



IUSS

Scuola Universitaria Superiore Pavia

University School for Advanced Studies IUSS Pavia

Vita-Salute San Raffaele University Milan

PH.D. PROGRAM IN COGNITIVE NEUROSCIENCE  
AND PHILOSOPHY OF MIND

XXXVII Cycle

Doctoral Dissertation

“IT TAKES TIME TO BE A BUTTERFLY”:

METAPHOR DEVELOPMENT FROM EARLY ACHIEVEMENTS TO  
LATER EFFECTS IN FRAMING ENVIRONMENTAL AWARENESS

**Candidate:** Chiara Pompei

**Supervisor:** Prof. Valentina Bambini

**Co-supervisor:** Dr. Luca Bischetti

Academic year 2023-2024

« [...]These scholars, who tend to stress the daringness, spontaneity, and invention involved in the creation of tropes, see the years of early childhood as a time of *poetic genius*. Indeed, in their admiring eyes, the process of metaphoric development is often marred by dissolution, with the older child straining to recapture what was effortlessly available to the innocent young.»

(Gardner & Winner, *The Development of Metaphoric Competence: Implications for Humanistic Disciplines*, 1978)



A one-and-a-half-year-old girl holding a pen for the first time.

The image rights of the minor fall within the scope of parental authority who provide consent for publication.



## ABSTRACT

The developmental trajectories of metaphorical skills are currently outlined in a piecemeal fashion. While it is well-known that in middle childhood metaphorical abilities are not fine-tuned yet, earlier achievements are not clearly charted. Moreover, metaphors are considered as possible tools for crafting messages related to scientific debates and social issues and it is unclear whether this is true for children and to what extent. The overarching goal of this dissertation is to detail metaphor development in childhood by providing answers to the following questions: How does receptive and expressive metaphorical competence develop in early childhood? Which cognitive and socio-cognitive factors are related to metaphorical skills in the children's cognitive architecture? Can metaphors be used to drive proactive sustainable behavior? This dissertation presents the results of four studies that, combining methods from developmental psychology and psycholinguistics, investigated the developmental trajectories of metaphorical skills in both the comprehension and production dimensions in early and middle childhood, as well as the use of metaphors for persuasive and conceptual purposes and in particular to influence attitudes and behavior toward environmental awareness. In *Study 1*, we aimed at mapping the development of metaphor understanding in early childhood in a sample of 176 children aged 4 to 6. To assess metaphor comprehension skills, we adapted the Physical and Mental Metaphors (PMM) task, in an age-appropriate multiple-choice fashion. By distinguishing the interpretation of physical (i.e., metaphors capitalizing on perceptual properties, e.g., Dancers are feathers) and mental (i.e., metaphors capitalizing on mental properties, e.g., The teacher is an icicle) metaphors, results showed that mental metaphors remain challenging during early childhood and that children need to master Theory of Mind to appreciate these metaphors. Conversely, physical metaphors are easier to understand, and their comprehension is supported by vocabulary skills. However, children start to grasp physical metaphors only at 6 years of age. The struggle in metaphor comprehension at earlier ages revealed that children are engaged in climbing over a literal step. *Study 2* investigated metaphor production abilities in 171 children, adopting an elicitation task where children were asked to produce metaphors given literal prompts (e.g., describing a dog as being white without using the word white). Results revealed that metaphor production abilities increase between 4 and 5 years of age, reaching an almost adequate level at 6 years of age. Moreover, while vocabulary boosts metaphor production throughout development, mind-

reading skills refine these abilities later on. Furthermore, adopting approaches from Natural Language Processing (NLP), namely computing the semantic distance between the literal prompt (e.g., white dog) and the metaphor produced (e.g., cloud dog), we observed that as age increases children use words more distant to the literal prompt as metaphors vehicle, reflecting broadening semantic network exploration. Notably, the exploitation of the semantic network for metaphorical purposes is promoted by Theory of Mind abilities, rather than vocabulary size. *Study 3* had the goal of assessing the contribution of metaphorical competence to specific aspects of social communication, namely persuasion. Adapting a test for evaluating persuasive skills in early childhood in Italian (e.g., Try to convince Mattia to eat broccoli), we studied the type of arguments provided (i.e., positive, negative, and psychological) as a function of cognitive and metaphorical skills. Results showed that the ability to produce metaphors supports children in selecting psychological persuasive arguments (e.g., Broccoli is candy), at the expense of more basic arguments such as rewards or punishments. Notably, metaphor comprehension abilities did not explain the quality of persuasive argumentation. *Study 4* examined the conceptual function of metaphors during middle childhood and built an innovative tool for fostering environmental awareness via metaphors. A novel and enriched version of the MetaCom training program was developed, named the MetaCom-Green, targeting metaphorical competence while presenting climate change-related issues. A control group was involved in a comparable training program on the same topics, promoting climate change awareness without using metaphors. Results showed that the MetaCom-Green group outperformed the control group not only in metaphor understanding, but also in general reading comprehension skills and climate change related knowledge, suggesting that training metaphorical competence may promote knowledge enhancing inferential skills. Moreover, the MetaCom Green training proved effective at modulating those psychological aspects contributing to climate change awareness, namely increasing feelings of hope toward the phenomena, decreasing despair and promoting pro-environmental behaviors. Overall, the results of this thesis provide a comprehensive picture of the development of metaphorical skills and their use for social purposes. During early childhood important acquisitions take place for the later refinement of metaphorical skills. These achievements are supported by other linguistic, cognitive and socio-cognitive abilities that contribute to a different extent according to the dimension investigated (comprehension vs. production), the type of metaphor (physical vs. mental) and the stage of development. At later stages, when metaphor

comprehension is more settled, metaphor can take an important role in the social world, being crucial for becoming an effective communicator, for persuasive intents, and especially for navigating complex topics relevant to education and social debates. In addition, this work provides the scientific landscape with a tool for including children in green policies, the MetaCom-Green, positively influencing their language and social functioning as well as their pro-environmental behavior.



# TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xiv
<b>ACKNOWLEDGEMENTS</b>	<b>xvii</b>
<b>INTRODUCTION</b>	<b>17</b>
1. <i>The ontogenesis of metaphors: on the acquisition of figurative meanings</i>	17
2. <i>How to do things with metaphors: linguistic, cognitive, and psychological effects</i>	21
3. Aims and thesis roadmap	22
<b>STUDY 1: Climbing the literal and physical steps: developmental trajectories of metaphor comprehension in early childhood</b>	<b>26</b>
<b>1.1. Introduction</b>	<b>27</b>
1.1.1. The present study	31
<b>1.2. Methods</b>	<b>32</b>
1.2.1. Participants	32
1.2.2. Procedure	32
1.2.3. Measures	32
1.2.2.1. <i>The Physical and Mental Metaphor task - multiple-choice</i>	32
1.2.2.2. <i>Vocabulary</i>	35
1.2.2.3. <i>Working Memory</i>	35
1.2.2.4. <i>Theory of Mind</i>	35
1.2.3. Statistical Analysis	36
1.2.3.1. <i>Accuracy Analysis</i>	37
1.2.3.2. <i>Response Types Analysis</i>	38
<b>1.3. Results</b>	
1.3.1. Developmental trajectories of metaphor comprehension: analysis of Accuracy and Error Types	39
1.3.1.1. <i>Accuracy Analysis</i>	39
1.3.1.2. <i>Response Types Analysis</i>	43
1.3.2. Interpretation Analysis	43
1.3.3. Cognitive substrates of metaphor understanding across types	44
<b>1.4. Discussion</b>	<b>45</b>

<b>STUDY 2: How a white dog becomes a cloud: a cross-sectional and distributional semantics investigation of metaphor production skills in early childhood</b>	<b>56</b>
<b>2.1. Introduction</b>	<b>56</b>
2.1.1. The present work	59
<b>2.2. Methods</b>	<b>60</b>
2.2.1. Participants	60
2.2.2. Procedure	60
2.2.3. Measures	61
2.2.3.1. <i>Vocabulary</i>	61
2.2.3.2. <i>Inhibitory control</i>	61
2.2.3.3. <i>Theory of Mind</i>	61
2.2.3.4. <i>Metaphor Comprehension</i>	62
2.2.3.5. <i>Metaphor Production</i>	63
2.2.4. Statistical analysis	64
<b>2.3. Results</b>	<b>66</b>
2.3.1. Development of Metaphor Production skills and the role of individual differences	68
2.3.2. Effects of Age and Individual Differences on Semantic Similarity	70
2.3.3. Comparison with adults	72
2.3.4. Exploratory analysis with Generalized Additive Models	72
<b>2.4. Discussion</b>	<b>76</b>
<b>STUDY 3: ‘Broccoli is candy’: The role of metaphors in children’s persuasive communication</b>	<b>82</b>
<b>3.1. Introduction</b>	<b>83</b>
3.1.1. The present study	87
<b>3.2. Methods</b>	<b>87</b>
3.2.1. Participants	87
3.2.2. Procedure	88
3.2.3. Measures	88
3.2.3.1. <i>Vocabulary</i>	88
3.2.3.2. <i>Working Memory</i>	89
3.2.3.3. <i>Metaphor Comprehension</i>	89
3.2.3.4. <i>Metaphor Production</i>	90
3.2.3.5. <i>Persuasion</i>	90
3.2.4. Statistical analysis	91
<b>3.3. Results</b>	<b>94</b>
3.3.1. Developmental effects of linguistic and cognitive covariates	96
<b>3.4. Discussion</b>	<b>101</b>

<b>STUDY 4: Metaphors as the earth’s parachute: the efficacy of the MetaCom-Green program in fostering environmental awareness in children</b>	<b>109</b>
<b>4.1. Introduction</b>	<b>111</b>
4.1.1. The present study	113
<b>4.2. Methods</b>	<b>114</b>
4.2.1. Participants	115
4.2.2. Design and procedure	115
4.2.3. Assessment	116
4.2.3.1. <i>MacArthur Scale of Subjective Social Status</i>	116
4.2.3.2. <i>Vocabulary</i>	117
4.2.3.3. <i>Working Memory</i>	117
4.2.4. Focus variables	118
4.2.4.1. <i>Theory of Mind</i>	118
4.2.4.2. <i>Reading Comprehension</i>	118
4.2.4.3. <i>Physical and Mental Metaphors task</i>	119
4.2.4.4. <i>Climate Change-related attitudes</i>	120
4.2.4.4.1. <i>Hope scale</i>	121
4.2.4.4.2. <i>Despair scale</i>	121
4.2.4.4.3. <i>Pro-environmental behavior scale</i>	121
4.2.4.5. <i>Climate Change Knowledge scale</i>	122
4.2.4.6. <i>Connection to Nature Index</i>	122
4.2.5. Training programs	123
4.2.5.1. The MetaCom-Green	123
4.2.5.1.1. Guided-play session	126
4.2.5.2. Control training	127
4.2.6. Statistical Analysis	128
<b>4.3. Results</b>	<b>129</b>
4.3.1. Descriptive Statistics	129
4.3.2. Main Analysis	131
<b>4.4. Discussion</b>	<b>133</b>
<b>CONCLUSION</b>	<b>137</b>
1. The development of metaphors in early childhood: the <i>chrysalis years</i>	137
2. Milk pragmatics: toward a novel developmental account of metaphorical meanings	141
3. Future directions	143
<b>BIBLIOGRAPHY</b>	<b>203</b>
<b>APPENDIX</b>	

## LIST OF TABLES

### STUDY 1:

<i>Table 1.1. Descriptive Statistics of children's performance in each age group</i>	23
<i>Table 1.2. Output of the Generalized Additive Model with Accuracy as dependent variable and age as predictor</i>	25
<i>Table 1.3. Output of the Generalized Additive Model with Interpretation as dependent variable and Age and Accuracy as predictors.</i>	29
<i>Table 1.4. Output of the Generalized Additive Model with Accuracy as dependent variable and cognitive predictors</i>	31

