Metaphorical Developing Minds: The role of multiple factors in the development of metaphor comprehension

Simona Di Paola [a, *], Filippo Domaneschi [b], Nausicaa Pouscoulous [c]

[a] Department of Educational Sciences, Psychology Unit, University of Genoa, Genoa, Italy

[b] DAFIST, University of Genoa, Italy

[c] Division of Psychology and Language Sciences, University College London, London, United Kingdom

Abstract

Metaphor understanding is traditionally thought to emerge late in childhood. Although recent findings suggest that even pre-schoolers can understand metaphor in more age appropriate paradigms, it is still unclear which skills scaffold the development of metaphorical abilities. Metaphor comprehension is a complex process relying on multiple cognitive abilities with different developmental paths, such as Alternative Naming (i.e., accepting two labels for the same referent) and Analogy Perception (i.e., detecting similarities across objects). These two abilities may prove crucial to explain the development of metaphoric competence because difficulties in one or both of them may impose additional demands linked to children's general cognitive development. This might contribute to slow down the development of metaphor understanding. This study aims at teasing apart the contribution of Alternative Naming and Analogy Perception in the development of metaphor understanding. Using a unified picture-matching paradigm, we tested 3- and 4-year-olds in three tasks: Metaphor Comprehension, Alternative Naming and Analogy Perception. Results reveal that children with better alternative naming and analogical abilities show a better performance in the metaphor task, suggesting that both Alternative Naming and Analogy Perception play a role in the development of metaphoric competence.

Keywords

Metaphor Comprehension; Pragmatic Development; Alternative Naming; Analogy Perception.

1. Introduction

In a metaphor such as *Babies are irresistible marshmallows,* the speaker communicates something about the *target* (e.g., babies) by means of the *vehicle* (e.g., marshmallows) in a non-literal way, drawing on an implicit relation between the two − i.e., the metaphor ground.

Classical studies suggest that metaphor understanding emerges late and that children interpret metaphor literally before age 8 (e.g., Billow, 1975; Cometa & Eson, 1978; Asch & Nerlove, 1960; see Winner 1988/1997 for a review). However, recently, a different trend has emerged: with more child-friendly experimental paradigms (e.g., by reducing metalinguistic demands; Pouscoulous 2011, 2014), metaphorical competence is visible already at age 3 (Pearson, 1990; Pouscoulous & Tomasello, 2019), starkly improves at 4 (Özçalişkan, 2005; Deamer, 2013) and progresses until 7 (Declercq et al., 2010). Overall, children's ability to grasp metaphors appears to be more precocious than traditionally thought, albeit having limitations. Difficulties persist especially in pre-school years, when children may interpret metaphor literally. Moreover, some evidence suggests that pre-schoolers exhibit greater difficulties with metaphor than with other non-literal meaning shifts such as hyperbole (Deamer, 2013) and metonymy (e.g., Falkum et al., 2017).

Where do children's difficulties stem from? Here, we investigate some of the sources of difficulty, which represents an essential missing piece in our understanding of figurative language development. Metaphor comprehension involves multiple cognitive skills and looking at the development of the abilities that scaffold metaphoric competence may help identify some of the sources of difficulty. We focus specifically on alternative naming and analogy perception.

1.1 Metaphor processing: multiple factors and sources of difficulty

Pragmatic theories offer several accounts of metaphor comprehension (see Gibbs & Colston, 2012 for a review) based on a variety of processes - from mapping of properties (Lakoff & Johnson, 1980/2003) to the online construction of occasion-specific concepts (Wilson & Carston, 2007). Despite this variety, all accounts presuppose two basic features. First, dual reference: in a metaphor, two labels can refer to one and the same target in the

same communicative context (e.g., a baby is a *baby* as well as a *marshmallow*). Second, the target and the vehicle generally belong to different domains (i.e., metaphor generally involves category violation). This calls for some mental operations on the properties of the vehicle for it to apply meaningfully to the target. Whatever the nature of this operation - e.g., mapping or broadening – it presupposes identifying relevant properties to recognise the metaphorical ground.

These two features involve two cognitive abilities fundamental for metaphor comprehension: alternative naming and analogy perception (see Pouscoulous, 2014; Rubio-Fernández & Grassmann, 2016). Alternative naming is the metalinguistic ability to assign a second label to an entity and accept that the same referent can take more than one label. Metaphor involves alternative naming since dual reference presupposes accepting two labels for the same entity within the same communicative context, a literal and a figurative one. Metaphor also involves analogy perception since deriving the metaphor ground presupposes detecting shared properties across entities. The cognitive substratum for this is analogical reasoning – i.e., the ability to transfer one mechanism from a known entity to another entity by identifying the relation that the two entities share in spite of their surface differences (i.e., the *relational similarity constraint*; Goswami, 2001). Theories of metaphor comprehension take different positions with respect to analogy. Two paradigmatic cases are Bowdle and Gentner (2005) and Relevance Theory (Wilson & Carston, 2007). In Bowdle & Gentner's (2005) view, metaphor is understood by mapping analogically the system of relations among the properties of the vehicle to the system of relations among the properties of the target. According to Wilson & Carston (2007), metaphor is a loose use of language and is interpreted by narrowing and broadening the lexically encoded concept of the vehicle to construct online an occasion-specific concept. While the role of analogical reasoning is core to Bowdle & Gentner's (2005) account, others (e.g., Wilson & Carston, 2007) reject their view and do not treat metaphor processing as a theory of analogy. Nonetheless, even these accounts require identifying the set of relevant properties shared by the metaphorical target and vehicle, thereby involving some kind of analogy perception. For our purposes, it suffices to commit

to the claim that analogical skills represent a cognitive substratum in metaphor comprehension: whatever the specific mental operations (analogical mapping or broadening/narrowing), one could not access the metaphor ground without first being *equipped* with the ability to detect similarities across entities (Pouscoulous, 2014; Rubio-Fernández & Grassmann, 2016).

Alternative naming and analogy perception follow different developmental trajectories and difficulties in one or both may impose additional demands on metaphor processing and impede pre-schoolers' metaphorical abilities. Rubio-Fernández & Grassmann (2016) are the first to address this issue. They tested pre-schoolers (age: 3 and 4) within a pretence scenario in which children were asked to hand the experimenter objects referred to with literal labels. Their baseline condition involved only spatial analogy, but the target referring expression required children to integrate both implicit spatial reasoning and alternative label [1]. Children exhibited difficulties when the two skills were required: 3-year-olds were at chance, while 4-year-olds performed significantly worse compared to the baseline condition. According to the authors, these findings suggest that pre-schoolers' difficulty with alternative naming might compromise metaphor comprehension.

These results provide initial evidence for a link between pre-schoolers' metaphoric difficulties and general cognitive development; specifically, they suggest alternative naming and analogy perception play a crucial role in metaphor development. However, Rubio-Fernández & Grassmann (2016) tested children on literal language, leaving the direct implication of these two abilities in metaphor comprehension uncertain. Moreover, Rubio-Fernández & Grassmann investigate pre-schoolers' ability to *combine* alternative naming and analogy perception. Further research is needed to tease apart the independent contribution to metaphor development of the two abilities.

[1] In this task, building blocks and toys were used as props for the referential task. In the baseline condition, the pretended object was referred to in a way that involved only spatial analogical reasoning (e.g., "the train with the hat", where a building block was placed either on the front or on the back of a train toy). In the experimental condition, both spatial analogy and alternative naming were required (e.g., "the train with the hat", where a dog toy was placed on the front or on the back of a train toy).

1.2 Alternative Naming and Analogical Reasoning

When occurring within the same communicative context, alternative naming – giving two labels to the same object – is problematic until age 4. Classic studies show that 3-year-olds exhibit severe problems accepting synonyms (e.g., rabbit - bunny) or superordinate categories (e.g., rabbit - animal) as second labels presented within the same context. This pattern improves starkly from age 4 (Doherty & Perner, 1998; Perner et al., 2002). Recent studies also showed that accepting a new conventional label for an object previously labelled using another conventional term is more demanding for pre-schoolers, particularly if a referential pact is violated. Matthews, Lieven & Tomasello (2010) tested 3-and-4-year-olds in a referential communication task in which a conventional label (e.g., "car" for a toy car) was used by one experimenter. Subsequently, an alternative label for the same object (e.g., "truck" for the toy car) was used either by the original experimenter or by a new experimenter. Children in both age groups selected the correct object above chance. Additionally, reaction times revealed children's sensitivity to alternative naming, as well as a developmental pattern: first, all children were slower when a new term was used; second, the difference in reaction times between New Term ("truck") and Same Term ("car") was significantly greater with the original experimenter than with a new one; finally, this pattern significantly decreased between 3- and 4-year-olds.

Perner and colleagues maintain that pre-schoolers difficulties with alternative naming in the same context stem from their difficulties to take an alternative perspective – i.e., understanding that the same entity can be described and mentally represented differently from different perspectives (Perner et al., 2002; Perner et al., 2003; Perner et al., 2007)[2]. This ability correlates with the development, around the age of 4, of theory of mind as tested in a standard false belief task (e.g., Doherty & Perner, 1998).

