physical attributes (“How much does this metaphor express physical attributes?”) and for mental attributes (“How much does this metaphor express mental attributes?”) with an online task and by means of a 7-point Likert scale. Metaphors in the physical set scored significantly higher on physical attributes than metaphors in the mental set. Metaphors in the mental set scored significantly higher on mental attributes than metaphors in the physical set. Interestingly, within each set, when considering the difference between the two interpretations (the absolute value of the difference between the mean score on the physical attributes and the mean score on the mental attributes), the value was significantly higher for metaphors in the physical compared with the mental set. Overall, these data support the distinction between the physical and the mental sets, and also indicate that metaphors in the physical set tended to have a straightforwardly physical interpretation, while metaphors in the mental set were also open to physical interpretations.

Children were required to explain the meaning conveyed by each metaphor, after an example item considered together with the examiner. Questions and answers were oral. Children’s answers were coded according to the level of (i) accuracy (defined as the ability to articulate the link between the topic and the vehicle) and (ii) interpretation (either physical or mental). Distinguishing these two levels – often blended in previous studies – allowed us to make a fine-grained analysis on interpretation rather than simply on the ability to understand a metaphor. Specifically, this scoring allowed us to take into account cases in which children gave plausible, yet physical interpretations of metaphors in the mental set, and vice versa. Accuracy was coded on a three-step scale following previous studies in the field of figurative language comprehension (Melogno, Pinto, & Di Filippo, 2017;

Vulchanova, Vulchanov, & Stankova, 2011): Incorrect responses scored 0, correct metaphorical responses scored 2, and responses where the participant had attempted a metaphorical albeit somewhat incomplete interpretation received a score of 1. More specifically, a 0 score was assigned when the child: (a) declares he/she does not know/understand, (b) interprets the metaphor literally (e.g., for the metaphor "Soldiers are lions", "They hunt well"), or (c) gives an incorrect response ("They are careful"). A 1 score was assigned when children's answer: (a) is incomplete ("They are very good") or (b) refers to a non-salient feature of the metaphor vehicle ("They run as fast as lions"). A 2 score was assigned when children's answer is complete and refers to the salient features of the metaphor vehicle ("They are strong"; "They are courageous"). Accuracy total scores ranged from 0 to 6 points for both the physical and the mental metaphor sets. Interpretation was coded on a two-step scale. A 0 score was given if children's answer referred to physical attributes ("They are strong") or actions ("They run fast") of the topic. A 1 score was given if children's answer referred to psychological attributes of the topic ("They are courageous"; "They are aggressive"). Interpretation total scores ranged from 0 to 3 points for both the physical and the mental metaphor sets. A second rater independently coded 25% of the responses and inter-rater agreement was established using Cohen's kappa and $k = .96$ for accuracy and $k = .88$ for interpretation; for the mental set, $k = .98$ for accuracy and $k = .94$ for interpretation).

Data analysis

We performed three sets of analyses. The first set was intended as a validation of the materials in the metaphor task, to ensure high accuracy for all items and different interpretations across the physical and the mental sets. The second set included a series of ANOVAs to assess developmental changes for each study-variable. The scores on control (verbal ability and working memory) and focus variables (ToM and metaphor interpretation for the physical and the mental sets) were submitted to an ANOVA with Age-Group (four levels) as between-subject variable. The third set of analyses comprised correlations and partial correlations (controlling for verbal ability, working memory, and socioeconomic status) between ToM and metaphor interpretation for the physical and the mental sets.

Results

Metaphor understanding and interpretation

We first checked the level of accuracy in the children's answers in the Physical and Mental Metaphors task. As expected, the great majority

of children scored high on this index, with 98.2% and 92.2% of children showing a performance that was equal or bigger than 3 on the total score for the physical and the mental metaphor set, respectively. This result allowed us to focus on interpretation, rather than accuracy, in the main analyses. As expected, children scored higher in interpretations of metaphors in the mental set (total score = 1.92, SD = .77) than in the physical set (total score = 0.07, SD = .26), that is they were more likely to refer to mental states when interpreting a mental metaphor than a physical one ($t(216) = 32.88$, $p = .000$). The correlation between interpretation scores for physical and mental metaphors was not significant ($r(217) = -0.04$, $p = .56$), as a further indication that the type of interpretation was different for each set.