---

### STUDY 2:

<i>Table 2.1. Descriptive Statistics of children's performance in each age group.</i>	67
<i>Table 2.2. Output of the Linear Model with Metaphor Production Accuracy as the dependent variable.</i>	69
<i>Table 2.3. Output of the Linear mixed-effects model with Semantic Similarity as the dependent variable.</i>	70

---

### STUDY 3:

<i>Table 3.1. Descriptive Statistics of children's performance in each age group.</i>	93
<i>Table 3.2. Output of the Generalized Linear mixed-effects model with the Persuasion scores as the dependent variable</i>	98

---

### STUDY 4:

<i>Table 4.1. Descriptive statistics of the two training groups at pre- and post-training on control and dependent variables.</i>	127
<i>Table 4.2. Independent sample t-tests comparing the two training groups on control and dependent variables at pre-training.</i>	129

## LIST OF FIGURES

### STUDY 1:

<i>Figure 1.1.</i> EXAMPLES OF PHYSICAL AND MENTAL METAPHORS ITEMS	33
<i>Figure 1.2.</i> CORRELOGRAM BETWEEN SCORES IN THE PHYSICAL AND MENTAL METAPHORS ACCURACY AND LINGUISTIC, AND PRAGMATIC VARIABLES	38
<i>Figure 1.3.</i> PHYSICAL AND MENTAL METAPHORS COMPREHENSION	40
<i>Figure 1.4.</i> EFFECT PLOTS OF CHANGES IN TYPE OF ANSWERS WITH AGE	41
<i>Figure 1.5.</i> MENTAL INTERPRETATIONS FOR ACCURATE AND INACCURATE ANSWERS	43
<i>Figure 1.6.</i> ASSOCIATION BETWEEN VOCABULARY AND PHYSICAL AND MENTAL METAPHORS ACCURACY	45
<i>Figure 1.7.</i> ASSOCIATION BETWEEN COGNITIVE AND SOCIO-COGNITIVE VARIABLES AND PHYSICAL AND MENTAL METAPHORS ACCURACY	46

---

### STUDY 2:

<i>Figure 2.1.</i> CORRELOGRAM BETWEEN METAPHOR PRODUCTION SCORES, SEMANTIC SIMILARITY VALUES, AND LINGUISTIC, COGNITIVE, AND SOCIO-COGNITIVE VARIABLES	64
<i>Figure 2.2.</i> DISTRIBUTION OF SEMANTIC SIMILARITY ON METAPHOR PRODUCTION ACCURACY LEVELS	65
<i>Figure 2.3.</i> ASSOCIATION BETWEEN METAPHOR PRODUCTION AND LINGUISTIC AND SOCIO-COGNITIVE PREDICTORS	66
<i>Figure 2.4.</i> SEMANTIC NETWORKS REPRESENTATION OF THE COSINE MATRIX AMONG LITERAL PROMPT AD VEHICLES PRODUCED ACROSS AGE GROUPS	68
<i>Figure 2.5.</i> METAPHOR PRODUCTION SKILLS AND SEMANTIC SIMILARITY VALUES ACROSS AGE GROUPS	69
<i>Figure 2.6.</i> VISUALIZATION OF THE EXPLORATORY ANALYSIS	72

---

### STUDY 3:

<i>Figure 3.1.</i> SCHEMATIC REPRESENTATION OF THE PERSUASION TASK ADAPTED FROM PETERSON ET AL., (2018)	90
<i>Figure 3.2.</i> BOX PLOT OF THE LINGUISTIC, COGNITIVE, AND SOCIO-COGNITIVE MEASURES FOR EACH AGE GROUP	92
<i>Figure 3.3.</i> CORRELOGRAM BETWEEN PERSUASION SCORES AND LINGUISTIC, AND PRAGMATIC VARIABLES	93
<i>Figure 3.4.</i> DEVELOPMENT OF PERSUASIVE STRATEGIES AND FOCUS ON THE USE OF METAPHORS	95
<i>Figure 3.5.</i> ASSOCIATION BETWEEN PERSUASIVE STRATEGIES AND COVARIATES	98

---

### STUDY 4:

<i>Figure 4.1.</i> COMPUTERIZED PROCEDURE FOR CLIMATE CHANGE-RELATED ATTITUDES ASSESSMENT	118
<i>Figure 4.2.</i> PICTURE MATCHING TASK	122

<i>Figure 4.3. MATERIAL FOR THE GUIDED-PLAY SESSION</i>	123
<i>Figure 4.4. CORRELATION BETWEEN OUTCOMES AND COVARIABLES AT PRE-TRAINING</i>	127
<i>Figure 4.5. EFFECT PLOT OF CHANGES IN ACCURACY LEVELS FOR THE PMM TASK</i>	129
<i>Figure 4.6. EFFECTS OF THE METACOM-GREEN TRAINING ON FAR TRANSFER AND GREEN VARIABLES</i>	130

---



## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Prof. Valentina Bambini, who has been a caring mentor. She provided me with a true *zone of proximal development*, and I owe her a lot.

I want to acknowledge the contribution of all my colleagues in the Neurolinguistics and Experimental Pragmatics (NEP) Lab at IUSS who offered their expertise for the studies included in this thesis. Some of them deserve a special mention. I want to express my gratitude to Dr. Luca Bischetti, my co-supervisor. Without witnessing his passion, five years ago, I would have never considered doing research. I also want to thank Dr. Paolo Canal for being an invaluable source of knowledge and inspiration as a scientist and human being. Finally, I could not feel any luckier having two of the most brilliant minds and generous hearts as colleagues and friends, Dr. Federico Frau and Veronica Mangiaterra, who have been by my side when work - or life - got tough.

I am also thankful to the Fondazione Treccani Cultura, in particular my tutor, Loredana Lucchetti, for providing me with inspiring training on scientific dissemination, especially with the young population.

I want to thank all the people from other institutions who collaborated with us to realize the studies presented in this work, particularly Prof. Serena Lecce (University of Pavia) for sharing her knowledge with me and providing insightful comments that crucially inspired this work.

I am also grateful to all the schools and the teachers who joined my research projects. I am deeply thankful to Gaia Gentili, Giuseppe Marchetti and all the teachers from the Comprehensive Institute Cupra Marittima and Ripatransone, as well as to Nicoletta Guglielmi, Letizia Galandra, Raffaella Grignani, Laura Mariani, Sabrina Fiorani, Maria Mazzoni, and all the teachers from the San Giorgio Institute in Pavia, for their collaboration and trust. I promise that my research will be always moved by the same love and care that you devote to children.

Importantly, thanks to my little scientists, namely the hundreds of children I worked with, who are the real mentors of this work. No elaborate technology, artificial intelligence, or sophisticated method enchanted me more than you did.

Finally, I want to thank the ones who loved me and believed in me before I was a *butterfly*, when my anger led me down painful paths. If I am now privileged, it is because your trust offered me a better fortune.

This work was conducted with financial support from the FSE-REACT-EU programme, PON “Ricerca e Innovazione” 2014-2020 (D.M. 1061/2021).

# INTRODUCTION

## 1. The ontogenesis of metaphors: on the acquisition of figurative meaning

What makes a smile different from a yawn? Infants navigate the social world actively, immediately after birth, thanks to a conspicuous range of behaviors capable of eliciting feedback from their caregivers. However, only those behaviors capable of evoking salient and relevant responses from adults survive across the lifespan (Rochat, 2007; Yoder & Warren, 1998), becoming persistent, intentional communicative tools, as is the case with the smile. This indefeasible mechanism governs language acquisition and suggests that children learn *meaning* in the cradle of interaction.

Imagine a child uttering *Spaghetti!*. If the child refers to her meal, she performs a correct - and literal - association of a word to a referent. But what if the child is pointing to her classmate's hair?

*Figurative meanings*, particularly *metaphors*, are fascinating aspects of human communication that children experience early on. Adults often use metaphors while interacting with children: a caregiver might refer to her toddler as a *little star*, and a teacher would persuade his students that books are *goldmines*. However, figurative meanings carry a certain degree of complexity, as suggested by a cornucopia of studies coming from Experimental Pragmatics and Neuropragmatics (Bambini & Frau, 2022). Specifically, metaphors are built upon a complex cognitive and socio-cognitive platform, are extremely fragile in psychiatric and neurological conditions, and draw in the brain their own neurophysiological pattern (Bambini, Pompei, & Scalingi, 2022). While obtaining a certain degree of agreement on the trajectory of acquisition of several phenomena (e.g., scalar implicatures, Wilson & Katsos, 2020), for other phenomena, such as metaphor, the developmental arm of pragmatics is facing the challenge of integrating studies reporting inconsistent developmental starting points, with the paradox of being separated by a gap of ten years (Cacciari & Padovani, 2012; Pouscoulous & Tomasello, 2020). I argue that the cause of the fueling of this rift is threefold. First, these studies lacked a clear definition of *metaphor* and *metaphorical skills*. Furthermore, while pragmatic research in fields other than development was beginning to

implement standardized tests strongly grounded in neuropsychology, and using sophisticated statistical methods to analyze the collected data, the study of metaphor acquisition remained anchored to tasks with major psychometric problems and weak analysis tools. Finally, the lack of a true developmental perspective has led studies to pay little attention to the contribution of other cognitive and socio-cognitive abilities, which have focused on *when* neglecting *how* metaphors gain a foothold in children's communicative repertoire. In this introduction, I will briefly retrace the key events that characterized the study of metaphor in acquisition, while trying to account for the psychometric, methodological, and individual difference issues in the experimental works presented in this dissertation.

Within the framework of Relevance Theory, *metaphors* stand at the far end of a continuum of word uses that goes from literal to figurative meanings (Sperber & Wilson, 2008). Metaphor constitutes a case of loose use of words (in a continuum with approximation, category extensions, hyperbole, and other phenomena) in which the intended meaning is not blatantly enclosed in the utterance, but rather is derived on the basis of inferential processes capable of integrating lexical, contextual and psychological information (Carston, 2010; Wilson & Carston, 2006). To make a long story short, metaphor is the use of a word that moves apart – to some extent – from its literal, conventional use. The nuanced nature of these meanings makes it hard to determine the minimum unit for a metaphor to be defined as such, and this has been the first challenge for the study of figurative language development.

Early observational studies reported that during pretend play, children may refer to their airplane as a *bird-car* or call *egg-ball* a ball with an ovoid form (Barrett, 1986). These expressions are usually produced within the communicative strategy of relying on familiar words to fill a gap, capitalizing on perceived similarities, and might vanish as soon as the child acquires the requisite word (e.g., *airplane* for the *bird-car*; Clark, 2020). Intriguingly, such over-extensions caught the eye of scholars, who glimpsed in these expressions the first signs of *figurative meaning* (Billow, 1981; Winner, 1979). Starting from this arm of over-extensions, a conspicuous number of studies tried to investigate the

developmental trajectory of metaphors. After initial observational studies, scholars tried to develop simple tasks to investigate metaphorical production (Gardner et al., 1975), asking children to complete some vignettes to find an ending for a simile:

*Things don't have to be huge in size to look that way.*

*Look at that boy standing over there. He looks as gigantic as ...*

Children as young as 4 years produced a high number of metaphors, also compared to their older peers. However, most of those metaphors were inappropriate (e.g., *gigantic as a clock* or *sad as a shirt*), questioning a genuine metaphorical use of these expressions (Gardner et al., 1975). More importantly, these authors warned about an unusual phenomenon. In facing a renaming task, while very young children often performed metaphorical renaming (e.g., *a park with grass* for a hairbrush), school-aged children showed a sort of tension in dealing with this task, strictly preferring to *call things by their name* (Gardner & Winner, 1978). In the same years, the very first studies on metaphor comprehension, which asked children to verbally explain the meaning of metaphors, described a different story: young children are unable to produce appropriate explanations until young adolescence (Winner et al., 1976). The U-shaped pattern described by studies on production, paired with the persistent inability to comprehend metaphors until later stages, elicited important doubts about the genuine *metaphoricity* of younger children's expressions. These doubts were further justified by contemporary evidence on the study of child cognitive development, which started deriving important cognitive implications from the systematic study of pretend play. Specifically, psychologists concluded that pretend play seemed to free children to participate in other realities that temporarily override the real world (Lillard, 1993) and, within this framework, children may flexibly turn a pen into a *rocket* for pretense purposes. These considerations sharply raised the question of whether the ease in renaming objects would count as a metaphorical ability or as a reflection of the tendency to override reality (Gardner & Winner, 1978). For decades, metaphorical skills were described as a late achievement (Cacciari & Padovani, 2012).