Alternative naming may tax pre-schoolers' metaphor interpretation. Metaphor involves using an unusual label to refer to a known object (e.g., *marshmallows* for *babies*), which

[2] Here, alternative naming is taken not just to imply the use of two words for the same thing, but also to imply and require taking an alternative perspective on one single entity, a perspective that is cued by the speaker's use of the second label. The reasoning behind this is that for this type of referential expression to be understood, one needs to understand (i) that something can be described in different ways and (ii) that speakers can take alternative perspectives on one and the same entity (e.g., see Perner et al., 2003).

implies taking two different perspectives on the same entity. While alternative naming and metaphor differ in many respects, it is hard to understand metaphor without first accepting that two known labels – a literal and a figurative one – are used for one and the same referent. Alternative naming might then represent a source of difficulty for children's metaphorical competence; considering its developmental trajectory, we expect this would be the case for 3-year-olds, but not by age 4.

Analogical reasoning is a relatively early-developing ability. The standard test to assess its development in pre-schoolers is the picture sequencing task (Goswami, 1989; Goswami & Brown, 1989), which requires mastery of the relational similarity constraint to complete an analogical sequence. In this task, the child is presented with two items, A and B; a third item, C, is then presented and the child has to select a fourth picture, D, from a range of alternatives to complete the sequence. The correct D picture shows an object that holds the same relation to C as B has to A (e.g., "Cat is to kitten as dog is to?" with "puppy" being the correct response). Three-year-olds already succeed in this task. For example, using this paradigm, Goswami & Brown (1989) tested 3-, 4- and 6-year-olds' analogical reasoning skills (i.e., driven by the relational similarity constraint). In their experiment 1, children saw two critical conditions: Explanation and Induction[3]. In Explanation, the strategy to complete the sequence was made explicit to the children (i.e., at the end of each trial, the analogical relation was spelled out and the experimenter explained what the correct choice was and why). In Induction, the strategy was induced (i.e., no explanation and only feedbacks on the kids' choice were given). This study revealed three main findings. First, all age groups were above chance in all conditions, but performance improved with age: 3-year-olds completed significantly fewer analogies than 4-year-olds, and in turn 3- and 4-year-olds completed significantly fewer analogies than 6-year-olds, who were at ceiling. Second, interestingly, in the Induction condition (as compared to Explanation), both 3- and 4-year-olds' performance dropped significantly. Third, pre-schoolers' most frequent errors were pictures showing either the correct transformation on the wrong object or the wrong transformation on the correct object,

[3] For each of the experimental condition (i.e., Explanation and Induction), there was a control condition that tested children's knowledge of the higher-order relations used in the analogies.

which the authors interpret as unsuccessful attempts at reasoning by analogy: children were trying to solve the analogies based on higher-order relations, but they had problems combining the two relations that were required to succeed in the task[4] (i.e., the source of the transformation and the correct transformation of the object). More recent evidence corroborates that pre-schoolers' analogical abilities are constrained by task complexity (e.g., Richland et al., 2006; Tunteler & Resing, 2007).

The effect of task complexity on the development of analogical skills may be relevant to pre-schoolers' metaphorical abilities. Indeed, metaphor comprehension requires grasping the metaphor ground, where the relation between target and vehicle is implicitly conveyed (see Rubio-Fernández & Grassmann, 2016). Thus, even though pre-schoolers possess some analogical abilities, their perception of analogical relations might not be sufficiently developed to identify the relevant implicit property and apply it to the correct object. In this respect, analogy perception might still represent a source of difficulty for children's metaphor comprehension throughout their pre-school years.

1.3 The present study

The main goal of this study was to extend Rubio-Fernández & Grassmann's (2016) findings and tease apart the contribution of alternative naming and analogy perception in pre-schoolers' metaphor understanding. Assuming that metaphor comprehension involves multiple cognitive abilities, we hypothesized that some of pre-schoolers' difficulties in metaphor interpretation might stem from difficulties with alternative naming and analogy perception, which makes the development of these two abilities possible predictors for metaphor development.

This study is the first investigating directly the role of alternative naming and analogy perception in pre-schoolers' metaphor understanding. We developed a new experimental procedure in which 3- and 4-year-olds were tested in three tasks independently assessing their metaphorical, alternative naming and analogical skills. Importantly, these were all built within the same scenario based on a general picture-

[4] Two more types of errors were possible in this analogy paradigm (i.e., children could choose the D picture among 5 alternatives): 1) mere appearance match and 2) thematic/category match).

selection procedure. Each task was inspired by previous experimental works and adapted so that all three tasks could be used within the same paradigm. Children were asked to move pictures around a grid following the experimenter's instructions. In the metaphor task, the experimenter referred to the pictures either metaphorically or literally. In the alternative naming task, the experimenter used either a previous label or a new one. In the analogy perception task, children completed a sequence of two pictures by choosing a third picture from a range of alternatives. Additionally, a separate pointing and naming picture book task assessed children's understanding and production of the vocabulary used in the metaphor comprehension task. This was done to ensure that correct answers on the metaphor task were driven by metaphorical interpretations rather than overextension of the metaphorical term (Pouscoulous, 2011, 2014).

Testing children's alternative naming and analogical abilities within the same experimental scenario as metaphor understanding enables us to investigate each of these abilities *per se* while simultaneously reducing the discrepancies between experimental methods between tasks. Additionally, assessing children's metaphorical abilities with a task including a literal condition is crucial because it allows pre-schoolers' difficulties with metaphor understanding – if any – to emerge more clearly.

For metaphor comprehension, we predicted pre-schoolers' greater difficulties with metaphoric than literal language. For alternative naming and analogy perception, we predicted pre-schoolers' metaphoric proficiency to increase as a function of their proficiency in these skills. Both abilities should act as enhancing or impeding factors depending on their developmental trajectories. Based on previous literature, alternative naming should represent a source of difficulty for metaphor comprehension at age 3, but it should be resolved by age 4. As for analogy perception, by age 4 – although not necessarily at 3 – children should at least identify the salient property to solve an analogy, even though this might not be sufficient for metaphor understanding.

2. Methods

2.1 Participants

Thirty-nine Italian speaking children participated in the experiment: 19 3-year-olds (*M*=3;4, range 3;1 to 3;11, 11 girls) and 20 4-year-olds (*M*=4;5, range 4;1 to 4;11, 9 girls). They were recruited from and tested in the nursery and pre-school sections of a single school in a middle-sized Italian city. The school is located in a middle-SES area of town and participants all lived in the area. Informed consent was obtained from the children's parents/guardians. This study was approved by the Research Ethics Committee of XXX (Project ID number: XXX) and the XXX.

One 3-year-old was excluded from the analysis for failing to perform correctly in the familiarization phase of the grid game after being corrected more than twice.

2.2 Materials and Procedure

Children saw four tasks, in the following order: first, the metaphor comprehension task, second the alternative naming task and third the analogy perception task, all of which were administered within the general procedure of a grid game; finally, a vocabulary assessment, was given at the end of the experiment. The tasks were administered in fixed order, a common practice in research on individual differences to control for fatigue and transfer effects (e.g., Carlson & Moses, 2001; Wiebe et al., 2011). The specific order was constructed so as to maximally reduce the differences across tasks and to minimize the risk of confusion for the children[5]. Materials and instructions for all tasks are available on the Open Science Framework web platform (XXX).

2.2.1 The Grid Game

The Grid Game was a picture-matching paradigm inspired from Morriseau, Davies & Matthews (2013). For each task, a set of 22x15cm magnetic laminated pictures was created that children could attach on and detach from a 100x79cm wooden grid. Children were told they would play a game: they would have to arrange pictures on the grid

[5] We gave priority to the metaphor task and administered it first in case some children could not complete the experimental session (which never happened). The alternative naming task was administered second because its procedure is the most similar to the metaphor task. Analogy perception came last within the Grid Game because its procedure diverges more from the other two. Vocabulary was assessed last of all because the procedure is entirely different from the Grid Game and because it was essential to assess vocabulary after the metaphor task, to avoid priming effects on metaphor comprehension.

according to a configuration that they could not see and the experimenter would give them instructions to achieve this. The children could look at the picture configuration booklet after they completed each task. All tasks included a warm up and a test phase. During the warm up phase children attached the set of pictures on the grid following the experimenter's instructions (i.e., to match configuration 1). This familiarized children with the picture items and the procedure. In the test phase, the pictures on the grid had to be arranged according to a new configuration (i.e., configuration 2). This time, the experimenter would tell the children which picture to remove from the grid; their task was to detach the pictures referred to and put them in a basket.

Depending on the task, the experimenter referred to target items in a way that would test the ability under scrutiny. The metaphor and alternative naming tasks shared the same procedure (i.e., moving the pictures in the grid to match a final configuration). Crucially, the experimenter referred to the target picture metaphorically or literally in the metaphor task and with the same or a different label in the alternative naming task. In the analogy task, children completed a sequence of pictures using the same attaching/detaching procedure.

For all tasks, the experimenter never looked at the pictures in the grid or at the child while instructing the child, but kept her eyes on the picture configuration booklet (lying flat in front of them). This was done to avoid biasing children's choices through non-verbal cues.