Developmental effects

Descriptive analyses for all age-groups are reported in [Table 2](#). The ANOVA showed a main effect on verbal ability² ($F(3,213) = 23.98$, $p < .001$, $\eta_p = .25$). Post-hoc Bonferroni comparisons showed that verbal ability of nine-year-olds was significantly lower than that of each of the other age-groups, and that verbal ability of ten-year-olds was significantly lower than that of twelve-year-olds. No other ²contrast reached significance. The ANOVA also showed a significant effect on working memory ($F(3,213) = 5.98$, $p = .001$, $\eta_p = .08$). Post-hoc Bonferroni comparisons showed that the working memory of nine-year-olds was significantly lower than that eleven- and twelve-year-olds. No other contrast reached significance.

Table 2. Descriptive Analyses for All Study Variables

		Age 9	Age 10	Age 11	Age 12
Control variables	Verbal ability	15.56 (5.49)	21.19 (6.39)	22.67 (7.44)	25.32 (6.64)
	Working memory	2.52 (0.83)	2.92 (0.99)	3.12 (0.93)	3.11 (0.80)
Physical and Mental Metaphors task	Physical set – accuracy	4.45 (1.36)	4.60 (1.18)	4.24 (1.41)	4.59 (1.33)
	Physical set – interpretation	0.03 (0.18)	0.08 (0.28)	0.14 (0.35)	0.5 (0.23)
	Mental set – accuracy	4.98 (1.17)	5.31 (1.03)	5.16 (0.92)	5.25 (1.0)
	Mental set – interpretation	1.68 (0.76)	1.90 (0.75)	2.08 (0.72)	2.04 (0.84)
ToM task	Mental Strange Stories	6.39 (1.37)	7.15 (1.66)	7.53 (1.47)	7.44 (1.48)
	Physical Strange Stories	3.06 (1.30)	3.19 (1.50)	3.63 (1.67)	3.66 (1.46)

Note. For each task, mean total scores and standard deviations in parenthesis.

Regarding ToM, we found a main effect of mental (F(3,213) = 7.09, $p < .001$, $\eta_p^2 = .09$), but not control Strange Stories (F(3,213) = 2.36, $p = .07$, $\eta_p^2 = .03$). Bonferroni comparisons showed that ToM of nine-year-olds was significantly lower than that of each of the other age-groups. No other contrast reached significance.

Regarding metaphor interpretation, our analyses showed a main effect on mental (F(3,213) = 3.59, $p = .01$, $\eta^2 = .04$) but not physical metaphors (F(3,213) = 1.67, $p = .18$, $\eta^2 = .02$). Post-hoc analyses showed that the mental interpretation of nine-year-olds was significantly lower than that of eleven- and twelve-year-olds. No other contrast reached significance. Here it is also important to note that, in each age-group, accuracy in the mental set was actually significantly higher than accuracy in the physical set ($t \geq 2.66$, $p = .01$).

Associations between ToM and metaphors

Results of the correlational analysis on each age-group between individual differences in interpreting mental vs. physical metaphors and the Strange Stories task are reported in Table 3. Individual differences in interpreting mental metaphors were significantly correlated with those in mental stories of the Strange Stories in nine-year-olds ($r(62) = 0.37$, $p = .003$), but not in any other age-group. This association (interpretation of mental metaphors with mental Strange Stories) remained significant even when we controlled for verbal ability, working memory, and socioeconomic status ($r(57) = 0.34$, $p = .008$). Importantly, the association remained significant also when we controlled for the above factor plus accuracy in the interpretation of mental metaphors ($r(56) = 0.32$, $p = .01$). Table 3 also shows that in nine-year-olds and in eleven-year-olds there was a significant association between the interpretation of mental metaphors and control stories of the Strange Stories task; however, the value of these correlations fell below significance when we controlled for verbal ability, working memory, and socioeconomic status ($r(57) = 0.27$, $p = .06$, and $r(57) = 0.23$, $p = .12$, respectively). Overall, in nine-year-olds, the association between metaphor interpretation and Strange Stories was specific for mental items and did not extend to the physical/control ones. There was no significant correlation between the interpretation of mental metaphors and control stories of the Strange Stories task and, on the same lines, no significant association between the interpretation of physical metaphors and mental stories of the Strange Stories task. No significant correlation between ToM and metaphor interpretation was found in any other age-groups.