However, some authors didn't give up on the idea that metaphors flourish very early in development, holding tightly to early observations of spontaneous metaphor generation (Pouscoulous, 2011), despite – intriguingly – discontinuing the study on metaphor production. Crucially, this perspective provided the literature with new evidence on metaphor comprehension: children as young as three years can understand metaphorical expressions when tested with age-appropriate tasks (Di Paola et al., 2020; Pouscoulous & Tomasello, 2020). These new tasks promised to catch early sparks of metaphor comprehension while capturing pure inferential abilities (Pouscoulous, 2011) and were structured as *act-out* tasks, in which children were asked to point to *The tower with the hat* choosing between a tower with a pointy roof and a tower with a balcony (Pouscoulous & Tomasello, 2020). These studies have the merit of having brought to light important methodological limitations of previous tasks, such as the high metalinguistic load of verbal explanation tasks (Winner et al., 1976) and the use of conventional metaphors, which rely on more experience-based lexical knowledge (Cacciari & Padovani, 2012). Nevertheless, it is not clear whether act-out tasks actually require *pragmatic inference*, namely the mechanism that allows for integrating information with the context to understand the intended meaning (Sperber & Wilson, 2002). Specifically, items like *The tower with the hat* seem to resonate with some children's production and over-extensions and highly capitalize on the *shape bias* (i.e., the idea that objects sharing the same shape also share the same name; Diesendruck & Bloom, 2003). Children may solve these tasks by accepting over-extensions (e.g., *hat*) as labels for unfamiliar objects (e.g., *pointy roof*). Against this argument, authors of the early stance claim that this kind of over-extensions might be an *intentional* misuse of words when the correct word is not possessed (Pouscoulous, 2011). Hence, according to this view, children would exploit the same underlying mechanisms of adults' figurative uses described by Sperber and Wilson (2008): *The ad hoc concepts built on the basis of most metaphorical terms are genuinely ad hoc: that is, they are adjusted to the precise circumstances of their use and are therefore unlikely to be paraphrasable by an ordinary language expression.*

At this point it seems clear to me that a large portion of literature has approached the ontogenesis of metaphor using the same lenses used to investigate metaphor processing in the adult. In this fuzzy picture, the need for a developmental account of pragmatics cries out.

Within this work, I will try to bring empirical evidence on how children navigate the metaphorical extreme of the literal/non-literal continuum. To do this, I will resume the study of metaphor acquisition in both receptive and expressive modalities, building original assessment and training materials, as well as employing innovative and adequate statistical methods.

## **2. *How to do things with metaphors: linguistic, cognitive, and psychological effects***

To frame, to gain, to illuminate. Metaphors are not linguistic trappings for human beings, rather, they serve a variety of cognitive and psychological functions. In 1996, Mio and Katz listed the main applications of metaphors, spanning from their conceptual functions to emotional ones. I will briefly review these applications, integrating previous considerations with recent evidence.

Initial insights on the power of figurative language come from the work of Lakoff and Johnson (1980) who, through the lens of Cognitive Linguistics, described metaphorical associations as conceptual mappings reflecting human experience. Along these lines, a conspicuous body of research has pursued the study of metaphors as conceptualization tools capable of driving our thoughts and behaviors, particularly in uncertain situations (Flusberg et al., 2017; Hanne & Hawken, 2007; Scherer et al., 2015; Thibodeau & Boroditsky, 2011, 2013). Complementary, metaphors seem to solve strategic functions such as increasing the memorability of concepts (Beard et al., 2024; Katz, 1992) as well as improving the social background during conversation, increasing intimacy and Theory of Mind (Bowes & Katz, 2015). Metaphors also play a major role in persuasion (Sopory & Dillard, 2002b), populating advertising, which capitalize on the multimodal nature of these figurative meanings (McQuarrie & Phillips, 2005; Pérez-Sobrino, 2016). Moreover, in spite of common sense, metaphors inhabit science as much as poetry: figurative expressions are brief, effective tools to communicate technical concepts (Deignan & Semino, 2019; Knudsen, 2003; Ortony, 1975). Finally, metaphors are recognized as a tool to the access of mental world also in

clinical settings, where they are defined as the *language of change* (Muran, 1990; Witztum et al., 1988). The difficulty in tracing the developmental trajectories of metaphors, as well as the lack of a developmental perspective, hinder information on when metaphors are converted from developmental challenges into linguistic, cognitive, and psychological tools. If it is true that metaphors are *necessary and not just nice* (Ortony, 1975), it seems crucial for the study of metaphor acquisition to be oriented not only in the understanding of when a child is able to utter or comprehend a metaphor *per se*, but how and when it is possible to *do things with metaphors*. This work delves into this gap, trying to provide initial empirical evidence on when children begin to benefit from metaphorical skills for socio-communicative purposes (i.e., persuasion) and deep conceptualization of complex issues.

### **3. Aims and thesis roadmap**

The present dissertation discusses the results of four experimental works, motivated by the ambition of providing current literature with new theoretical insights on metaphor competence as well as novel assessment, training, and statistical tools for a robust renovation of the study of metaphor acquisition.

Following this ambition, the aim of this work was two-fold. The first aim was to investigate the metaphorical abilities in early childhood in both receptive and expressive modalities, by testing which developmental phases are crucial for figurative meaning and which other linguistics, cognitive, and socio-cognitive skills support this acquisition. The second aim of the work was to further explore when children can start exploiting metaphors for specific communicative (i.e., persuasion), conceptual (i.e., learning), and psychological (i.e., attitudes modulation) purposes.

To reach these overarching goals, attention has been drawn to several aspects. First, I considered the insights from the study of psychometric properties of tasks in the developmental population (Haladyna et al., 2019; Rodriguez, 2005; Williams et al., 2014), as well as the guidelines and principles for the assessment and training of fragile individuals (American Psychological Association, 2022). Moreover, I have adopted statistical approaches apt to detect possible *wobbles*,

*bumps, and sudden jumps* during acquisition (van Dijk & van Geert, 2007), following the most recent applications of Additive modeling (Van Rij et al., 2022; Wieling, 2018), and combining it with more conservative approaches, without assuming apriori either the linearity of developmental patterns, or linear relationships with other cognitive and socio-cognitive variables. Along these lines, it was not possible to neglect the role of Individual Differences in linguistic, cognitive and socio-cognitive variables, as deeply intertwined with pragmatics skills in the lifespan and in fragile populations (Ceccato et al., 2025; Frau et al., 2024; Lecce et al., 2019; Matthews et al., 2018).

The empirical arm of this work is built upon four studies:

- **Study 1** aimed to adapt a multiple-choice tool to evaluate the developmental trajectory of physical and mental metaphor comprehension in early childhood. The study brings attention to the possibility that children face a multistep developmental trajectory to grasp different metaphor types, characterized by the literal and the physical steps. Individual differences in vocabulary and Theory of Mind play a role during this climbing.
- **Study 2** investigated the developmental trajectories of metaphor production combining cross-sectional and distributional semantics methods. The study provides novel information on metaphorical expressive abilities in early childhood, with distributional semantics being diagnostic of children's semantic network exploration for metaphorical purposes. Comparing the performance of children and two groups of adults revealed that while the developmental endpoint of productive skills is a long way to go, at 6 years of age children already find the adequate semantic distance to cover for optimal metaphor generation. An exploratory analysis via additive modeling disentangled the complex non-linear relationships between metaphor production and cognitive and socio-cognitive substrates, alongside its relationship with metaphor comprehension.
- **Study 3** aimed to detect the first advantages of mastering metaphors, examining children's persuasive communication development. The study contributes to delineating the profile of

the young *metaphorical thinker*, who employs more refined persuasive strategies than his peers, relying on metaphor production for psychological negotiation.

- **Study 4** approached the most high-level nature of metaphors, namely their power to shape thoughts and guide behavior during complex situations. Providing climate change communication literature with a tool for enhancing metaphor skills and pro-environmental behavior, this work highlights that children can benefit from metaphors' cognitive and psychological effects only at later stages and when appropriately accompanied during metaphor sense-making.

Across these four studies, the experimental work incorporated data collected using behavioral tasks and undergoing distributional semantic analysis. The methodological and theoretical background of this dissertation is rooted in the field of Developmental Pragmatics (Matthews, 2014) and, more broadly, Psycholinguistics (Traxler & Gernsbacher, 2011) and Experimental Pragmatics (Noveck, 2018). However, Developmental Psychology profoundly contaminated this collection of studies, and several disciplines, such as Applied Linguistics, Educational Psychology, and Distributional Semantics contributed to the transdisciplinary profile of this work.

The most relevant consideration has to do with the population involved in the study, and with the psychological nature of metaphor. During data collection, I have witnessed the emergence of delicate and precious cores of children's inner world. In research practice, these emotional overflows are often considered side effects of the assessment in the experimental setting. However, these eventualities have profoundly guided the interpretation of the data presented in this dissertation and strengthened my conception of metaphors as irreducible bridges to the child's symbolic life. Some of the children I have worked with have reported to me abuse, traumas, and strong emotions experienced in the shadow of domestic and school roofs. While I believe that children's emotional sharing is fostered by proper communication practice and the creation of a safe environment (that cannot go without constant training and a genuine gravitational shift toward the child), I attribute to metaphors the residual degree of causality. The majority of these delicate

conversations occurred during the testing of metaphorical competence. While activating channels of sharing, metaphors also helped me in giving something back to children, in a non-clinical context, in which *we cannot open drawers we cannot close*.<sup>1</sup> Hence, this is a work about metaphors, played with metaphors.

---

<sup>1</sup> Is a typical expression used by clinicians to refer to those situations in which there is not enough time to properly clinically address a patient need (i.e., a patient that we see just once or, precisely, in the experimental setting). In these cases, to make a long story short, the general guideline is not to go through that need, limiting to an active listening.

# STUDY ONE

## CLIMBING LITERAL AND PHYSICAL STEPS: DEVELOPMENTAL TRAJECTORIES OF METAPHOR COMPREHENSION IN EARLY CHILDHOOD<sup>2</sup>

### Abstract

Attempts to describe the early acquisition of metaphor skills led to inconsistent results, possibly due to differences in metaphor type. We implemented a multiple-choice version of the Physical and Mental Metaphors task, accounting for literal and figurative interpretations of different metaphor types. The task was administered to 175 children aged 4 to 6, alongside verbal and socio-cognitive tests. The analysis revealed a curvilinear trend, with the age of six marking a significant improvement in physical, but not mental, metaphors, driven by a drop of literal interpretations. A further analysis limited to mental metaphors indicated that children at this age still tend to prefer physical interpretations. In terms of cognitive substrates, vocabulary skills supported physical metaphors, while theory of mind aided mental metaphor comprehension. Overall, results highlight that overcoming literal, and physical preferences are two important stages in the developmental trajectory of metaphor comprehension, interacting with verbal and socio-cognitive factors.

---

<sup>2</sup> This chapter is a manuscript currently in preparation for submission to a peer review journal with the following provisional authorship “Lecce, S.\*, Pompei, C.\*, (co-first authors), Del Sette, P., Canal, P., Bischetti, L., & Bambini, V., *Climbing physical and literal steps: developmental trajectories of metaphor comprehension in early childhood*”.