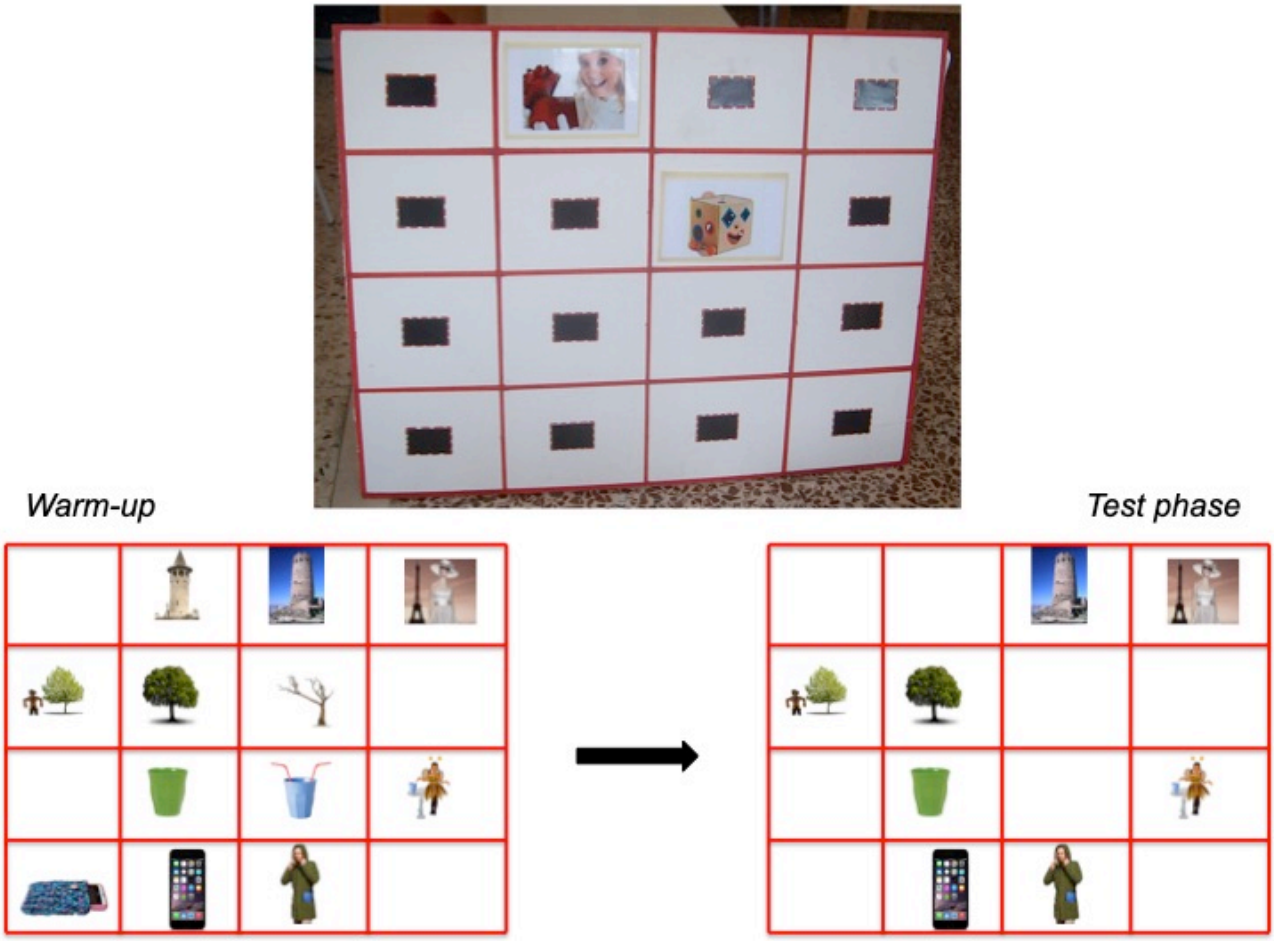

Figure 1. The Grid Game and two sample pictures for Configurations 1 (i.e., Warm-up) and 2 (i.e., Test phase) used in the metaphor task.

2.2.2 The Metaphor comprehension task

The metaphor task was adapted from Deamer (2013). Participants saw three pictures in each of eight trials. Eight pairs of novel metaphors and corresponding literal expressions were created (Table 1)[6]. In each trial, the experimenter referred to the target picture either using a metaphor of the form [The X with the Y] or literally, in the metaphor and literal condition respectively (e.g., *the glass with the antennae/straws* for a glass with two straws). The children's task was to take the referred picture off the grid. Children were randomly assigned to one of two material lists each containing 4 metaphorical and 4 literal scenarios; metaphors and literal counterparts were never in the same list.

[6] Similarly to Rubio-Fernández & Grassmann (2016), who tested literal language based on spatial analogy, we constructed metaphors based on body parts, the understanding of which requires a perception of spatial analogical relations relative to body parts. This was done to ensure the material was age-appropriate.

Table 1. List of the metaphorical and literal expressions used in the Metaphor Task. *Original expressions in Italian*. Trials 1, 3 and 5 were taken from Pouscoulous & Tomasello (2019).

	Metaphor	Literal
1.	The car with the backpack *Ita: La macchina con lo zaino*	The car with the box on the top *Ita: La macchina con la scatola nel tetto*
2.	The kitten with the socks *Ita: Il gattino coi calzini*	The kitten with the white paws *Ita: Il gattino con le zampe bianche*
3.	The carrots with the hair *Ita: Le carote coi capelli*	The carrots with the leaves *Ita: Le carote con le foglie*
4.	The bottle with the big belly *Ita: La bottiglia col pancione*	The round bottle *Ita: La bottiglia rotonda*
5.	The tower with the hat *Ita: La torre con il capello*	The tower with the pointy roof *Ita: La torre col tetto a punta*
6.	The tree with the arms *Ita: L'albero con le braccia*	The tree with the branches *Ita: L'albero con i rami*
7.	The glass with the antennae *Ita: Il bicchiere con le antenne*	The glass with the straws *Ita: Il bicchiere con le cannucce*
8.	The mobile phone with the coat *Ita: Il cellulare col cappotto*	The mobile phone with the cover *Ita: Il cellulare con la custodia*

In each trial, children could choose among three pictures: a target and two controls – *Target-Literal* and *Vehicle-Literal*. The target picture illustrated the target object referred to either metaphorically or literally (e.g., a glass with two straws for *The glass with the antennae/straws*). The *Target-Literal* control showed the metaphor target without the relevant property (e.g., a glass with no straws). The *Vehicle-Literal* control was a literal competitor literally showing both target and vehicle (e.g., a glass and a girl wearing an antennae-headband) - Figure 2.

During warm-up, the experimenter introduced the pictures using literal language (e.g., Item 1: "*Look! What a wonderful glass!";* Item 2: "*Look! One more glass!";* Item 3: "*Look! Another glass!"* for the *glass with the antennae* trial). In the test phase, the experimenter asked the child to remove a picture from the grid using either a metaphor or a literal expression depending on the condition (i.e., "*take the glass with the antennae/straws'*). If the child did not choose, the experimenter repeated the instruction a second time. If the child still failed to make a choice, the experimenter named the target object literally.

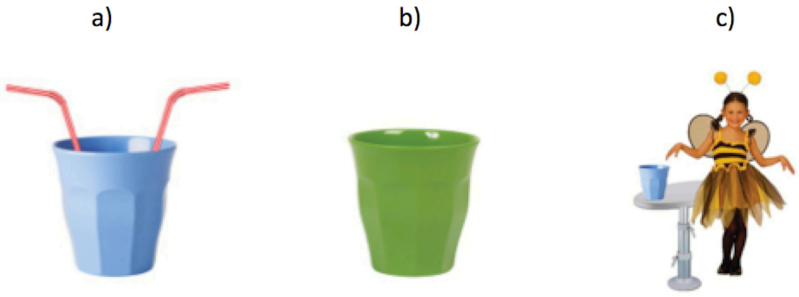

Figure 2. Picture alternatives for the metaphor item *The glass with the antennae* and the literal item *The glass with the straws*: (a) target picture, (b) *Target-Literal* distractor, (c) *Vehicle-Literal* distractor.

2.2.3 The Alternative Naming Task

The alternative naming task was adapted from Matthews, Lieven & Tomasello (2010)[7]. Children saw thirteen pictures (8 targets and 5 fillers) on the grid that they reconfigured according to the experimenters' instructions. The target pictures were referred to using the same label in the Same Term condition and two different labels in the New Term condition (4 pictures/trials per condition). In the Same Term condition, conventional literal labels were used (e.g., *Backpack* for the picture of a backpack). In the New Term condition, four pairs of literal expressions were used: the first term was the conventional label (e.g., *Lollipop,* for the picture of a lollipop), while the second was an alternative label the target picture can felicitously be referred to in Italian (e.g., *Candy* for lollipop) - Table 2.

The Same Term condition was administered before the New Term condition to avoid spill-over effects. During warm-up, items were named three times (e.g., 1: "*Look! A lollipop!;* 2: *I like this lollipop*; 3: *Do you like this lollipop?"*). In the test phase, the experimenter told the child which picture to take off the grid. In the New Term condition the target object was referred to with a different label (e.g., "*take the candy off*"). In the Same Term condition, the experimenter used the same label as in the warm-up (e.g., "*take the backpack off*"). If the child did not pick a picture, the instruction was repeated. If she still failed to make a choice, the target object was named using the original label.