Discussion

This study investigated the relationships between ToM and metaphor understanding by distinguishing between two types of metaphors in the Physical and Mental Metaphors task. The first aim was to explore developmental changes in metaphor comprehension and in ToM across middle childhood. Although all children scored high in accuracy, our findings showed that nine-year-olds performed lower than the other age-groups (who did not differ significantly from one another) in the interpretation of mental, but not physical metaphors. This confirms our hypothesis and is in line with previous evidence suggesting that metaphor comprehension starts to be accurate at about ten years of age (Winner et al., 1976). Interestingly, idiom comprehension also starts approximating adult-like knowledge around age ten (Kempler, VanLancker, Marchman, & Bates, 1999; Vulchanova et al., 2011), indicating that this stage might

Table 3. Correlations between ToM and Physical and Mental Metaphors task

		Metaphor Physical set – interpretation	Physical Strange Stories	Mental Strange Stories
Metaphor Mental set – interpretation	Age 9	0.08 (0.04)	0.29* (0.26)	0.37** (0.34*)
	Age 10	-0.05 (0.03)	0.25 (0.10)	0.27 (0.19)
	Age 11	-0.21 (-0.13)	0.29* (0.23)	0.24 (0.14)
	Age 12	0.02 (0.04)	0.18 (0.08)	0.13 (0.09)
Metaphor Physical set – interpretation	Age 9		0.06 (0.02)	0.02 (-0.10)
	Age 10		-0.09 (-0.14)	-0.03 (-0.03)
	Age 11		-0.04 (0.01)	-0.15 (-0.07)
	Age 12		0.01 (0.03)	0.21 (0.27)
Physical Strange Stories	Age 9			0.25* (0.20)
	Age 10			0.39** (0.22)
	Age 11			0.31* (0.31*)
	Age 12			0.26* (0.09)

Notes. Partial correlations controlling for verbal ability, working memory, and socioeconomic status are given in parentheses. * $p < .05$; ** $p < .01$ (all significance tests are two-tailed).

be critical for different forms of figurative language. Moreover, we observed a similar result for ToM development, where nine-year-olds performed significantly lower than the other age-groups. Again, this result is in line with existing literature showing that the development of ToM, measured via the Strange Stories, does not follow a linear pattern of change across middle childhood (Lecce et al., 2017; Lecce, Zocchi, Pagnin, Palladino, & Taumoepeau, 2010). The important and novel finding here is the observation of a parallel in developmental timing, with the age of ten marking a significant difference with respect to the age of nine both in ToM and in the interpretation of mental metaphors. This parallel in development suggests that there may well be a relationship between ToM and the interpretation of the psychological attributes of metaphors.

This consideration brings us to the individual differences issue, and the second main result of this study. Independently of verbal ability and working memory, nine-year-olds (but not older children) who are better in ToM are also more likely to better understand mental, but not physical, metaphors. This suggests that: (i) the association between metaphor comprehension and ToM is not general, but rather specific for those metaphors whose interpretation requires an inference on mental states, and (ii) the association between interpreting mental metaphors and ToM changes across development, being stronger in earlier developmental phases. We believe that these findings have important implications on the relation between metaphor and ToM and, more generally, on the relation between pragmatics and ToM. About (i), our results support the theoretical distinction between physical and mental metaphors, and suggest that not all metaphors are interpreted through the same strategy, i.e., not all metaphors recruit ToM skills to the same degree. The level of the inference to be drawn (e.g., physical vs. mental) might indeed play a role in modulating the involvement of the ToM system. Although we acknowledge that pragmatic interpretation always requires a certain degree of attribution of intentions (Sperber & Wilson, 2002), our data suggest that this might be a basic level of the communication process, beyond which ToM involvement might vary depending on the metaphor's type. Metaphor capitalizes on the context-based enrichment of lexical meanings through promoting and elaborating some properties of the encoded concepts, while dropping others (Carston, 2010). When the relevant properties refer to mental aspects (as, e.g., in the case of *Daddy is a volcano*), the ToM load is likely to increase. A similar view is supported by a recent study on children with Autism Spectrum Disorder, where cognitive training differently affected psychological and sensory metaphor comprehension (Melogno et al., 2017). This suggests that the interpretation of the two types of metaphor requires, at least in part, different skills. Crucially, in everyday communication, metaphors often refer to emotions and mental states, whereby they illuminate

psychological experiences through concrete concepts (Kövecses, 2003). Studies on adults processing metaphors have often pointed to the involvement of ToM brain circuitry (Bambini, Gentili, Ricciardi, Bertinetto, & Pietrini, 2011), which – we argue – might be especially driven by the mental aspects of metaphorical interpretations. Future studies employing a larger number of stimuli than the sample used here, and possibly measuring brain correlates, might further elucidate how specific the involvement of ToM is for certain metaphors.