## 1.1 Introduction

*Are metaphors departures from a norm of literalness?* This question opens the reflections of Sperber and Wilson (2008) on the pragmatic nature of metaphors. According to Relevance Theory, utterances can be ascribed on a continuum moving from literal uses (e.g., *That candy is sweet*) to metaphorical meanings (e.g., *Those players are elephants*). What happens along this continuum is a progressive increase in the freedom of interpretation of the utterance left to the receivers, departing from the literal concept (Sperber & Wilson, 2008). In facing metaphors such as *Those players are elephants*, the receiver constructs a novel *ad hoc* concept that incorporates the emergent property of *clumsiness*, which is not blatantly enclosed neither in the concept of *players*, nor in the one of *elephant*, capitalizing on inferential processes capable of integrating lexical meaning, contextual information and, possibly, information regarding the interlocutor's mental state (Carston, 2010a; Sperber & Wilson, 2002). While conventionalized metaphorical meanings (e.g., *Traffic jam*) are strongly lexicalized and load less on these processes, more novel metaphors (e.g., *That wound is a fjord*) are close to the pole of the continuum, being associated with a wide range of weak implicatures.

How do children learn to move along this continuum? Evidence from language development studies suggests that the literal/metaphorical distinction may also reflect different stages of acquisition toward metaphorical competence (see Clark, 2020). For instance, some studies on metaphor comprehension reported that figurative meanings are acquired later than literal ones, as suggested by a preference for more literal interpretations at early stages, with some metaphors being thorny until middle childhood (Gardner et al., 1975; Lecce et al., 2019; Winner, 1997; Winner et al., 1980). Other studies, however, disconfirm the *late* acquisition view (Di Paola et al., 2020; Pouscoulous, 2014). For instance, three years old children were reported to understand metaphors when presented with act-out tasks (e.g., for the metaphor *The tower with the hat*, being able to prefer a tower with a pointy roof over a tower with a balcony; Pouscoulous & Tomasello, 2020). Another strand of recent research capitalized on the insights on literal preferences suggested by first studies, building on the idea that young children can grasp metaphors, but when faced with literal options

they fall for more conventional interpretations over metaphorical ones (Neff et al., 2023). These works opened up the possibility that children's difficulties with metaphorical language may not be due to a lack of metaphor skills *per se*, but rather to a genuine preference for literal rather than metaphorical statements (Falkum, 2022; Long et al., 2021; Martín-González et al., 2024; Neff et al., 2023). This scenario originated a long-standing debate that has fuelled the field of experimental and developmental pragmatics (for an overview of the *late* stance see Cacciari & Padovani, 2012; for an overview of the *early* stance see Pouscoulous, 2014; for a general overview see Bambini & Domaneschi, 2024). At present, however, literature provides a fuzzy picture of the development of metaphorical skills, especially during early childhood. A deeper exploration of the error types and a shift towards a developmental perspective are needed to overcome this variability.

Among the possible reasons for these discrepancies in the literature, there is certainly the fact that metaphors vary along many dimensions, related to novelty as well as to their content. For instance, some studies (Asch & Nerlove, 1960; Winner et al., 1976) distinguished between psychological metaphors (e.g., *Daddy is a Volcano*), sometimes referred to as *abstract*, which require deriving a link between a physical (e.g., *volcano*) and a mental (e.g., *anger*) object, and physical/perceptual metaphors, which require to grasp a link between two physical ones (e.g., *My brother is a skyscraper*). This distinction is potentially loading on inferential efforts, with early studies showing earlier grasping of the physical type (Cicone et al., 1981; Winner et al., 1976) and later studies confirming these findings, in addition to pointing out a greater involvement of Theory of Mind resources for the psychological type (Del Sette et al., 2024; Lecce et al., 2019).

Moreover, studies vary in the consideration they paid to the type of answers given by children. In this respect, pioneering studies from developmental psychology suggested paying attention to error analysis (Bowerman, 1982; James, 2013). For instance, Winner, Rosenstiel and Gardner (1976) show that children, while not being metaphorical yet, can vary in the type of answers used to verbalize metaphorical meanings, ranging from magical (e.g., *The king had a magic rock and transformed the guard in another rock*, as an answer to *The prison guard was a hard rock*), to metonymic (e.g., *The guard*

*worked in a prison that had hard rock walls*) or primitive metaphorical (e.g., *The guard had a hard, though muscle*) content. Notably, a specific trend was observed for the different types of interpretations, with the magic and metonymic errors dominating between ages 6 and 7 and decreasing after 8, and the primitive metaphorical responses (or rather, particularly physical interpretations of mental contents) increasing at 8 years of age. Only at the age of 12, metaphorical answers became dominant, after dropping the more literal (magical and metonymic answers) and then the more physical (primitive) types. More recently, authors used to detect children's metaphor comprehension using forced choice tasks comparing literal vs. metaphorical answers and non-related vs. metaphorical answers, showing that, when present, the literal option is preferred but in the latter case children are able to choose metaphors, suggesting a “literal” preference (Neff et al., 2023; Neff & Falkum, 2022). Across studies, however, aspects such as the type of metaphor and the type of answer are not systematically considered.

Other sources of heterogeneity may be due to methodological aspects and specifically task properties. To begin with, verbal explanation tasks – used for instance by Winner and colleagues (1976) - allow for capturing the nuances of comprehension machinery in great detail (Kalandadze et al., 2019). However, these tasks heavily load on metacognitive skills, which might be not mature and blur the possibility of observing early metaphorical development (Pouscoulous, 2011). Act-out tasks such as those used in Pouscoulous and Tomasello (2020), while catching some precursors of metaphor understanding, might be biased in other ways that do not fully ascertain the level of metaphorical competence. Specifically, the item used in these tasks (e.g., *The tower with the bat* for a tower with a pointy roof) can be defined as *perceptual compounds* rather than proper metaphors, and children may accept overextensions (e.g., *bat*) for labeling unfamiliar objects (e.g., *pointy roof*) capitalizing on simple analogical reasoning (Rubio-Fernández & Grassmann, 2016), often based on the *shape bias* (i.e., the idea that objects sharing the same shape also share the same name; Diesendruck & Bloom, 2003). A handful of studies adopted a viable middle ground assessment, namely multiple-choice tasks (Long et al., 2021; Martín-González et al., 2024; Mashal & Kasirer,

2011; Nippold et al., 1984) mitigating the issues of too complex and too easy tasks. While trying to go beyond the *early-late* debate to provide more fine-grained insights, these studies fell short in exploiting their potential. Specifically, some of these studies used a suboptimal number of alternatives, namely only two options (Nippold et al., 1984) resulting in a force-choice task, or too many options (Martín-González et al., 2024). A growing literature evaluating the validity of tasks suggests that three alternatives constitute the optimal paradigm (Rodríguez, 2005), and using too many distractors may support *strategic guessing* (Haladyna et al., 2019), thus increasing the possibility of overestimating children's ability. The way alternatives are presented is also problematic: some do not support alternatives with images (Nippold et al., 1984), with the risk of an extra loading on other cognitive skills (e.g., working memory, DeStefano & LeFevre, 2007; Lindner, Lüdtke, Grund, & Köller, 2017); others use images (Long et al., 2021) not controlling for visual parameters (e.g., visual complexity) that may interfere with children's selection (Madan et al., 2018; Sun & Firestone, 2021). An additional concern regards the use of pictures as answer options in absence of a verbalization of the linguistic content of the image (e.g., for the metaphor *That child is a grasshopper* showing the images of a running child, a jumping child, a bug, and a grasshopper, Martín-González et al., 2024). Showing graphical representations only does not consider the possible mediating role of mental imagery (for an overview see Carston, 2010b; Gibbs, 2006), namely, the possibility that children select the picture that resembles the mental image activated by the metaphorical prompt. Specifically, children might choose the grasshopper's image because the vehicle of the metaphor elicited that mental image (Paivio & Walsh, 2012), running the risk of intercepting a different phenomenon from metaphor comprehension.

Another issue concerns the non-modular nature of metaphorical competence and the role of other abilities in contributing to metaphor comprehension development. Most of the literature focused on structural linguistic abilities (i.e., vocabulary; Rundblad & Annaz, 2010), showing that these strongly affect metaphor skills. However, evidence from studies on adolescents, adults, and clinical populations suggests a broader involvement of cognitive and socio-cognitive skills in supporting

metaphor understanding development (Bambini et al., 2020, Bambini, Frau et al., 2025; Carriedo et al., 2016; Ceccato et al., 2025; Chiappe & Chiappe, 2007; Norbury, 2005). These abilities might play a key role, especially in early childhood, and this is not systematically accounted for to date. Moreover, these cognitive skills might differently affect different types of metaphors, as in the case of Theory of Mind reported to be involved to a greater extent in metaphors conveying psychological, compared to perceptual, content (e.g., *Daddy is a volcano*), requiring Theory of Mind (ToM) abilities to a greater extent (Canal et al., 2022; Tonini et al., 2023).

#### 1.2.4. The present study

The present study aims to capture the developmental trajectories of metaphor understanding in early childhood, taking into account (i) the distinction between physical and mental metaphors, (ii) different types of errors, including literal and unrelated answers, and (iii) role of cognitive and socio-cognitive skills. To disentangle these conceptual and methodological issues, we implemented a novel multiple-choice version of the Physical and Mental Metaphors task (Lecce et al., 2019), consisting of a sentence-picture association task assessing both physical and mental metaphors and including different response options, which was administered to children together with the Peabody Picture Vocabulary Test-Revised (PPVT-R; Stella et al., 2000), two First-order False Belief tasks (Sullivan et al., 1994; Wimmer & Perner, 1983) and the Backward Word Recall task (Lanfranchi et al., 2004). In addition, we adopted a comprehensive statistical approach able to detail eventual curvilinear trends of metaphor comprehension achievements (via Additive models) and to provide an ordinal classification (via ordinal logistic regressions) of error types during development.

Our hypotheses were that: a) physical and mental metaphors display different developmental trajectories; b) the types of errors (literal and unrelated) change throughout development, and c) metaphor understanding is related to a different extent to specific aspects of linguistic, cognitive, and socio-cognitive substrates, expecting in particular a link between mental metaphors and Theory of Mind.

## 1.2. Methods

### 1.2.1. Participants

A sample of 246 children ranging in age from four to six years was enrolled in the present study. Children were recruited from local schools and kindergartens located in Lombardy, Italy. Before running the data analysis, we excluded children who met one or more of the following criteria: being diagnosed with intellectual disabilities or neurodevelopmental disorders, hearing deficits, and not having acquired the Italian language before 3 years of age. The final sample included 175 children aged 4 to 6 years (Age:  $M = 5;4$ ,  $SD = 0;8$ ). Both parents signed written informed consent, and children were provided with age-appropriate information about their participation in the study. The study was approved by the Local Ethical Committee of the Department of Brain and Behavioral Sciences of the University of Pavia and followed the principles of the Declaration of Helsinki.

### 1.2.2. Procedure

Each child took part in three individual sessions administered in a silent room during school time. Each session lasted approximately 20 minutes. During the first session, children were assessed for their vocabulary and working memory skills, the second session was dedicated to the assessment of Theory of Mind, while the third session investigated receptive and productive metaphor skills.