[7] While being faithful to their overall paradigm, our procedure involved a single experimenter, rather than two as theirs.

Table 2. List of the target expressions used in the Alternative Naming task for the Same Term and the New Term condition *(original expressions in Italian).* The New Term pairs Truck/Car and Book/Story were taken from Matthews et al., (2010).

Same Term	New Term
Kitten (*Micio*)	Truck/Car (*Camion/Macchina*)
Boots (*Stivali*)	Book/Story (*Libro/Favola*)
Spaghetti	Lollipop/Candy *(Leccalecca/ Caramella)*
Backpack (*Zaino*)	Colours/ Pencils *(Colori/Matite)*

2.2.4 The Analogy Perception task

This task was adapted from Goswami & Brown (1989, Experiment 1). In each of four trials, children saw a sequence of two pictures (A & B) and were asked to complete this by choosing a picture, C, from three alternatives. The analogies were based on the relational similarity constraint: the items of each sequence always shared a relational feature on which an analogy could be based. We avoided surface similarity (i.e., based on the object's appearance), which would bias children's analogical reasoning (Gentner, 1989). All relations and objects depicted in the pictures were age-appropriate. The following relations were used (one per trial): (i) Housing*;* (ii) Family; (iii) Open; (iv) Animals flying.

The three picture choices included one target and two controls. The target picture showed the relevant property applied to the relevant object – e.g., for *The animals that fly*, the A and B items were a flying butterfly and a flying bird; the target C item was a flying bee. To complete the sequence with the correct C picture, children were required to (i) grasp the relation holding between A and B and identify the shared feature (e.g., flying); and (ii) apply this to the correct picture. Control 1 - *Correct property/wrong object* - illustrated the relevant property applied to the wrong object (e.g., a helicopter). Control 2 - *Wrong property/correct object* - depicted an irrelevant property on the right object (e.g., a sleeping bee) - Figure 3. Both controls presuppose reasoning about higher-order relations. However, selecting the *Wrong property/correct object* indicates a substantial lack of

analogical abilities because the salient relational property is missing. Conversely, choosing the *Correct property/wrong object* shows that children identify the salient property, but that their analogical skills are not yet sufficiently developed to complete the analogy. This is crucial since we wanted to assess the extent to which analogy perception is developed to enable pre-schoolers accessing the metaphor ground – which requires both identifying a relevant property of the vehicle and applying it to the target.

Children would place two pictures (A & B) on two slots of the grid with the experimenter. After this, their task was to choose the picture which completed the sequence. The experimenter introduced pictures A and B one at a time by drawing the children's attention to a given salient property. For A, this was phrased as follows: "*Look! This is a butterfly that flies, put it in this slot of the grid"* (pointing to the dedicated slot). For B: "*Look! This is a bird and it flies too. So, put it here* (pointing to the slot in the grid), *close to the butterfly that flies*". Then, the three alternative pictures were introduced and the child was asked to name the depicted objects to ensure she recognized them. Finally, since the relation between metaphor target and vehicle is implicit, the analogical problem was inducted (i.e., more implicitly presented). For example, *the animals that fly* trial was phrased as follows: "*if the flying butterfly is close to the flying bird, what do you put here* (pointing to the empty slot of the sequence) *to complete the sequence? This* (pointing to alternative 1), *this* (pointing to alternative 2) *or this* (pointing to alternative 3)?" If the child did not choose, the analogical problem was repeated. If the child still failed to make a choice, the experimenter solved the problem.

To ensure the materials in all three tasks were adequate, fifteen native Italian speakers (age: 20 to 40) were tested on all tasks with all the materials. They chose the target picture 100% of the time, indicating that the material was understandable and suitable. Additionally, five 9-year-olds were tested on the analogy task and chose the target item 100% of the time.

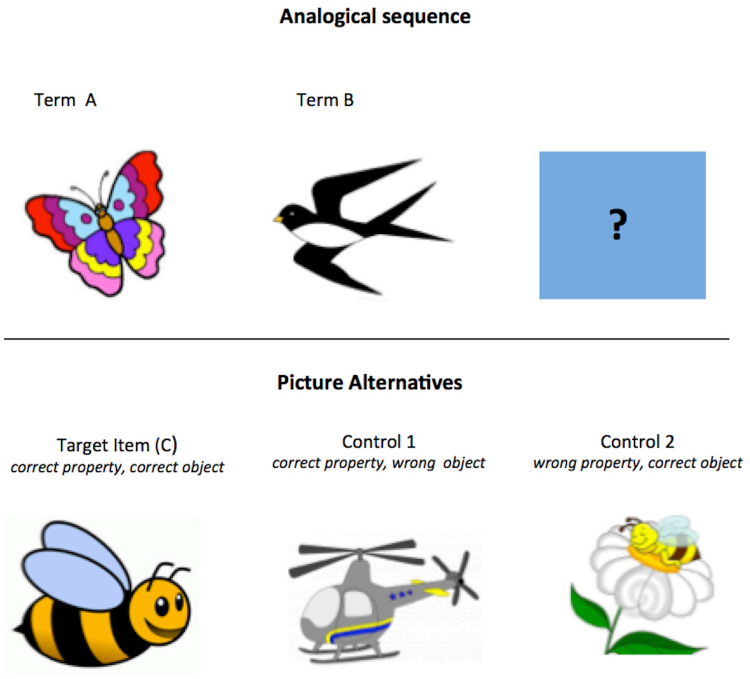

Figure 3. Example of the material used in the analogy task and location of the items in the grid. Here, the material for the *Animals that fly* trial is illustrated. To create the sequence in the grid, A and B were arranged on 2 contiguous slots of the grid. The triplet of alternative choices was arranged in 3 contiguous slots of the grid on the line below the slots occupied by A and B. The slots of the grid dedicated to A and B and to the three picture choices were the same across trials and participants. The order of presentation of the picture triplets was randomized across trials and participants.

2.3 Vocabulary assessment

A naming-and-pointing picture book adapted from Pouscoulous & Tomasello (2019) assessed children's knowledge of the literal vocabulary used in the metaphor task. The book consisted of a comprehension and a production part. In each of the two parts, four pictures tested the vocabulary of each metaphor. These showed, first, the literal referent of the metaphor vehicle (picture 1; e.g., the antennae of a snail for *The glass with the antennae* trial) to assess children's comprehension/production of the literal label for the target referent; second, the intended referent of the metaphor vehicle (picture 2; e.g., a straw) to control for overextension; and finally two distractors (pictures 3 and 4). Different pictures were used for comprehension and production and they all differed from the pictures in the metaphor task. Comprehension was assessed first by asking the child to point to the objects the experimenter named (e.g., *where are the antennae?).* In the production part, the child was asked to name the objects the experimenter pointed to (e.g., *what is this?*).

2.4 Coding and Statistical Analyses

Response accuracy was coded for all tasks as 1 or 0 depending on whether the child picked the correct picture. For metaphor and alternative naming, the picture the child picked from the grid was considered her choice. For the analogy task, the child's choice was considered made when the child attached a picture to the sequence. In the metaphor and alternative naming tasks, participants could achieve a maximum score of 8 correct responses (4 in each condition). In the analogy task, the maximum score was 4 correct responses. Additionally, incorrect responses in the metaphor and analogy task were then recoded as a binary variable to analyse the type of error.

Reaction times were also collected in the metaphor and alternative naming tasks to measure processing times; this has not previously been done in metaphor studies, but it is standardly used in alternative naming tasks to assess sensitivity in a referential paradigm (Matthews et al., 2010). Reaction times were collected for all responses (both correct and incorrect) from the onset of the target sentence to when the child provided her response by picking a picture. They were computed by subtracting the time of the experimenter's utterance to this time interval (ELAN 5.0.0-beta software; Lausberg & Sloetjes, 2009). Children always selected a picture and did not need additional prompts in the metaphor task. In the alternative naming task, children sometimes produced a verbal response in incorrect trials (see below). In these cases, reaction times were calculated up to when the child produced her verbal response (rather than when she picked a picture). Trials were excluded from coding if the child spoke during the experimenter's instructions or if she did (or spoke of) something else before picking the target object during a trial. This occurred with 2 trials in the metaphor task, 1 trial in the alternative naming task and 1 trial in the analogy perception task.

For vocabulary assessment, each correct answer was assigned one point. The responses were deemed correct when the child pointed to the target picture (i.e., comprehension, 16 items) and when she produced a literal label for the object (i.e., production, 16 items); correct literal labels included synonymous, dialectical expressions and words with slightly imprecise denotations, as long as these were not overextensions of the target expression.

All experimental sessions were filmed (with parental consent). The experimenter coded all video-recordings and a blind coder coded 20% of the data. Agreement was 100%.

As a general procedure, accuracy was analysed with Generalized Linear Mixed Models (GLMMs) and reaction times with Linear Mixed Models (LMMs). The random structure of the models included random intercepts for subjects and items and by-subjects random slope for condition. Age group, condition and their interactions were the fixed factors. To break down (marginally) significant interactions, simple effects analyses of variance (for LMMs) and two-sided Z-Tests with continuity correction for the analysis of proportion in each group (i.e., accuracy) were conducted. Kendall's Tau correlations were also conducted to better identify developmental effects. Errors in metaphor and analogy perception were analysed with chi-squared statistics with Yate's continuity correction. To tease apart the contribution of alternative naming and analogy perception to metaphor comprehension, GLMM was used. Here, alternative naming and analogy perception data were treated as predictors and responses in the metaphor task were the outcome measures.