About (ii), i.e., a stronger relation between the interpretation of mental metaphor and ToM at an earlier age, our results offer a novel developmental angle to the debate over the relation between metaphor (and pragmatics) and ToM. A closer inspection to the literature on ToM and other non-literal language aspects reveals that evidence in favor of a relation between the two is usually based on data from young children (Caillies & Le Sourn-Bissaoui, 2008). Thus, a likely scenario is that the link between metaphor interpretation and ToM is stronger in earlier ages, but it becomes looser in later stages. A similar pattern can be found considering the association between ToM and language (receptive and expressive), which is strong in preschoolers (Milligan, Astington, & Dack, 2007) and seems to become weaker across middle childhood (Lecce et al., 2010).

An alternative interpretation might be that our results are due to the fact that all age-groups are good in physical metaphors, suggesting that they have mastered them, and it is only in the domain of mental metaphors that differences can be observed. However, this interpretation is unlikely for two reasons. First, our main analyses focused on interpretation rather than on accuracy, and thus our results cannot be simply explained by a difference in the level of difficulty between the two sets of metaphors. Second, when we checked differences in the level of accuracy between the two sets, we found that, in each age-group, the accuracy in the mental set was actually significantly higher than accuracy in the physical set. Hence, the explanation that the link between ToM and metaphor interpretation becomes weaker over development seems more likely.

More generally, our result is in accordance with Karmiloff-Smith's model of the progressive modularization and specialization of cognitive skills (Karmiloff-Smith, 1995). Interestingly, the literature on metaphor processing in adults is compatible with this view, as it does not offer compelling evidence of a strong relationship with ToM, while in contrast emphasizing the role of other components such as working memory and executive control (Chiappe & Chiappe, 2007; Columbus, Sheikh, Côté-Lecaldare, Häuser, Baum, & Titone, 2015). Very few studies explicitly tested the

relation between adults' performance in metaphor tasks and in ToM. In an eye-movement study, participants' reading patterns were shown to be influenced by cognitive factors but not by socio-cognitive skills such as those measured in the Emotional Stroop task (Olkonemi, Ranta, & Kaakinen, 2016). A study on the Reading the Mind in the Eyes test showed that individual differences in ToM were related to figurative language comprehension skills, but this relation was not independent from the general intelligence level (Peterson & Miller, 2012). It is possible that the relation between metaphor comprehension and ToM is stronger in neurodevelopmental disorders and other clinical conditions such as schizophrenia (Bosia et al., 2016; Champagne-Lavau & Stip, 2010; Happé, 1993), although the issue is debated (Andrés-Roqueta & Katsos, 2017; Norbury, 2005), but this does not seem to explain individual differences in the healthy population. To properly test this account, and the hypothesis that the link between metaphor interpretation and ToM progressively loosens from middle childhood into adult full-fledged competence, future studies involving younger and older child groups, and possibly adopting a lifespan perspective, are needed.

Our study has important limitations that should be acknowledged and that correspond to future research directions. The first regards the ToM measure used: the Strange Stories task. Here, it is important to note that, for time reasons, we have included only three physical/control stories, and this may have reduced the variability in the total score. Future studies with a higher number of items are therefore needed. In addition, as recently outlined by Bosco and colleagues, the Strange Stories task is a verbal task that may imply the ability to understand language in the given context (Bosco, Tirassa, & Gabbatore, 2018). Although we acknowledge that, generally speaking, there may be an overlap between ToM and pragmatics in the Strange Stories, we argue that this is not a major issue in the present study for at least two reasons. First, we selected a subset of Strange Stories items, avoiding the items more closely related to non-literal speech. Second, this overlap cannot explain our finding with respect to the change in the association between ToM and metaphor across ages. Future studies are certainly needed to explore the link between ToM and metaphor understanding using non-verbal ToM tasks, such as the Triangle task (Castelli, Happé, Frith, & Frith, 2000). The second limitation regards the use of a vocabulary task to assess children's language. The literature – especially on atypical development – offers much evidence on the impact of language comprehension abilities, including competence in structural aspects, on figurative language understanding (Gernsbacher & Pripas-Kapit, 2012; Vulchanova et al., 2015). Although the Physical and Mental Metaphors task was developed to be easy to

understand, future studies using more comprehensive measures of language (including grammar) are needed to disentangle the relationship between the development of linguistic competence and of metaphor understanding. The third limitation concerns our participants, who belong primarily to affluent families. Given that ToM development, at least in preschool years, is associated with family background (Slaughter & De Rosnay, 2017), future research on ToM and metaphors in low-income children would be interesting.