### 1.2.3. Measures

#### 1.2.3.1 *The Physical and Mental Metaphor task - multiple-choice*


Metaphor comprehension skills were assessed using a novel multiple-choice task, modeled after the PMM task created for middle childhood (Lecce et al., 2019). In order to make the task suitable to use in early childhood, the open task format (verbal explanation in the original version) was changed to a closed format (sentence-picture matching task) by creating three multiple choice options. Six items were extracted from the original PMM task, while four were created de novo, for a total of ten metaphors (five physical and five mental). The six pre-existing items were selected among those metaphors for which, in the study of Del Sette and colleagues (2021), children scored

more than 0 in at least 80% of the cases and obtained the maximum score in at least 60% of the cases. One physical and five mental metaphors met the selection criteria. The four novel metaphors were built to balance the physical-mental set. The quality of these four novel items was evaluated against the original set of metaphors, assessed in the multiple-choice fashion: by correlating single novel items' comprehension scores with the six-item score, we found a good item-total correlation ( $.37 < r_s < .57$ ) indicating a similar structure of the addendum against the original set. Items were orally presented to the participants who were asked to choose the best and most apt (visually supported) explanation for the meaning of each sentence, among a set of alternatives. Following the indications on multiple-choice formats (e.g., Rodriguez, 2005), three alternatives were included and these included both literal and unrelated options, in addition to the correct metaphorical one. Importantly, in creating the alternatives, we controlled for the physical-mental dimension, focusing on the literal/nonliteral distinction. Specifically, for physical metaphors, all options capitalize on physical properties (i.e., "Dancers are feathers" means that: a) "They are dressed in white"; b) "They are light"; c) "They are short"), while for mental metaphors all answers capitalize on mental properties (e.g., "The teacher is an icicle" means that: a) "She likes cold things"; b) "She is strict," or c) "She likes singing"), keeping the physical or mental trait constant within options set. Each option was a) verbally provided, and b) associated with an image (See Fig.1.1). Specific criteria were set for both literal and non-literal incorrect alternatives: literal alternatives are based on salient encyclopedic features of the literal concept, adjusted to be plausible with respect to the topic. Unrelated alternatives are based on features that can be sensibly associated with the topic but do not in any way integrate vehicle features. Moreover, they are never antonyms (i.e., they never carry the opposite) of the correct meaning of the metaphor. Images accompanying the three alternatives were controlled for photometric dimensions, visual complexity, self-similarity, and symmetry. Answers were coded as correct (score: 1) and incorrect literal and incorrect unrelated (score: 0). Furthermore, in addition to the question on Accuracy, Mental metaphors only were followed by an additional question on Interpretation. Mental metaphors can have both a mental (appropriate,


*The teacher is mean*) and a physical (under-informative but plausible, *The teacher is freezing*) interpretation. While in the Accuracy question we controlled for possible literal/unrelated preference, in the Interpretation questions we controlled for the mental vs. physical preference, independently of the option selected during Accuracy evaluation. Thus, mental items only were followed by a forced-choice question about the interpretation, among a physical and a mental one (e.g., "She is freezing", coded as 0 points, vs "She is mean", coded as 1 point). The total score range for Accuracy was 0-10. The total score range for interpretation was 0-5.

**Example of Physical and Mental metaphors items**


a) **Physical Metaphor**



*Literal*




*Correct*




*Unrelated*


b) **Mental Metaphor**



*Literal*



*Correct*



*Unrelated*

**Interpretation question:**  
The teacher is an icicle means more that she is mean or that she is freezing?

*Figure 1.1. Examples of Physical and Mental Metaphors items. Panel A illustrates an example of physical metaphor (Dancers are feathers) with three possible alternatives, each accompanied by an image. Panel B: Mental metaphor (The teacher is an icicle) with three possible alternatives each accompanied by an image followed by the interpretation question with the mental (i.e., "She is mean") and physical (i.e., "She is freezing") options.*

#### 1.2.3.2. *Vocabulary*

The Italian version of the Peabody Picture Vocabulary Test-Revised (PPVT-R; Stella, Pizzoli, & Tressoldi, 2000) was used to assess children's vocabulary skills. The PPVT-R is a robust tool, with excellent split-half reliability ( $r = .88$ , Dunn & Dunn, 1981) and strong validity compared to the McCarthy Scales of Children's Abilities ( $r = .79$ , Naglieri & Maxwell, 1981). The PPVT-R consists of 175 verbal stimuli and assesses receptive vocabulary through a picture selection task, where children are asked to choose the image that best represents the meaning of a word spoken by the experimenter. Following standard procedures, each child's test starting point is determined based on their chronological age. If a child gives an incorrect answer to any of the first 8 benchmark items, the starting point is adjusted accordingly. Additionally, six consecutive errors in an 8-item block result in the test being discontinued. The total score is based on the number of correct responses, with a score range of 0-175.

#### 1.2.3.3. *Working Memory*

Working memory was assessed using the backward word recall task (Lanfranchi et al., 2004), a widely used measure that is part of a reliable verbal working memory battery (overall Cronbach's alpha = .56). The task involves presenting children with lists of two to five words, which they are asked to recall immediately in reverse order. For each of the four levels of difficulty (i.e., 2-, 3-, 4-, and 5-word trials), two items are administered. If a child provides incorrect responses to both items at a given level, the test is discontinued. The total score is based on the number of levels for which the child correctly recalls the words in reverse order, with a score range of 0-8.

#### 1.2.3.4. *Theory of Mind*

Theory of Mind abilities were assessed using four first-order false belief (FB) tasks. Two of these tasks, adapted from Sullivan and colleagues (Sullivan, Zaitchik, & Tager-Flusberg, 1994), were presented in Italian using a picture book. Each task included a first-order FB question and a control question. Children's responses were considered correct (1 point) only if they answered both the FB question and the control question correctly, with a score range of 0-2. Additionally, two first-

order FB tasks developed by Wimmer and Perner (1983) were adapted into Italian and administered using puppets. Each story included a first-order FB question and two control questions. Children's responses to the first-order FB question were considered correct only if they also answered the reality control question correctly, with a score range of 0-2. Given that these tasks are likely to measure the same underlying construct (see Beaudoin et al., 2020), raw scores (0/1) from both task types were summed, and composite scores with a score range of 0-4 were transformed into proportions, with a score range of 0-1.

#### 1.2.4 Statistical analysis

Statistical analyses aimed at evaluating the trajectories of acquisition of metaphor comprehension abilities first accounting for age effects and then gauging the role of other linguistic and cognitive skills, for the different metaphor types (Mental vs. Physical). As a preliminary step Pearson's correlations among all variables were performed. Descriptive summary data for each linguistic, cognitive, and socio-cognitive variable are reported for sequential age groups. In light of the well-known relationship between chronological age and vocabulary skills (e.g., Keuleers et al., 2015), we privileged an analytical approach where we assessed age-related and linguistic/cognitive-related effects separately. All statistical analyses were performed in the R environment (v. 4.2.3; R Core Team, 2023), with the R Studio editor (v. 2023.09.1+494).

##### 1.2.4.1 *Accuracy analysis*

Moving to the main analysis, accuracy and interpretation scores were analyzed with mixed-effects regressions using Generalized Additive Models (GAMs; Hastie & Tibshirani, 1986), a family of models suitable for studying complex non-linear relationships between outcome and predictive variables. GAMs were preferred over linear regressions for their ability to describe developmental effects through smooth functions, relaxing the assumption of either linear or non-linear effects in the regression framework. In GAMs, the effect of a predictor refers to a smooth function of the expected value of the response, which in turn may follow any exponential family distribution, allowing the use of a quasi-likelihood approach (Wood, 2017). Considering the nature of our

dependent variables (correct vs. incorrect, score 1 vs. score 0), the binomial family of GAMs was used, and random intercepts for individual-level (Subjects) and material-level (Items) variability were estimated. Accuracy scores were modeled as a function of multivariate smooth functions across each level of the Metaphor Type categorical factor and estimated areas of significant difference between smooths using *itsadug* package version 2.4.1 (van Rij et al., 2022). For recent guidelines in using GAMs in language acquisition research see Wieling, 2018. The package *mgcv* version 1.9-1 (Wood, 2023) was used to implement GAM models. We used the package *mgcViz* version 0.1.11 (Fasiolo et al., 2023) for data visualization.

To meet our aims, we ran three different models. The first GAM examined metaphor comprehension abilities (Accuracy) as a (smoothed) function of age and included the Metaphor Type (Mental vs Physical; sum-coded) factor in interaction with the smooth function. The second GAM tested metaphor comprehension as a (smoothed) function of Vocabulary, ToM, and WM, including the Metaphor Type interaction. The third GAM examined the physical or mental preference (Interpretation) for mental metaphors only as a smoothed function of age and included the measure of correct understanding (Accuracy) as an item-level factor (Correct, 1 vs. Literal/Unrelated, 0; sum-coded). The final random structure for each model included by-subjects and by-items random intercepts only, as more complex structures did not lead to model convergence.

#### 1.2.4.2. Response Types Analysis

To examine response types, we considered the three responses on an ordinal scale (correct > literal > unrelated). We thus analyzed the ordered accuracy scores as a function of Age and Metaphor Type in interaction, with Cumulative Link Mixed Models (CLMMs) with the *clmm()* function from the ordinal package, v. 2019.12-10 (Christensen, 2019). To account for non-linear relationships, we introduced both the linear and quadratic terms of Age, and we included random intercepts for subjects and items. To assess the contribution of fixed effects we compared models of increasing complexity via likelihood ratio test: the hierarchical approach revealed that adding the quadratic

term to the model did not improve the model's fit. The final model formula was: Accuracy ~ Age \* Metaphor\_type + (1 | Participants) + (1 | Items). Post-hoc comparisons were computed with the emmeans package, v. 1.10.6-090001 (Lenth, 2024).

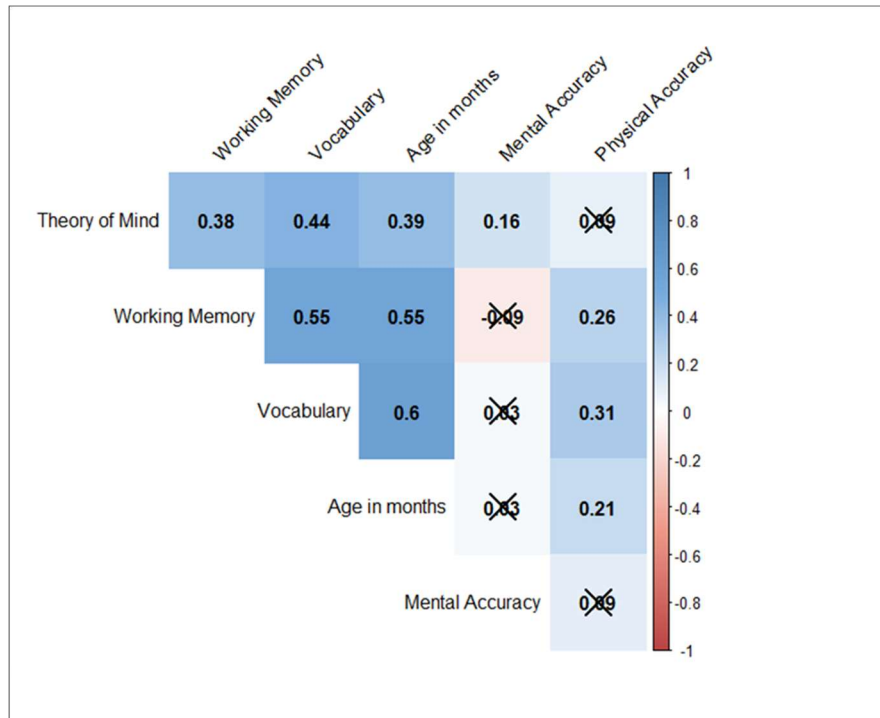
### 1.3. Results

Raw scores for each variable are reported across age groups in Tab. 1.1.

Table 1.1: Descriptive Statistics of children's performance in each age group.

<b>Measures</b>	<b>Range of possible scores</b>	<b>Whole sample N = 175</b>		<b>4 years N = 58</b>		<b>5 years N = 78</b>		<b>6 years N = 39</b>	
<b>Vocabulary</b>	0-175	68.4	(25.2)	52. 2	(18.6)	68. 9	(22.5)	91.7	(19.9)
<b>Working Memory</b>	0-8	2.01	(1.76)	1.0 9	(1.53)	1.9 9	(1.70)	3.44	(1.19)
<b>Theory of Mind</b>	0-1	0.54	(0.29)	0.4 2	(0.26)	0.5 6	(0.28)	0.66	(0.30)
<b>Mental Metaphors</b>	0-5	1.74	(0.99)	1.8 3	(0.88)	1.5 6	(1.03)	1.95	(1.05)
<b>Physical Metaphors</b>	0-5	2.65	(1.12)	2.5 0	(1.08)	2.4 7	(1.12)	3.21	(1.00)

Correlations among all variables of interest are reported in Fig. 1.2.



**Figure 1.2. Correlogram between scores in the physical and mental metaphors accuracy and linguistic, and pragmatic variables.** The plot shows correlations between physical and mental metaphor accuracy scores and age, vocabulary, working memory and theory of mind. Positive correlations are displayed in blue and negative correlations in red. The color intensity is proportional to the magnitude of correlation coefficients. Crossed cells indicate associations at  $p$ -value  $> .05$ .