3. Results

Table 3 reports the proportion of correct responses and mean reaction times for all tasks. In what follows, we report only the relevant results. See Appendix A for all results. Data are available on the Open Science Framework web platform (XXX).

Table 3. Proportion of correct responses (Mean Accuracy and Lower CI - Upper CI) and Reaction Times in seconds (mean and standard deviation) for all experimental tasks. Metaphor and Alternative Naming Tasks: Accuracy and Reaction Times; Vocabulary and Analogy Task: Accuracy.

Metaphor Task				
	Accuracy		Reaction Times	
	Metaphor	Literal	Metaphor	Literal
3-year-olds	0.45 (0.34 - 0.57)	0.77 (0.67 - 0.87)	7.68 (6.74)	6.12 (3.92)
4-year-olds	0.56 (0.45 - 0.67)	0.93 (0.88 - 0.99)	4.72 (1.78)	4.52 (2.18)
Vocabulary Assessment				
	Accuracy		//	
	Comprehension	Production		
3-year-olds	0.88 (0.83 - 0.93)	0.86 (0.80 - 0.92)	N.A.	
4-year-olds	0.92	0.93	N.A.	

	(0.88 - 0.96)	(0.89 - 0.97)		
Alternative Naming Task				
	Accuracy		Reaction Times	
	New Term	Same Term	New Term	Same Term
3-year-olds	0.77 (0.68 - 0.87)	0.98 (0.95 - 1)	7.13 (6.30)	4.06 (2.79)
4-year-olds	0.96 (0.92 - 1)	1 (1 - 1)	3.73 (2.59)	3.06 (2.41)
Analogy Perception Task				
	Accuracy		//	
3-year-olds	0.29 (0.18 - 0.40)		N.A.	
4-year-olds	0.40 (0.29 - 0.50)		N.A.	

3.1 Metaphor

Children from both age groups performed above chance (i.e., 0.33) in the literal and metaphorical conditions as their probability of selecting the correct picture falls outside the 95% Wald CI of the estimated mean proportion of correct answers (Table 3). The GLMM on accuracy revealed that both age groups were significantly less accurate at correctly interpreting metaphors than literal language, as shown by a significant main effect of Condition ($\chi^2(1) = 18.97$; $p<0.0001$). This analysis also revealed a marginal interaction ConditionXAge Group ($\chi^2(1) = 2.90$; $p=0.08$), but the Z-Test confirmed that the proportion of correct choices in metaphor vs. literal significantly differed in both groups (3-year-olds: $\chi^2(1)=13.37$; $p<0.0005$; 4-year-olds: $\chi^2(1)=28.03$; $p<0.0001$).

The error analysis on incorrect metaphoric trials (Figure 4) showed that the most common error was the *Vehicle-Literal* control for both 3-year-olds ($\chi^2(1) = 23.21$; $p<0.0001$) and 4-year-olds ($\chi^2(1) = 32.91$; $p<0.0001$), with no significant group difference ($\chi^2(1) = 0.20$; p = n.s.). No relevant effects emerged on reaction times (see Appendix A).

Overall, these results suggest that both age groups could interpret novel metaphors. Yet, children exhibited more difficulties interpreting metaphorical than literal expressions.

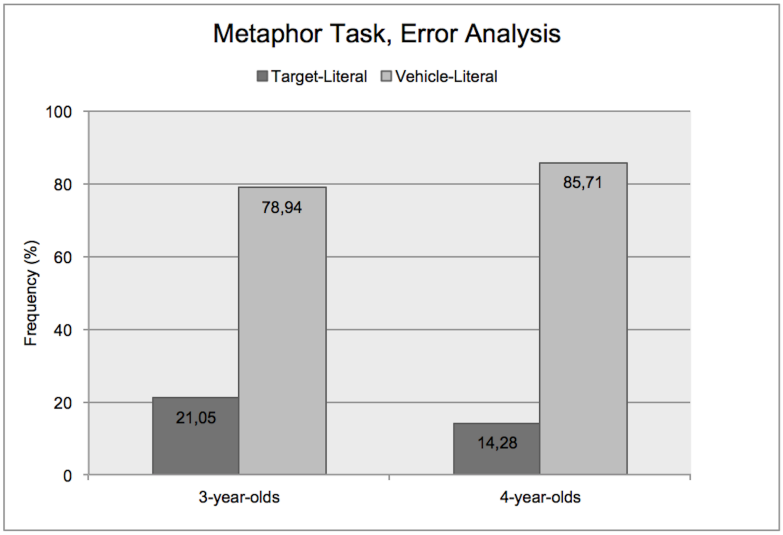

Figure 4: Frequency (%) of type of choice in metaphorical trials with accuracy = 0 per age group.

3.1.1 Vocabulary assessment

Both 3- and 4-year-olds performed very well on the vocabulary assessment (Table 3), with no group difference for comprehension ($\chi^2(1) = 1.73$; p=0.18) or production ($\chi^2(1) = 1.45$; p=0.22). A GLMM with children's comprehension and production scores as covariates, metaphor accuracy as the dependent variable and condition as the independent variable[8] revealed that accuracy in the metaphor task did not correlate with children's lexical knowledge (Comprehension: $\chi^2(1) = 0.80$; p=0.36; Production: $\chi^2(1) = 0.04$; p=0.82; ConditionXComprehension: $\chi^2(1) = 0.50$; p=0.47; ConditionXProduction: $\chi^2(1) = 0.55$; p=0.45). Lexical difficulty, then, cannot account for any difficulty children might have had with metaphor interpretation.

3.2 Alternative Naming

Both age groups were above chance since the expected probability of picking the correct picture (i.e.,0.50) falls outside the 95% Wald CI of the estimated mean proportion of correct answers (Table 3). GLMM on accuracy revealed a significant effect of Condition

[8] Age group was not included as a fixed term in these models for two reasons. First, the statistics on vocabulary scores did not reveal any group-related effect for both comprehension and production. Second, when Age Group was inserted, the models did not converge (see Barr et al., 2013 for issues about convergence of GLMMs). However, since no group differences emerged in vocabulary assessment, keeping the Age Group out of the GLMMs fixed structure still produces reliable results.

($\chi^2(1) = 9.75$; $p<0.0001$) and a significant interaction ConditionXAge Group ($\chi^2(1) = 14.55$; $p<0.0001$). The analysis of proportion of correct responses revealed that, contrary to 4-year-olds ($\chi^2(1) = 1.38$; $p=0.23$), 3-year-olds were significantly less accurate in condition New Term ($\chi^2(1) = 13.07$; $p<0.0005$).

In incorrect trials, children could not find the referred object 36.84% of the time and protested/corrected the experimenter 63.15% of the time. These types of incorrect responses mostly occurred in the group of 3-year-olds and only in the New Term condition: 3-year-olds could not find the referred object 25% of the time and protested/corrected the experimenter 75% of the time; 4-year-olds provided incorrect responses only in 3 trials, in which they could not find the referred object.

The between-group analysis on reaction times revealed a significant interaction ConditionXAge Group ($F(1, 43.77)=6.84$; $p<0.05$). A simple effects ANOVA confirmed that, contrary to 4-year-olds ($F(1, 9.54)=0.22$; $p=0.64$), 3-year-olds were significantly slower in condition New Term than Same Term ($F(1, 9.92)=5.22$; $p<0.05$). A Kendall's correlation indicated that the difference in reaction times (New Term minus Same Term) negatively correlated with Age Group ($r_t = - 0.21$; $p<0.001$).

Overall, 4-year-olds exhibited no difficulty with alternative naming, while 3-year-olds still showed some.

3.3 Analogy Perception

Both age groups completed the analogical sequences successfully below chance since the expected chance probability of selecting the correct picture (i.e., 0.33) falls within the 95% Wald CI of the estimated mean proportion for both groups (Table 3). A Kendall's Tau correlation revealed no significant effect of Age Group ($r_t = 0.09$; $p=0.18$).

Interestingly, the error analysis revealed that the proportion of choices significantly differed between age groups ($\chi^2(1) = 4.38$; $p<0.05$) and the difference between the two controls was significant only in 4-year-olds (4-year-olds: $\chi^2(1) = 15.04$; $p<0.0001$; 3-year-olds: $\chi^2(1) = 0.04$; $p=0.84$) – Figure 5.

In our task, children did not successfully complete a three-terms analogical sequence by age 4. However, 4-year-olds – but not 3-year-olds – chose the *Correct property/wrong object* control significantly more often than the *Wrong property/correct object* one.