Perhaps the most important limitation of the present study lies in its design and in the type of information provided by the correlational analysis, which can suggest association but cannot support claims on causality between ToM and metaphor understanding. Although a correlation study can be viewed as a first necessary step when investigating a new topic, this approach leaves open the question of whether ToM is a prerequisite for metaphor, or for specific types of metaphor only. Future research should investigate more deeply the nature and the direction of the relationship between ToM and the understanding of mental vs. physical metaphors. One obvious candidate is to use a longitudinal study; such a design would indeed enable us to understand if ToM promotes the development of pragmatics in typically developing children or if, to the contrary, pragmatics predicts the development of ToM. A developmental study would also help discriminate between divergent theoretical positions such as pragmatic accounts of ToM (Westra & Carruthers, 2017) and more ToM oriented accounts of pragmatics (Sperber & Wilson, 2002).

Overall, our study contributes to the clarification of the relations between ToM and pragmatics in middle childhood, with two main strengths. The first is the distinction between two types of metaphor, physical and mental. These two types require inferences on two different levels, an aspect which has been little considered in the literature on metaphor. The second is the developmental angle of the study, which covers a wide age-range and showed not only that there is a parallel in developmental timing between the interpretation of the mental aspects of a metaphor and ToM, but also that the association between the two is stronger in earlier ages. Besides the theoretical implications, our work might also have practical ones. Non-literal language is a fragile aspect of children's pragmatic competence, and our work might help shaping training programs based on a careful consideration of the role of ToM for the understanding of certain linguistic expressions.

References

- Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with autistic spectrum disorders. *Frontiers in Psychology*, 8, 996.

doi.org/10.3389/fpsyg.2017.00996

- Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin*, 86(3/4), 203–16.
- Bambini, V., Ghio, M., Moro, A., & Schumacher, P. B. (2013). Differentiating among pragmatic uses of words through timed sensibility judgments. *Frontiers in Psychology*, 4, 938. doi.org/10.3389/fpsyg.2013.00938
- Bambini, V., Resta, D., & Grimaldi, M. (2014). A dataset of metaphors from the Italian literature: exploring psycholinguistic variables and the role of context. *PLoS ONE*, 9(9). doi.org/10.1371/journal.pone.0105634
- Bosco, F. M., & Gabbatore, I. (2017). Sincere, deceitful, and ironic communicative acts and the role of the theory of mind in childhood. *Frontiers in Psychology*, 8, 21. doi.org/10.3389/fpsyg.2017.00021
- Bosco, F. M., Tirassa, M., & Gabbatore, I. (2018). Why pragmatics and Theory of Mind do not (completely) overlap. *Frontiers in Psychology*, 9, 1453. doi.org/10.3389/fpsyg.2018.01453
- Bosia, M., Arcara, G., Buonocore, M., Bechi, M., Moro, A., Cavallaro, R., & Bambini, V. (2016). Communication in schizophrenia, between pragmatics, cognition and social cognition. In A. M. Di Sciullo (ed.), *Biolinguistic investigations on the language faculty* (pp. 213–34). Amsterdam: John Benjamins.
- Caillies, S., & Le Sourn-Bissaoui, S. (2008). Children's understanding of idioms and theory of mind development. *Developmental Science*, 11(5), 703–11.
- Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., Ballestrino, P., & Sebastián, I. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLoS ONE*, 11(3), 1–20. doi.org/10.1371/journal.pone.0150289
- Carston, R. (2010). Lexical pragmatics, ad hoc concepts and metaphor: a relevance theory perspective. *Italian Journal of Linguistics*, 22(1), 153–80.
- 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(3), 314–25.
- Champagne-Lavau, M., & Stip, E. (2010). Pragmatic and executive dysfunction in schizophrenia. *Journal of Neurolinguistics*, 23(3), 285–96.
- Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language*, 56(2), 172–88.
- Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: an eye movement study of sentence reading. *Frontiers in Human Neuroscience*, 8, 1057. doi.org/10.3389/fnhum.2014.01057
- Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: the development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine*, 66(6), 1429–36.
- Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, 79(1), 126–38.
- Gernsbacher, M. A., & Pripas-Kapit, S. R. (2012). Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. *Metaphor and Symbol*, 27(1), 93–105.
- Grigoroglou, M., & Papafragou, A. (2017). Acquisition of pragmatics. In *Oxford Research Encyclopedia of Linguistics* (pp. 1–39). Oxford: Oxford University Press.
- Happé, F. G. E. (1993). Communicative competence and theory of mind in