### 1.3.1. Developmental trajectories of metaphor comprehension: analysis of Accuracy and Error Types

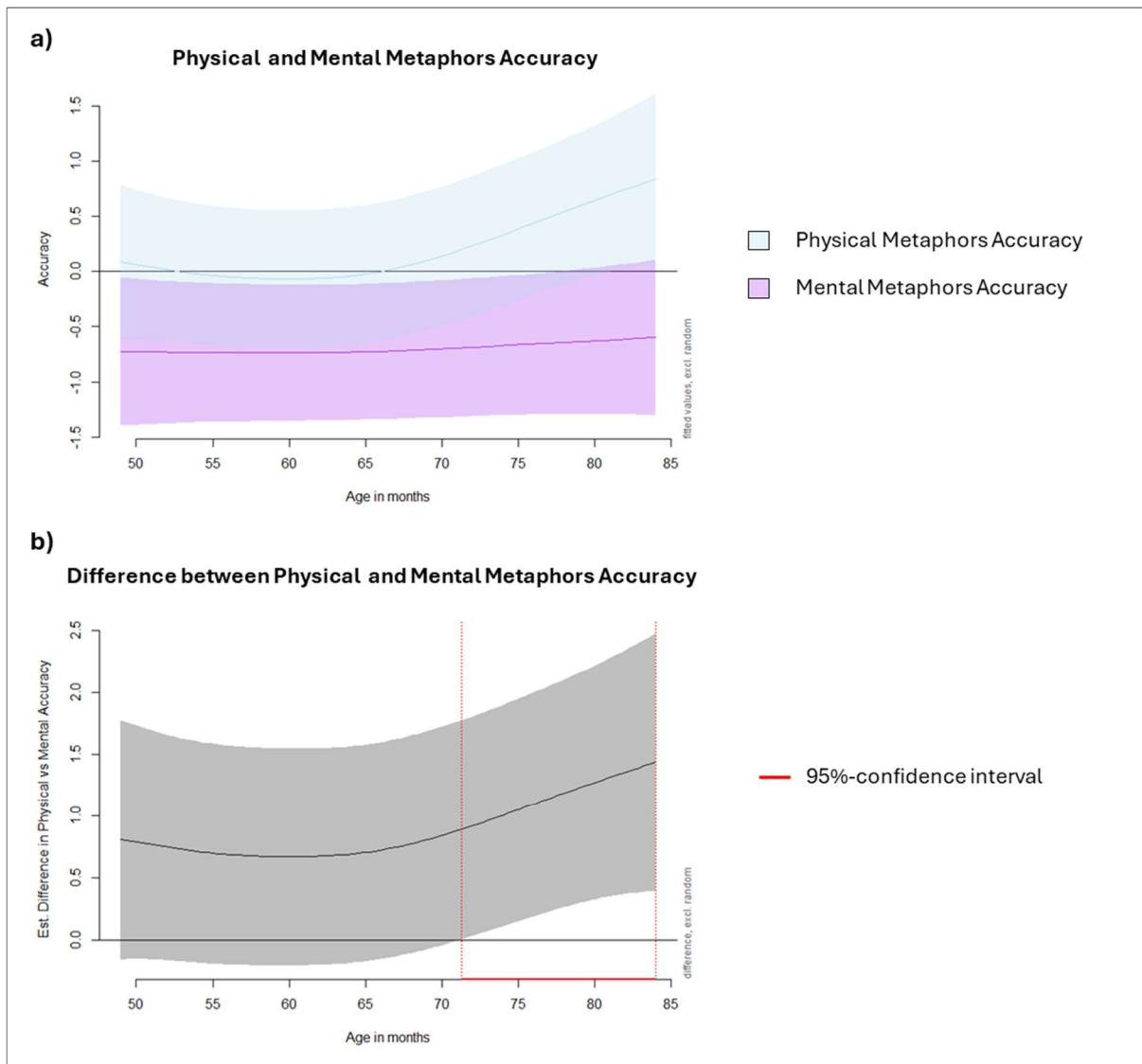
#### 1.3.1.1. Accuracy analysis

The first GAM showed an effect of Metaphor Type approaching significance ( $\beta = -0.83$ ,  $t = -1.89$ ,  $p = .058$ ), which revealed that Mental Metaphors were more difficult to understand than physical ones (across the age range). Moreover, the model showed specific trends for Metaphor Type. In Physical Metaphors, a significant growth of Accuracy as a function of Age emerged ( $s(\text{Age})$ :Physical metaphors,  $edf = 2.328$ ,  $F = 12.93$ ,  $p = .005$ ), showing that Age modulates a non-linear change in Physical Metaphors Accuracy. Visual inspection showed that the non-linear change described a rapid growth in Physical metaphor Accuracy with age (See Fig. 1a). In Mental Metaphors, a more flat, non-significant effect of Age was found ( $s(\text{Age})$ :Mental metaphors,  $edf = 1.274$ ,  $F = 0.52$ ,  $p = .767$ ; See Fig. 1.1a). Estimating differences between the two smooth effects, we found a significant

difference between Physical and Mental Metaphors smooths starting from 71 months of Age, signalling that from six years of age the performance in Physical metaphors departs from the one in Mental ones (See Fig. 1.1b). Model's formula and output are reported in Table 1.2.

Table 1.2. Output of the Generalized Additive Model with Accuracy as dependent variable and age as predictor

<i>Model's Formula:</i>				
Accuracy ~ Metaphor Type + s(Age, by= Metaphor Type) + s(Subjects, bs= "re") + s(Items, bs= "re")				
<i>Fixed Effects</i>				
Predictors	Odds Ratios	CI	edf	p
(Intercept)	0.75	0.49 – 1.15	-1.33	0.185
Metaphor Type[MentalvsPhysical]	0.44	0.18 – 1.03	-1.90	0.058
Smooth term (Age) × Metaphor Type[Mental]			0.52	0.767
Smooth term (Age) × Metaphor Type[Physical]			12.91	<b>0.005</b>
<i>Random Effects</i>				
s(Subjects)			8.60	0.310
s(Items)			119.64	<b>&lt;0.001</b>
Observations	1750			
R <sup>2</sup>	0.113			



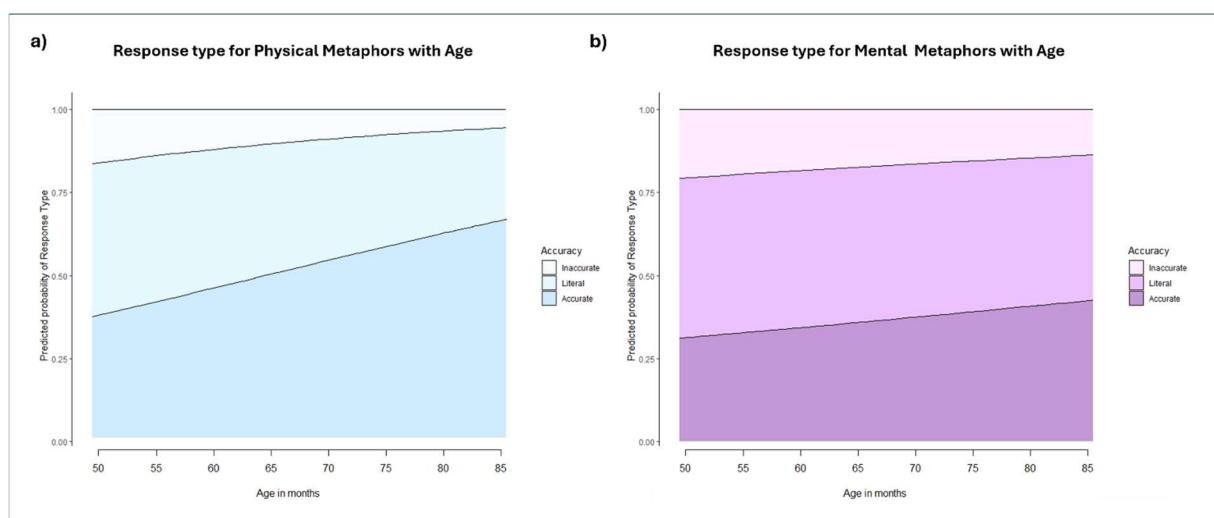
**Figure 1.3. Physical and Mental metaphors comprehension.** a) Smooths of Physical and Mental metaphors accuracy. Shaded bands represent the 95%-confidence interval. b) Smooth difference between the two (non-linear) smooths comparing the performance in Physical and Mental metaphors. The red line on the x-axis (and vertical dotted lines) indicates that the difference is significantly different from zero.

### 1.3.1.2. Response Types analysis

The CLMM showed a main effect of Age (OR = 1.25, 95% CI [1.13, 1.39],  $z = 4.27$ ;  $p < .001$ ), a main effect of Metaphor Type (OR = 0.54, 95% CI [0.30, 0.97],  $z = -2.06$ ;  $p = .040$ ), and a significant Age by Metaphor type interaction (OR = 0.83, 95% CI [0.69, 0.99],  $z = -2.02$ ;  $p = .044$ ), to probe this interaction we performed post-hoc analyses.

The analysis of simple trends of cumulative probabilities confirmed a different effect of Age on Metaphor comprehension. For Physical metaphors, Age increase was associated with less

Unrelated answers ( $\beta_{\text{prob}} = -0.03$ , CI [-0.05, -0.01],  $\chi = -3.51$ ,  $p < .001$ ), less Literal answers ( $\beta_{\text{prob}} = -0.05$ , CI [-0.08, -0.03],  $\chi = -4.00$ ,  $p < .001$ ) and more Accurate answers ( $\beta_{\text{prob}} = 0.08$ , CI [0.04, 0.12],  $\chi = 4.37$ ,  $p < .001$ ). Differently, for Mental metaphors, no changes in the probability of observing Accurate and Literal answers were observed. The result on Unrelated answers only was numerically in line with the ones on Physical metaphors but failed to reach significance ( $\beta_{\text{prob}} = -0.02$ , CI [-0.04, -0.00],  $\chi = -1.87$ ,  $p = .061$ ). To better examine this latter trend, we performed the analysis of the probability distribution of each Accuracy level between Metaphor Types as a function of Age. With this approach, we confirmed that, with Age, Unrelated answers between Physical and Mental decreased at a similar rate ( $\Delta_{\text{prob}} = -0.01$ , CI [-0.03, 0.01],  $\chi = -0.84$ ,  $p = .399$ ). Conversely, the analysis revealed a significant difference in the Age trend for Literal answers between Physical and Mental metaphors ( $\Delta_{\text{prob}} = -0.04$ , CI [-0.07, -0.01],  $\chi = -2.80$ ,  $p = .005$ ): while Mental metaphors have the same probability to obtaining literal answers with Age, the probability to observe literal answers for Physical metaphors decreases with Age. Moreover, the analysis indicated a significant difference in the effect of Age between Physical and Mental metaphors for Accurate answers ( $\Delta_{\text{prob}} = 0.05$ , CI [0.00, 0.09],  $\chi = 2.19$ ,  $p = .028$ ), with Physical metaphors showing more Accurate answers in comparison to Mental ones as Age increases (See Fig.1.4).