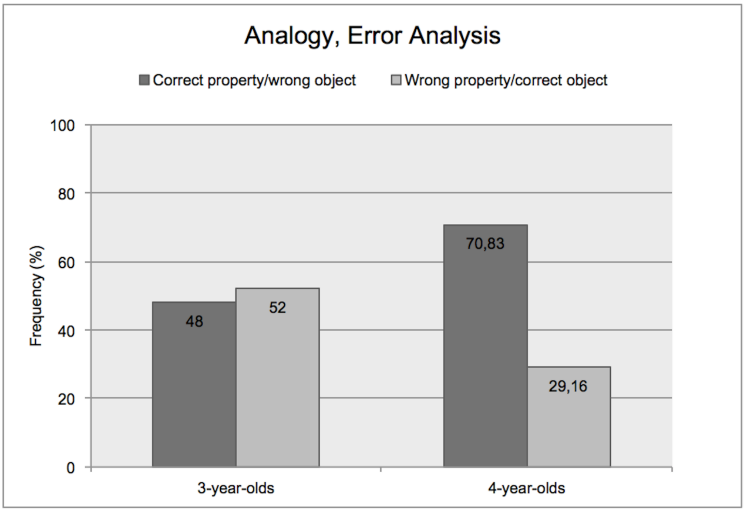

Figure 5: Frequency (%) of type of choice in analogical trials with accuracy = 0 per age group.

3.4 Predictors of metaphor development

A GLMM was conducted with participants' accuracy in the metaphor task as the outcome variable, while Condition (i.e., literal vs. metaphorical) and participants' mean accuracy scores for analogy and alternative naming as well as the averaged difference reaction times from the alternative naming task were the predictors[9]. Most importantly, this analysis revealed that the following interactions were significant: MetaphorXAnalogy ($\chi^2(1)$ = 153.518; $p<0.0001$; $\beta=0.04$; SE=0.003347; z= 12.4), MetaphorXAlternative Naming Difference Reaction Times ($\chi^2(1)$ = 32.08; $p<0.0001$; β= - 0.01; SE=0.003343; z= - 5.7) and MetaphorXAlternative Naming Accuracy ($\chi^2(1)$ = 503839; $p<0.0001$; β=- 2.37; SE=0.003347; z= - 709.8). A likelihood-ratio test indicated that removing the predictors from the model significantly decreased the goodness of fit ($\chi^2(7)$= 56.28; $p<0.0001$).

[9] Within this model, Age Group was not inserted as a predictor since (i) the results on the metaphor task revealed a main effect of age group for both the metaphor and literal condition; and (ii) this analysis mainly aimed at assessing whether alternative naming and analogical abilities influence pre-schoolers' metaphorical competence, independently of age group.

Overall, this suggests that pre-schoolers' alternative naming and analogical abilities predict their metaphor understanding.

4. Discussion

This study extended Rubio-Fernández & Grassmann (2016) and looked at the role of alternative naming and analogy perception in the development of metaphorical competence. We tested 3- and 4-year-olds in metaphor understanding, as well as alternative naming and analogy perception and hypothesized these abilities would predict metaphor development. The following main findings emerged:

(i) Both 3- and 4-year-olds were above chance in the metaphor comprehension task. Interestingly, they were significantly less accurate at picking the correct picture when it was referred to metaphorically than literally.

(ii) Both age groups performed very well in the alternative naming task. However, when a new term was used, 3-year-olds were slower and less accurate than 4-year-olds.

(iii) All children were below chance in the analogy task. Nonetheless, the error pattern was different between groups: while 3-year-olds indiscriminately chose one of the two controls, 4-year-olds mostly picked a picture showing the relevant property.

(iv) Pre-schoolers' performance on both the alternative naming and the analogy perception tasks significantly correlated with their metaphor understanding.

We now discuss these points in turn.

In the metaphor comprehension task, both age groups picked the target picture above chance in the metaphoric condition. Importantly, since they mastered the literal meaning of the words used metaphorically (according to the vocabulary assessment results), when choosing the correct picture, they were not overextending the metaphorical terms. Thus,

pre-schoolers can already grasp novel metaphors. This finding indicates an early development of metaphor understanding and disconfirms the traditional view that children do not access metaphor before age 8 (e.g., Winner, 1988/1997). Rather, our findings replicate recent studies showing signs of metaphor understanding from age 3 (e.g., Pouscoulous & Tomasello, 2019; Deamer, 2013; Bühler, Perovic & Pouscoulous, 2018). Importantly, this study brings to light a novel finding: pre-schoolers were significantly less accurate at selecting the target picture when it was referred to metaphorically than literally. This indicates that both 3- and 4-year-olds exhibited more difficulties interpreting metaphorical than literal language. Unlike previous studies, which did not include a literal condition (e.g., Pouscoulous & Tomasello, 2019; Deamer, 2013), this manipulation highlights an aspect of metaphor comprehension where pre-schoolers still experience difficulties and, therefore, contributes to refining our understanding of the phenomenon. Interestingly, when children failed to access the metaphorical meaning of an expression, they gave it instead a literal interpretation, choosing the picture literally depicting both the metaphoric target and vehicle (for similar pattern of errors see, Declercq et al., 2010; Deamer, 2013). Inhibitory control might contribute to this trend (and the relatively low mean accuracy in the metaphor condition). There is stark development in inhibitory control around 3-to-4-years of age (Carlson & Moses, 2001) and pre-schoolers may still lack appropriate inhibition skills to suppress the more salient literal meaning, as suggested by Deamer (2013). Future studies should explore this hypothesis.

Both age groups had high accuracy rates in the alternative naming task indicating that pre-schoolers can cope with alternative naming within the same context in a picture-matching paradigm. Nonetheless, contrary to 4-year-olds, 3-year-olds were slower and less accurate when a new term was used. We follow Matthews et al. (2010) in considering that the most appropriate measure to establish how children fare in the alternative naming task is reaction times and not accuracy. Indeed, even 3-year-olds chose the target object most of the time. Importantly, the pattern of children's behavior when they do not choose the correct object is not necessarily consistent with an incapacity to take into account alternative naming and identify the intended referent. 'Inaccurate' choices among 3-year-

olds corresponded to two distinct types of behaviours (observed by Matthews et al., 2010, as well): in case I (25% of the inaccurate responses), children said that the object referred to with the new term was not present, while in case II (75%) they corrected the experimenter and protested to the use of the new term (e.g., "no, this is not a candy, it's a lollipop"). These types of responses only occurred in the 3-year-old group and only in the New Term condition. Both cases I and II were coded with an accuracy of 0 because they prevented the child from picking the correct object. Note, nonetheless, that while case I responses indicate difficulties with alternative naming, case II responses (protesting), on the contrary, imply understanding the intended referent and being sensitive to the fact that the referential pact was broken. Therefore, overall we cannot claim difficulties with alternative naming based on the behavioural measure in the alternative naming task, we need to rely on reaction times for this. Therefore, the most important finding in this task is that 3-year-olds were slower when a new term is being used. Alternative naming remains somewhat taxing for 3-year-olds, while difficulties disappear by age 4. Overall, these results corroborate those found by Matthews et al. (2010)[10] and confirm the general trend reported in the literature: alternative naming develops between age 3 and 4 (Doherty & Perner, 1998; Perner et al., 2002; Perner et al., 2003). These findings are in line with Perner's hypothesis that difficulties with alternative naming are overcome by age 4, when children possess sufficient Theory of Mind to entertain multiple mental representations and integrate several perspectives on a single object within the same context. Three-year-olds' extra effort would then be due to their difficulties with alternative perspective. The findings also fit with Matthews et al.'s (2010) view that a change in perspective interferes with a mental representation that selectively encoded what someone has said to them.

Both age groups were at chance in analogy perception, suggesting that their analogy perception is not sufficiently developed to work out an implicit relation. This outcome is only partially consistent with Goswami & Brown (1989, Experiment 1), where 3-and-4-year-olds were above chance. We believe this discrepancy is due to differences in

[10] Matthews et al. did not conduct statistical analyses on kids' accuracy, but they report high accuracy rates for both age groups and a trend in protesting/correcting the experimenter when the alternative naming is not accepted by participants, both of which seem to match our findings.

task demands. Goswami and Brown always provided supporting cues to prompt analogical reasoning: either explanation of the correct choice (i.e., explanation condition) or feedback when making mistakes (i.e., induction condition). Over trials, this might have trained the child to reason analogically and/or to grasp the key mechanism to complete the sequence. Indeed, children's performance dropped significantly in the Induction condition, where less cues were provided. In our experiment, analogy was induced and not facilitated by any supporting hints, which might explain the low mean accuracy rates in both groups. Although this was an unintended consequence of our paradigm, this interpretation of our findings fits nicely with recent work showing that task complexity constraints pre-schoolers' analogical abilities. Indeed, these are significantly worse when relational complexity and featural distractors are increased (e.g., Richland et al., 2006) and when the analogical transfer is unaided/untrained in spontaneous elicitation scenarios (e.g., Tunteler & Resting, 2007). Furthermore, error analysis revealed a developmental pattern (consistent with Goswami & Brown): while 3-year-olds indiscriminately chose one of the two controls, 4-year-olds' most common mistake was the *Correct property/wrong object* control (e.g., the helicopter). Thus, by the age of 4 – but not 3 – children identify a shared property (i.e., helicopters fly too) but they cannot yet apply it to the correct object; they therefore appear to follow an analogical strategy, albeit imperfectly[11].

Our main aim was to investigate the role of alternative naming and analogy perception in pre-schoolers' metaphor understanding. We predicted that the more/less the two abilities are developed, the more/less metaphorical competence is developed too. We found that metaphor proficiency significantly correlated with pre-schoolers' performance for both abilities, which confirms our general prediction.