- autism: a test of relevance theory. *Cognition*, 48(2), 101–19.
- 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(2), 129–54.
- Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66(3), 843–55.
- Kalandadze, T., Norbury, C., Naerland, T., & Naess, K.-A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: a meta-analytic review. *Autism*, 22(2), 99–117.
- Karmiloff-Smith, A. (1995). *Beyond modularity: a developmental perspective on cognitive science*. Cambridge, MA: MIT Press.
- Kempler, D., VanLancker, D., Marchman, V., & Bates, E. (1999). Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology*, 15(3), 327–49.
- Kövecses, Z. (2003). *Metaphor and emotion: language, culture, and the body in human feeling*. Cambridge University Press.
- Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: a short-term longitudinal study. *Journal of Experimental Child Psychology*, 163, 69–86.
- 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(6), 1876–93.
- Marconi, L., Ott, M., Pesenti, E., Ratti, D., & Tavella, M. (1994). *Lessico Elementare. Dati statistici sull'italiano scritto e letto dai bambini delle elementari*. Bologna: Zanichelli.
- Massaro, D., Valle, A., & Marchetti, A. (2013). Irony and second-order false belief in children: What changes when mothers rather than siblings speak? *European Journal of Developmental Psychology*, 10(3), 301–17.
- Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, 14(3), 186–223.
- Melogno, S., Pinto, M. A., & Di Filippo, G. (2017). Sensory and psychophysiological metaphor comprehension in children with ASD: a preliminary study on the outcomes of a treatment. *Brain Sciences*, 7(7), 85. doi.org/10.3390/brainsci7070085
- Milligan, K., 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(2), 622–46.
- Norbury, C. F. (2005). The relationship between theory of mind and metaphor: evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, 23(3), 383–99.
- Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, 16(1), 109–21.
- Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(3), 433–50.
- Orsini, A. (1997). *WISC-R. Contributo alla taratura italiana [Contribution to Italian norms]*. Firenze: Organizzazioni Speciali.
- Peterson, E., & Miller, S. F. (2012). The eyes test as a measure of individual differences: How much of the variance reflects verbal IQ? *Frontiers in Psychology*, 3, 220. doi.org/10.3389/fpsyg.2012.00220
- Pinto, M. A., Melogno, S., & Iliceto, P. (2006). Test di comprensione delle

- metafore. Roma: Carocci. Pouscoulous, N. (2014). 'The elevator's buttocks': metaphorical abilities in children. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 239–60). Amsterdam/Philadelphia: Benjamins.
- Rubini, V., & Rossi, M. A. (1982). Analysis of the psycho-metric characteristics of the Primary Mental Abilities Tests (Level 4-sup-6, 1962 revision) applied to a group of Italian subjects *Bollettino Di Psicologia Applicata*, 161–164, 87–105.
- Slaughter, V., & De Rosnay, M. (Eds.). (2017). *Theory of mind development in context*. London: Routledge.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language*, 17(1/2), 3–23.
- Vosniadou, S., Ortony, A., Reynolds, R. E., & Wilson, P. T. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development*, 55(4), 1588–606.
- Vulchanova, M., Saldaña, D., Chahboun, S., & Vulchanov, V. (2015). Figurative language processing in atypical populations: the ASD perspective. *Frontiers in Human Neuroscience*, 9, 24. doi.org/10.3389/fnhum.2015.00024
- Vulchanova, M., Vulchanov, V., & Stankova, M. (2011). Idiom comprehension in the first language: a developmental study. *Vigo International Journal of Applied Linguistics*, 8(1), 207–34.
- Wang, C., & Dowker, A. (2010). A cross-cultural study of metaphoric understanding. In G. Low, Z. Todd, A. Deignan, & L. Cameron (Eds.), *Researching and applying metaphor in the real world* (pp. 105–22). Amsterdam: Benjamins.
- Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, 158, 165–76.
- White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: revealing mentalizing impairments in Autism. *Child Development*, 80(4), 1097–117.
- Whyte, E. M., & Nelson, K. E. (2015). Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders*, 54, 2–14.
- Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12(4), 289–97.