**Figure 1.4. Effect plots of changes in type of answers with age.** a) Effect plot of changes in type of answers for physical metaphors. Each band represents the proportion of Inaccurate (upper band), Literal (medium band) and Accurate (lower band) responses

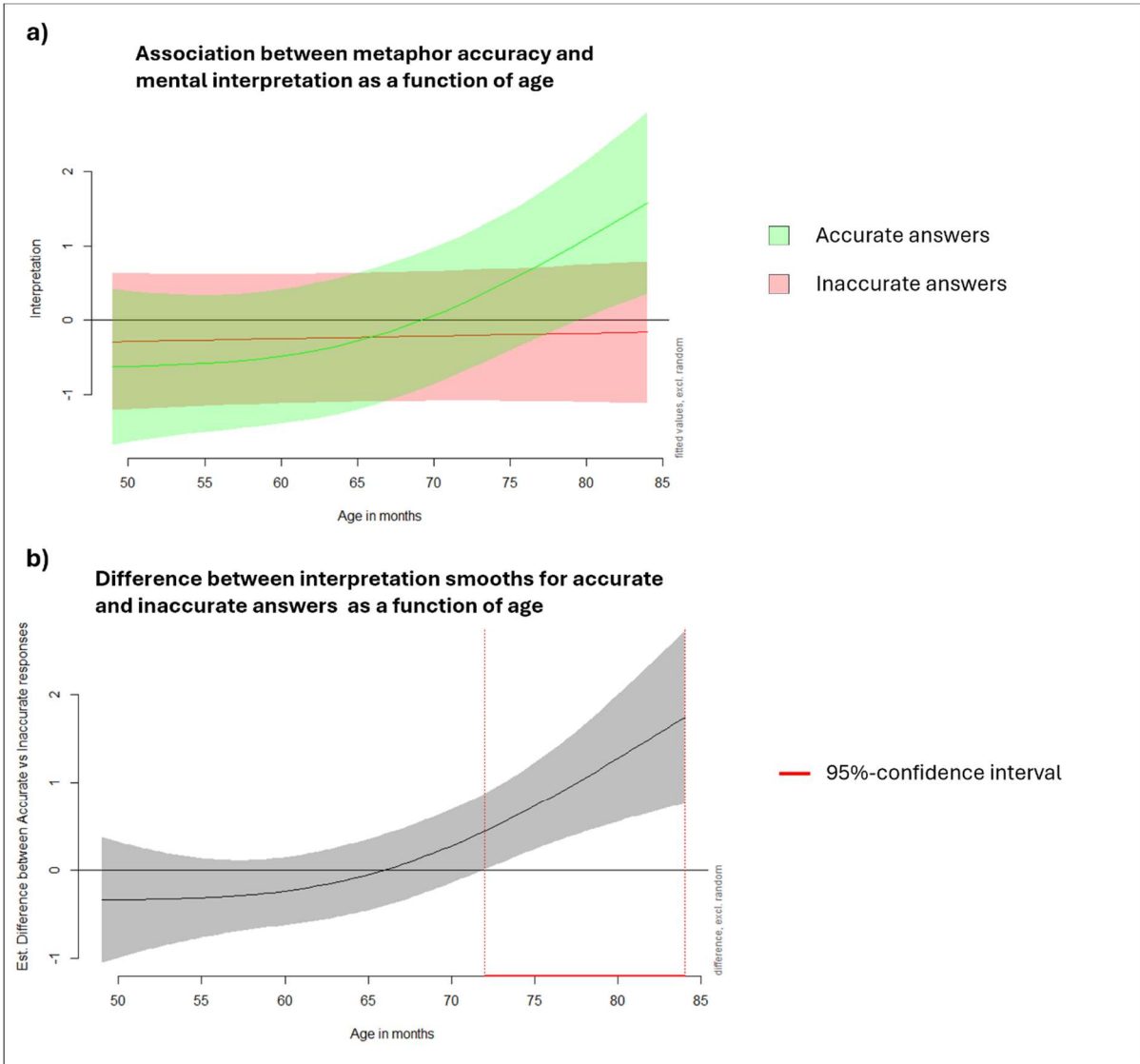
through age. B) Effect plot of changes in type of answers for mental metaphors. Each band represents the proportion of Inaccurate (upper band), Literal (medium band) and Accurate (lower band) responses through age.

### 1.3.2. Interpretation Analysis

Finally, we used a GAM to evaluate the preference for mental over physical interpretation in mental metaphors. This GAM showed a significant interaction between Accuracy (of Mental metaphors) and Age ( $s(\text{age}): \text{Accuracy}$ ,  $\text{edf} = 1.775$ ,  $F = 16.12$ ,  $p < .001$ ): only when children were accurate in understanding the meaning of the (Mental) metaphor, they preferred Mental over Physical interpretation (See Fig. 1.5a). By estimating differences between the two smooth effects, we found a significant difference between Accurate and Inaccurate answers starting from 72 months of Age, signaling that from about six years of age having understood mental metaphors directs only to mental interpretations (see Fig. 1.5b). The model's formula and output are reported in Table 1.3.

Table 1.3. Output of the Generalized Additive Model with Interpretation as dependent variable and Age and Accuracy as predictors

<i>Model's Formula:</i>				
Interpretation ~ Accuracy + s(Age, by= Accuracy) + s(Subjects, bs= "re") + s(Items, bs= "re")				
<i>Fixed effects</i>				
<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>edf</i>	<i>p</i>
(Intercept)	0.85	0.36 – 2.00	-0.38	0.706
Accuracy [AccuratevsInaccurate]	1.14	0.83 – 1.59	0.81	0.418
Smooth term (Age) × Accuracy[Inaccurate]			0.12	0.734
Smooth term (Age) × Accuracy[Accurate]			16.21	<b>&lt;0.001</b>
<i>Random Effects</i>				
s(Subjects)			39.95	<b>0.019</b>
s(Items)			105.51	<b>&lt;0.001</b>
Observations	868			
R <sup>2</sup>	0.197			



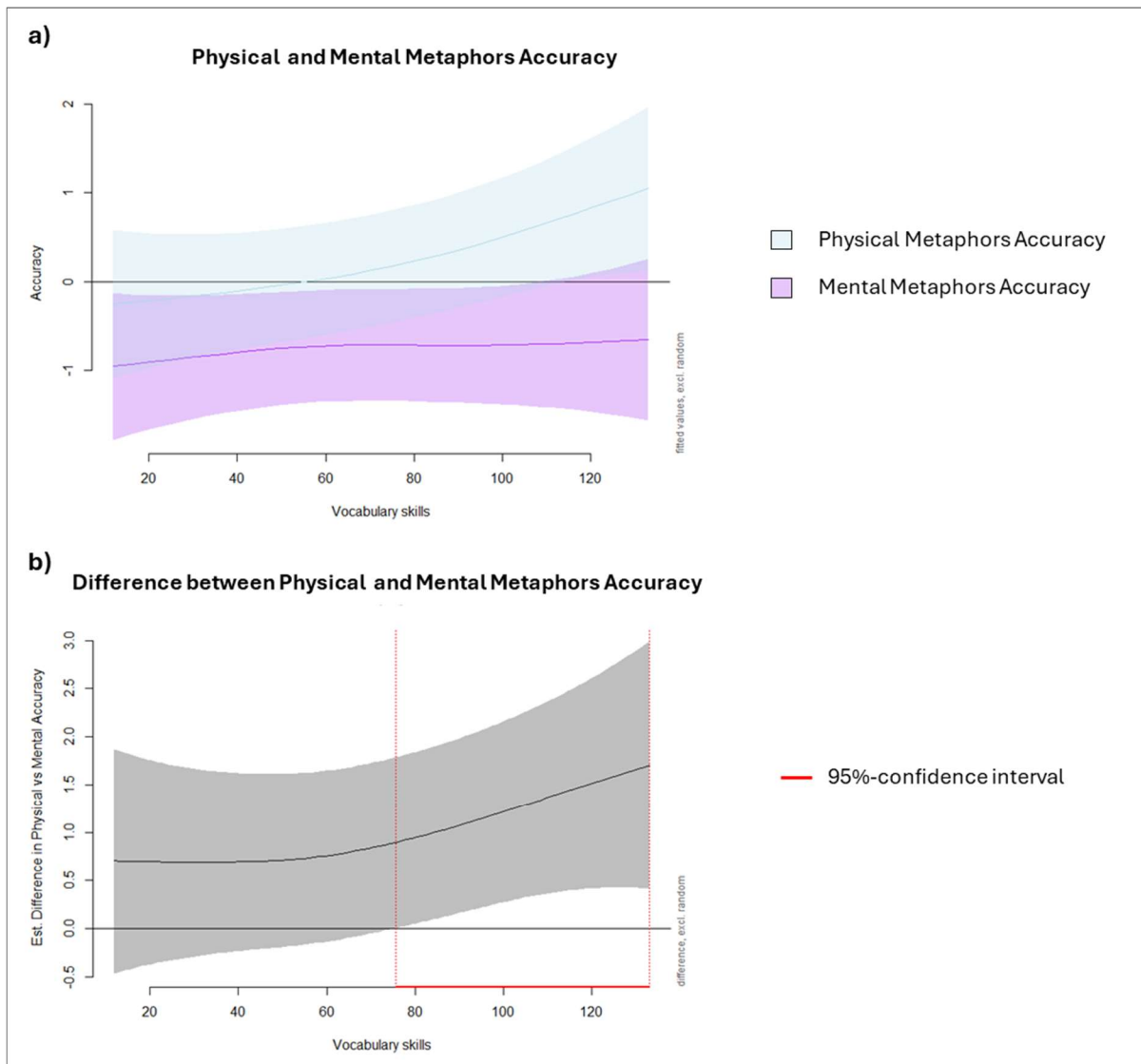
**Figure 1.5. Mental interpretations for Accurate and Inaccurate answers.** a) Smooths of Mental Interpretations for Accurate and Inaccurate answers for Mental metaphors. Shaded bands represent the 95%-confidence interval. b) Smooth difference between the two (non-linear) smooths comparing Accurate and Inaccurate answers for Mental metaphors. The red line on the x-axis (and vertical dotted lines) indicates that the difference is significantly different from zero.

1.3.3. Cognitive substrates of metaphor understanding across types

The role of cognitive variable was tested with a second GAM, which showed only a significant interaction between Physical Metaphor Type and the cognitive substrates, and namely with Vocabulary skills (s(Vocabulary):Physical metaphors,  $edf = 1.682$ ,  $F = 8.77$ ,  $p = .013$ ): as performance in Vocabulary increases, Physical Metaphors accuracy increases (See Fig.1.6). No other interactions with Physical metaphors emerged.

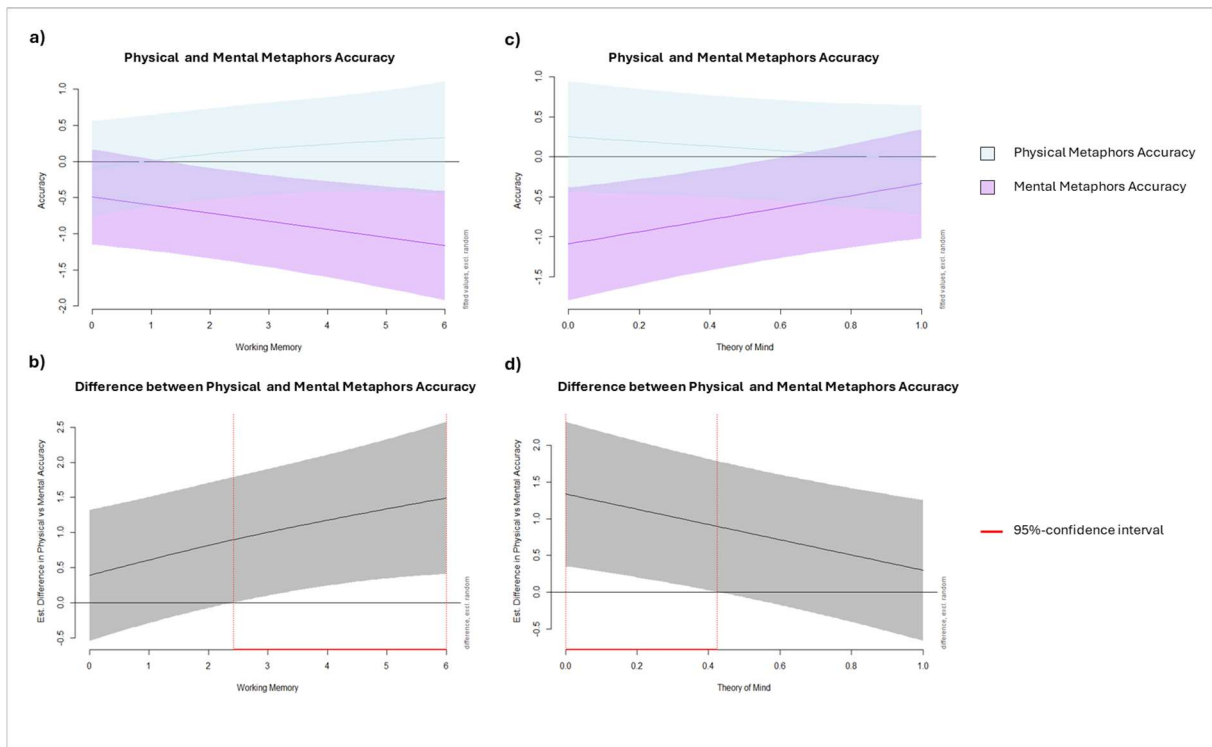
Table 1.4. Output of the Generalized Additive Model with Accuracy as dependent variable and cognitive predictors