[11] Note that the fact that children chose the target item at chance in our experiment does not necessarily suggest a total lack of analogy perception since all three items presupposed a certain degree of analogical abilities. In fact, the two control items were adapted from Goswami & Brown (1989, items E & F), who used those items to check whether, even though making mistakes, children were still attempting to reason by analogy. Based on this, chance performance in our study only suggests that, when elicited at a certain level of complexity, pre-schoolers' analogical abilities are not developed enough to simultaneously identify a shared property and apply it to the correct object within a context in which analogy perception is presented in a more implicit form.

Significant negative correlations emerged between children's performance on metaphor comprehension task and alternative naming – one with accuracy and the other with reaction times. First, children who were better at interpreting metaphors (i.e., more accurate in correctly picking the target metaphorical picture) were also more sensitive to the use of a second label for the same referent (i.e., less accurate in the alternative naming task, mostly by protesting and correcting the experimenter because the referential pact was broken). In the metaphor task, accuracy genuinely indicates children's comprehension of metaphor: if the child did not grasp the meaning of the metaphor, she just could not select the target picture (and would instead opt for one of the two controls which provided other – incorrect – referents). Yet, this does not necessarily apply for accuracy in alternative naming. Here, a child's response was deemed inaccurate not only when the child could not find the item (case I) but also when she protested/corrected the experimenter (case II). These cases were coded as inaccurate because they prevented the child from picking the target object and indicated difficulties with alternative naming (following Matthews et al., 2010). Importantly, however, protesting or correcting presuppose understanding what object is being referred to by the alternative label and only reveals that children have more difficulties *accepting* an alternative label (see discussion above). Accuracy in metaphor and alternative naming therefore mirrors different processes: genuine comprehension for the former, non-acceptance for the latter. This, in turn, may explain the negative correlation we found, which is still somewhat puzzling: one might have expected the opposite pattern. The explanation that seems the most plausible to us is that the effect is driven by the case II children who protested the new term (75% of the 3-year-olds' inaccurate answers). These children understand which object is referred to – since they object to it being referred to with a new label – but they appear to be particularly sensitive to the referential pact being broken. Sensitivity to the referential pact is arguably a pragmatic ability. We had no reason to anticipate a link between good metaphor understanding and sensitivity to referential pact in young children, but it is not absurd that these abilities should be linked – or at least correlate. Nonetheless, we would not want to make any big claims on the basis of an effect relying on only 12

incorrect trials in the group of 3-year-olds. Note that this reasoning does not apply to reaction times because they reflect the time needed to process the phenomenon at stake. They are, therefore, a more direct measure of the cognitive costs underlying alternative naming.

It is therefore more appropriate to focus on the second correlation between the two tasks: that between accuracy in metaphor understanding and reaction times in alternative naming. Interestingly, the shorter the difference in reaction times between New Term and Same Term, the higher children's accuracy in the metaphor task. Therefore, when alternative naming becomes easy enough and does not require additional processing costs, interpreting metaphor becomes easier too. This suggests that a sufficiently developed alternative naming ability enhances metaphor interpretation by providing the cognitive cornerstone to deal with its dual reference. Moreover, our results on alternative naming suggest it is a source of difficulty at age 3, but not 4. Hence, at age 3, the extra effort associated to alternative naming may tax metaphor comprehension, too.

Finally, children with more developed analogical abilities exhibited a better understanding of metaphor. This suggests that better developed analogical skills enhance pre-schoolers' metaphor interpretation. They presumably prompt the identification of the properties/relations shared by the metaphor target and vehicle, thereby facilitating the recognition of the metaphor ground. Even though both groups completed the analogical sequences at chance, error analysis suggested that at age 4, but not at 3, analogy perception seems more developed (i.e., children recognize relevant properties) and this might allow children to access the metaphor ground to some extent. Nonetheless, their analogy perception is not yet mature enough to apply the identified property to the correct object. Since both operations are needed to fully grasp the metaphor ground, this might contribute to explaining where their metaphorical difficulties stem from at this developmental stage.

Finally, both alternative naming and analogy perception improve at age 4: children could identify the relevant property in the analogy task and showed no difficulties with alternative naming. This is worth noting since the progress from 3 to 4 may be bound to inhibitory

control, which peaks between age 3 and 4 (Carlson & Moses, 2001) and is required by these tasks.

To sum up, alternative naming and analogy perception predict pre-schoolers' metaphorical proficiency. At age 3, both cognitive skills are not yet fully mature and may hinder metaphor interpretation. At age 4, alternative naming is not taxing anymore, but analogy perception remains costly. Thus, while alternative naming is not an impeding factor anymore, poor implicit analogy perception may still prevent children from fully appreciating metaphors. These findings echo recent research indicating that metaphor understanding depends on a cluster of cognitive abilities (Pouscoulous, 2014; Rubio-Fernández & Grassmann, 2016; Deamer, 2013). In addition to alternative naming and analogy perception, two abilities deserve special attention: inhibitory control and working memory. Inhibitory control appears to be crucial in suppressing the irrelevant literal properties of the vehicle. Indeed, metaphoric abilities are impaired in population with inhibitory control deficits (e.g., Amanzio et al., 2007) and some evidence shows that pre-schoolers' inhibition skills predict metaphor understanding (Deamer, 2013). As for working memory, metaphor interpretation requires maintaining and updating several information in working memory, which is limited in early childhood (Simmering & Perone, 2013) and correlates with adults' (e.g., Chiappe & Chiappe, 2007) and adolescents' (Carriedo et al., 2016) metaphor comprehension. Future research should further investigate the cluster of abilities involved in metaphor development. We know a lot about metaphor and development. Yet, our knowledge of what is in a metaphor is much deeper than our knowledge of which cognitive prerequisites allow a developing mind to jump into metaphoricity. Alternative naming and analogy perception are two such ingredients.

Acknowledgements

We would like to thank the participating children, their parents and the very helpful schools and nurseries.

References

Amanzio, M., Geminiani, G., Leotta, D. & Cappa, S. (2007). Metaphor comprehension in Alzheimer's disease: Novelty matters. *Brain and Language*, 107, 1–10.

Asch, S. & Nerlove, H. (1960). The development of double function terms in children: An exploratory investigation. In B. Kaplan & S. Wapner (Eds), Perspectives in Psychological Theory: Essays in Honor of Heinz Werner (pp. 47–60). New York, NY: International Universities Press.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.

Billow, R. M. (1975). A cognitive developmental study of metaphor comprehension. *Developmental Psychology*, 11, 415.

Bowdle, B. & Gentner, D. (2005). The career of metaphor. *Psychological Review*,112, 193 - 216.

Bühler, D., Perovic, A. & Pouscoulous, N. (2018). Comprehension of metaphor in young children with Developmental Language Disorder, *Autism and Developmental Language Impairments,3*:1-11

Carlson, S.M & Moses, L.J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development*, 72, 1032–1053.

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,: e0150289.

Chiappe, D. & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of memory and language*, 56, 172-188.

Cometa, M., & Eson, M. (1978). Logical operations and metaphor interpretation: A piagetian model. Child Development, 49, 649–659.

Deamer, F. (2013). *An investigation into the processes and mechanisms underlying the comprehension of metaphor and hyperbole* (PhD thesis). University College London, London, England.

Declercq, C., Baltazart, V., Didon, D. (2010). Les nuages sont un troupeau de moutons : a` quel aˆge les enfants comprennent-ils qu'il n'y pas de moutons dans le ciel? *Canadian Journal of Experimental Psychology*, 64, 142–152.

Doherty, M. J., & Perner, J. (1998). Metalinguistic awareness and Theory of Mind: Just two words for the same thing? *Cognitive Development*, 13, 279–305.

Falkum, I.L., Recasens, M., Clark, E. V. (2017). "*The moustache* sits down first": on the acquisition of metonymy. *Journal of Child Language*, 44, 87-119.

Gentner, D. (1989). The Mechanisms of Analogical Learning. In S. Vosniadou & A. Ortony (Eds.) *Similarity and Analogical Reasoning.* London: CUP, 199–241.

Gibbs, R.W. Jr. & Colston, H.L. (2012). *Interpreting Figurative Meaning.* NY: Cambridge University Press.

Goswami, U., & Brown, A.L. (1989). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition,* 35*,* 69-95.

Goswami, U. (1989). Relational Complexity and the development of analogical reasoning. *Cognitive Development*, 4, 251-268.

Goswami, U. & Brown, A.L. (1990). Higher-order structure and relational reasoning: Contrasting analogical and thematic relations. *Cognition,* 36*,* 207-226.

Goswami, U. (2001). Analogical reasoning in children. In K. Holyoak, D. Gentner, & B. Kokinov (Eds), *The Analogical Mind: Perspectives from Cognitive Science* (pp. 437–470). Cambridge, MA: The MIT Press.

Lakoff, G. & Johnsen, M. (2003). *Metaphors we live by*. 2nd edn. London: The university of Chicago press.

Lakoff, G. & Johnsen, M. (1980). *Metaphors We Live By*. Chicago, IL: The University of Chicago Press.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. Behavior Research Methods, Instruments, & Computers, 41, 841–849.

Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, 46, 749–760.

Morriseau, T., Davies, C., Matthews, D. (2013). How do 3- and 5-year-olds respond to under- and over-informative utterances? *Journal of Pragmatics*, 59, 26-39.

Nippold, M. A. (1988/1998). *Later Language Development: The School-Age and Adolescent Years* (2nd Ed.). Austin, TX: Pro-Ed.

Özçaliskan, S. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language*, 32, 291–318.

Pearson, B. (1990). The comprehension of metaphor by preschool children. *Journal of Child Language*, 17, 185–203.

Perner, J., Rendl, B., & Garnham, A. (2007). Objects of desire, thought, and reality: Problems of anchoring discourse referents in development. *Mind and Language*, 22, 475–513.

Perner, J., Brandl, J.L., Garnham, A. (2003). What is a Perspective Problem? Developmental Issues in Belief Ascription and Dual Identity. *Facta Philosophica*, 5, 355-378.

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of Mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development*, 103, 1–22.

Pouscoulous, N., Tomasello, M. (2019). Early birds: Metaphor understanding in 3-year-olds, *Journal of Pragmatics*, https://doi.org/10.1016/j.pragma.2019.05.021.

Pouscoulous, N. (2014). "The elevator's buttocks": Metaphorical abilities in children. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition* (pp. 239-259). John Benjamins.

Pouscoulous, N. (2011). Metaphor – for adults only? *Belgian Journal of Linguistics*, 25, 64-92

Richland, L.E., Morrison, R.G., & Holyoak, K.J., (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, 94, 249– 273.

Rubio-Fernández, P. & Grassmann, S. (2016). Metaphors as Second Labels: Difficult for Preschool Children? Journal of Psycholinguistic Research, 45, 931 - 944.

Simmering, V.R. & Perone, S. (2012). Working memory capacity as a dynamic process. *Frontiers in Psychology*, 3, 567.

Tunteler, E. & Resing, W.C.M. (2007). Change in Spontaneous Analogical Transfer in Young Children: A Microgenetic Study. *Infant and Child Development*, 16, 71–94.

Wiebe, A.S., Sheffield, T., Nelson, J.M., Clark, C.A.C., Chevalier, N., Espy, K.A. (2011). The structure of executive functions in 3-year-old children. *Journal of Experimental Child Psychology*, 108, 436 – 452.

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Advances in Pragmatics* (pp. 230–260). Houndmills, Basingstoke: Palgrave.

Winner, E. (1988/1997). *The Point of Words: Children's Understanding of Metaphor and Irony*. Cambridge, MA: Harvard University Press.

Appendix A

Results of all statistical analyses conducted for each experimental task and for the analysis of predictors of metaphor development (i.e., GLMM between accuracy in metaphor task and each of the measures collected for Alternative Naming and Analogy Perception). *Table A.1*: Metaphor Task (Accuracy and Reaction Times); *Table A.2*: Vocabulary Assessment (Accuracy); *Table A.3*: Alternative Naming Task (Accuracy and Reaction Times); *Table A.4*: Analogy Perception Task (Accuracy); Table *A.5*: GLMM Analysis between children's accuracy in metaphor task and their performance in each of the tasks for the predictors under scrutiny.

Table A.1

Metaphor Task						
Accuracy						
GLMMs						
	β	SE	z	χ^2	DF	$P(\chi^2)$
Condition	-2.84	0.6743	-4.22	18.97	1	<0.0001***
Age Group	-1.68	0.6639	-2.53	*3.85*	1	<0.05*
ConditionXAge Group	1.18	0.6971	1.70	*2.90*	1	0.08 .
Analysis of Proportion per age group: two-sided Z-Tests with continuity correction						
3-year-olds						
Condition	//	//	//	13.37	1	<0.0005**
4-year-olds						
Condition	//	//	//	28.03	1	<0.0001***
Kendall's tau correlation on metaphor trials only						
	z	r_t	p			
Age Group	- 1.27	- 0.10	0.20			
Error Analysis: Chi-squared tests with Yate's continuity correction						
				χ^2	DF	$P(\chi^2)$
3-year-olds						
Control choice				23.21	1	<0.0001***
4-year-olds						
Control choice				32.91	1	<0.0001***
3- vs. 4-year-olds						
Age Group				0.20	1	0.65
Reaction Times						
LMMs						
	β	SE	t	F	DF	p
Condition	0.20	0.8146	0.25	1.59	1, 16.60	0.22
Age Group	1.60	0.7027	2.28	7.93	1, 35.64	<0.01*
ConditionXAge Group	1.35	0.8909	1.52	2.31	1, 55.86	0.13

LMMs on trials with Accuracy = 1 only						
Condition	0.39	0.6576	0.60	0.72	1, 12.05	0.41
Age Group	1.02	0.4642	2.21	7.15	1, 31.69	0.01*
ConditionXAge Group	0.21	0.6631	0.32	0.10	1, 35.54	0.74

Table A.2

Vocabulary Assessment						
GLMMs						
Comprehension						
	β	SE	z	χ^2	DF	$P(\chi^2)$
Age Group	-0.64	0.4938	-1.31	1.73	1	0.18
Production						
Age Group	-1.10	0.9165	-1.20	1.45	1	0.22
Analysis of Covariance: Vocabulary Score & Accuracy in Metaphor Task						
GLMMs, Comprehension						
Condition	-3.22	1.442	-2.23	18.95	1	<0.05*
Comprehension	-1.28	1.175	-1.09	0.80	1	0.36
ConditionXComprehension	1.003	1.410	0.71	0.50	1	0.47
GLMMs, Production						
Condition	-1.53	1.1326	-1.35	19.86	1	<0.0001***
Production	0.25	0.7850	0.32	0.04	1	0.82
ConditionXProduction	-0.83	1.1229	-0.74	0.55	1	0.45
Kendall's tau correlations per age group: accuracy in metaphor trials only & vocabulary						
	z	r_t	p			
3-year-olds						
Comprehension	- 0.63	- 0.07	0.52			
Production	- 0.25	- 0.03	0.79			
4-year-olds						
Comprehension	0.17	0.02	0.85			
Production	- 0.88	- 0.09	0.37			

Table A.3

Alternative Naming Task						
Accuracy						
GLMMs						
	β	SE	z	χ^2	DF	$P(\chi^2)$
Condition	3.63	0.0028	125.794	9.75	1	<0.0001***
Age Group	-3.32	0.0027	-1.190	1.42	1	<0.0001***
ConditionXAge Group	-3.48	0.0028	-120.660	14.55	1	<0.0001***
Analysis of Proportion per age group: two-sided Z-Tests with continuity correction						

3-year-olds						
Condition	//	//	//	13.07	1	<0.0005***
4-year-olds						
Condition	//	//	//	1.38	1	0.23
Reaction Times						
LMMs						
	β	SE	t	F	DF	p
Condition	-0.61	1.2839	-0.47	2.26	1, 8.64	0.16
Age Group	3.40	0.9284	3.66	13.49	1, 36.87	0.0007**
ConditionXAge Group	-2.43	0.9327	-2.61	6.84	1, 43.77	<0.05*
Simple effects ANOVA						
3-year-olds						
Condition	-1.52	0.661	- 2.03	5.22	1, 9.92	<0.05*
4-year-olds						
Condition	- 0.30	0.642	- 0.47	0.22	1, 9.54	0.64
Kendall's tau correlation on difference reaction times (New Term – Same Term)						
	z	r_t	p			
Age Group	- 3.18	- 0.21	<0.001**			

Table A.4

Analogy Perception Task			
Kendall's tau correlation			
	z	r_t	p
Age Group	1.32	0.09	0.18
Error Analysis: Chi-squared tests with Yate's continuity correction			
	χ^2	DF	$P(\chi^2)$
3-year-olds			
Control choice	0.04	1	0.84
4-year-olds			
Control choice	15.04	1	<0.0001***
3- vs. 4-year-olds			
Age Group	4.38	1	<0.05*

Table A.5

Predictors of Metaphor Comprehension						
GLMMs: Accuracy in Metaphor Task ~						
	β	SE	z	χ^2	DF	$P(\chi^2)$
Condition	0.032	0.003347	9.9	95.06	1	<0.0001***
Alternative Naming Accuracy	3.33	0.003348	994.7	988.976	1	<0.0001***
Alternative Naming Difference reaction times	0.12	0.003343	36.4	1.326	1	<0.0001***
Analogy Accuracy	1.87	0.003348	560.7	314.334	1	<0.0001***
ConditionXAlternative Naming Accuracy	-2.37	0.003347	-709.8	503.839	1	<0.0001***

ConditionXAlternative Naming Difference reaction times	-0.018	0.003343	-5.7	32.087	1	<0.0001***
ConditionXAnalogy Accuracy	0.041	0.003347	12.4	153.518	1	<0.0001***

Models Comparisons: Likelihood Ratio test

Model 0: Met_Acc ~ 1 + (1|Subject)
Model 1: Met_Acc ~ Cond + Predictors + Cond*Predictors + (1|Subject)

	DF	AIC	BIC	χ^2	DF	P
Model 0	2	375.38	382.80			
Model 1	9	333.10	366.49	56.284	7	<0.0001***