<i>Model's Formula:</i>				
Accuracy ~ Metaphor Type + s(Vocabulary, by= Metaphor Type) + s(Theory of Mind, by= Metaphor Type) + s(Working Memory, by= Metaphor Type) + s(Subjects, bs= "re") + s(Items, bs= "re")				
<i>Fixed Effects</i>				
<i>Predictors</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>edf</i>	<i>p</i>
(Intercept)	0.74	0.48 – 1.15	-1.32	0.186
Metaphor Type[MentalvsPhysical]	0.43	0.18 – 1.03	-1.89	0.059
Smooth term (Vocabulary) × Metaphor Type[Mental]			0.35	0.787
Smooth term (Vocabulary) × Metaphor Type[Physical]			8.77	<b>0.013</b>
Smooth term (Theory of Mind) × Metaphor Type[Mental]			6.15	<b>0.013</b>
Smooth term (Theory of Mind) × Metaphor Type[Physical]			1.06	0.303
Smooth term (Working Memory) × Metaphor Type[Mental]			4.43	<b>0.035</b>
Smooth term (Working Memory) × Metaphor Type[Physical]			2.23	0.161
<i>Random Effects</i>				
s(Subjects)			11.66	0.254
s(Items)			120.78	<b>&lt;0.001</b>
Observations	1750			
R <sup>2</sup>	0.121			



**Figure 1.6. Association between Vocabulary and Physical and Mental metaphors accuracy.** a) Smooths of the relationship between Physical and Mental metaphors accuracy and Vocabulary scores. Shaded bands represent the 95%-confidence interval. b) Smooth difference between the two smooths comparing the effects of Vocabulary on Physical and Mental metaphors comprehension. The red line on the x-axis (and vertical dotted lines) indicates that the difference is significantly different from zero.

The same model highlighted two interactions between mental metaphors and the cognitive substrates: Linear effects of Theory of Mind and Working Memory were observed for Mental metaphors only ( $s(\text{ToM}):\text{Mental metaphors}$ ,  $df = 1.000$ ,  $F = 6.15$ ,  $p = .013$ ;  $s(\text{WM}):\text{Mental metaphors}$ ,  $df = 1.000$ ,  $F = 4.43$ ,  $p = .035$ ). As indicated in Fig. 1.7a, as Theory of Mind scores linearly increase, the performance in Mental Metaphors understanding increases and higher scores in Working Memory were associated with worse performance in Mental Metaphors.



**Figure 1.7. Association between cognitive and socio-cognitive variables and Physical and Mental metaphors accuracy.** a) Smooths of the relationship between Physical and Mental metaphors accuracy and Working Memory scores. Shaded bands represent the 95%-confidence interval. b) Smooth difference between the two smooths comparing the effects of Working Memory on Physical and Mental metaphors comprehension. The red line on the x-axis (and vertical dotted lines) indicates that the difference is significantly different from zero. c) Smooths of the relationship between Physical and Mental metaphors accuracy and Theory of Mind. Shaded bands represent the 95%-confidence interval. d) Smooth difference between the two smooths comparing the effects of Theory of Mind on Physical and Mental metaphors comprehension. The red line on the x-axis (and vertical dotted lines) indicates that the difference is significantly different from zero.

#### 1.4. Discussion

This study aimed to investigate the developmental trajectories of metaphor comprehension in early childhood using an age-appropriate tool taking into account different types of metaphor and response options and adopting a fine-grained statistical approach for metaphor development. We expected to observe a different pattern of acquisition for physical and mental metaphors, determined by modifications of the response types. We also assumed that individual differences in cognitive and socio-cognitive skills play a role in metaphor understanding at this age, with a specific contribution of ToM for mental metaphors. Our results showed that when applying a refined statistical approach and controlling for the psychometric properties of the task, namely the use of

metaphors rather than perceptual compounds, and the appropriate number and type of options, the study of metaphor development goes far beyond the *early-late* debate, showing different “challenges” or steps, that children have to climb in reaching full fledged metaphor skills.

Our analysis via additive modeling revealed that physical and mental metaphors were associated with different developmental processes within the same time window. Specifically, the developmental trajectory of physical metaphors departed from mental metaphors, drawing a curvilinear trend, with the age of six marking a significant improvement in physical metaphors only. This evidence must be considered in light of the pragmatic and cognitive achievements in this developmental period. During the whole preschool period, children collect an impressive amount of pragmatic tools and consolidate some pragmatic inferential abilities, strategically extracting information from the context (Bohn & Frank, 2019; Grigoroglou & Papafragou, 2019). At the end of this process, children can detect Gricean maxim violation at an adult-like level (Okanda et al., 2015) as well as derive implicatures through simple inferences (e.g., indirect answers; Bernicot et al., 2007), showing to have built the prerequisite pragmatic toolkit for metaphor comprehension. In this phase, children also learn to capitalize on perceptual experience to guide knowledge and language acquisition (Neill & Astington, 1992; Werker, 2018), thus exploiting this experience to enrich their conceptualization of objects (Juhasz et al., 2011; Ayhan et al., 2015). This is particularly relevant for analogical reasoning development, which might have a role in interpreting metaphor understanding (Holyoak & Stamenković, 2018). In this context, sensorial experience plays a pivotal role in metaphor acquisition: between five and seven years of age, children’s development of analogical perceptual reasoning is related to metaphor comprehension (Nippold & Sullivan, 1987). Crucially, in this phase, to determine similarities between objects children explore deeper properties that are less often available in the perceptual array such as labels, internal biological components, or motion capabilities (Rakison, 2005).

The first step to climbing seems to be represented by literal interpretations. Analysis of error types via multinomial ordinal regressions showed that the improvement in physical metaphors was driven

by a drop of literal interpretations, which conversely did not decrease for mental metaphors. Unrelated answers decreased at a similar rate in both Physical and Mental metaphors. This suggests that dropping literal interpretation is a necessary stage in developing metaphorical skills, while the reduction of unrelated answers does not determine an improvement. According to the *early* stance, literal answers reflect a general children's preference for literal interpretations, also shared with adults (Neff & Falkum, 2022). Conversely, the *late* stance considered literal answers as reflecting a pure deficit in metaphor comprehension (Winner, 1997). However, our results suggest that the tendency to prefer literal interpretation may play a more crucial role in metaphor development than previously described, reflecting an initial step toward metaphor understanding that needs to be overcome. Insights on the role of literal meaning in metaphor processing derive from the research involving the adult population (Carapezza & Garello, 2025; Carston, 2010b; Weiland et al., 2014), and may have ontogenetic implications. Several authors suggested that some activation of the literal meaning during metaphorical elaboration occurs (Al-Azary & Katz, 2021; Rubio Fernandez, 2007; Weiland et al., 2014) and that, intriguingly, prime words related to the literal concept denoted by the vehicle display a facilitation effect on lexical-semantic operations occurring during metaphor processing (Weiland et al., 2014). These insights suggest that literal meaning plays a role in the adult population, smoothing the process and possibly offering an alternative route to grasp the intended meaning. However, while this *literal path* allows reaching the correct interpretation for some metaphors, it may be unsuccessful when employed rigidly for all figurative phenomena. From a developmental perspective, literal answers observed in children reflect a stage in which children rigidly apply a possible – but not the best – strategy to interpret metaphorical meanings, as it happens for other forms of overregularization (Figuerola & Gerken, 2019; Marcus et al., 1992). The literal one is a necessary step during which children employ rudimental strategies to forge their pragmatic repertoire and exploit it flexibly.

Another step that children must climb seems to be the *physical* one. The additional analysis on mental metaphor interpretation (e.g., *The teacher is an icicle*) revealed that until six years of age, even

when the literal alternative is removed, children prefer physical (e.g., *She is freezing*) over mental interpretation (e.g., *She is mean*). This result suggests that in contrast with the predictions of the early stance (Neff & Falkum, 2022), removing the literal option may not be sufficient for specific metaphor types. Moreover, even at six years of age, only a few children who selected the correct option in the accuracy task did not fall for physical interpretation. While it might be the case for physical metaphors, which require children to overcome literal answers, mental metaphors seem to involve a further step. As suggested by electrophysiological evidence on the elaboration of physical metaphors via dual coding of meaning representation (Canal et al., 2022), children may benefit from activating perceptual knowledge of the vehicle during physical metaphor comprehension. However, trying to use this toolkit to process mental metaphors may lead to a *physical bias*. Specifically, activating the perceptual information of the vehicle for a physical metaphor (e.g., *Dancers are feathers*), may help in deriving the intended meaning (e.g., *being light*) (Al-Azary & Katz, 2021; Carston, 2010b; Paivio & Walsh, 2012). Conversely, we argue that the activation of the perceptual information of the vehicle during mental metaphor interpretation (e.g., *The teacher is an icicle*), may lead to an unappropriated physical interpretation (e.g., *She is freezing*) in early developmental phases. Literature on the development of mental processes highlights that mental content is physically grounded, and that sensory experience during the development of high-order knowledge plays a scaffolding role (L. E. Williams et al., 2009). As for the literal step, we suggest that the physical step may constitute an essential stage toward full-fledged metaphorical competence, during which children settle the pragmatic and cognitive acquisition for physical metaphor understanding.

Individual differences in linguistic, cognitive, and socio-cognitive variables further provide insights into the differences between physical and mental metaphors. The Additive model on individual differences showed that vocabulary abilities supported the development of physical metaphors, but not of mental ones. The role of vocabulary in metaphor development is in line with previous literature pointing out the involvement of structural linguistic abilities for metaphorical competence

(Lampri et al., 2024; Rundblad & Annaz, 2010; Stamenković et al., 2019). The specific contribution of receptive vocabulary for physical metaphors only hints at the assumption that metaphors may benefit from exploring the semantic features of the topic and the vehicle (Kintsch, 2000). Thus, a deeper vocabulary knowledge may foster feature exploration during metaphor comprehension. However, this may not be sufficient for mental metaphors, in which the metaphorical meaning is rarely inscribed in the palette of vehicle features (Tourangeau & Rips, 1991). Consistently with this hypothesis, our data reported a specific contribution of Theory of Mind in supporting mental metaphor understanding, in line with previous studies (Lecce et al., 2019; Tonini et al., 2023). As suggested by Lecce and colleagues (2019), in the preschool period the link between ToM and mental metaphor may be stronger than at later stages, as is true for language in general (Milligan et al., 2007). To begin with, to understand mental metaphors children have to master psychological lexicon (i.e., lexicon referred to inner mental states), which is linked to ToM rather than vocabulary (Grazzani & Ornaghi, 2012). Even if the difficulty in mental metaphors cannot be linked to psychological content only (Cicone et al., 1981), it may be a source of difficulty in early childhood when children are fine-tuning their mental lexicon (Moore & Davidge, 1989; Pascual et al., 2008). Furthermore, in these early phases, the ability to master false belief is linked to the awareness of *appearance-reality* distinction (i.e., the possibility to apply different representations of the same object; Gopnik & Astington, 1988), which may be necessary, even if not sufficient, for creating appropriate concrete-abstract links for grasping psychological content (Cicone et al., 1981). Along these lines, to derive a mental metaphorical meaning (e.g., *The teacher is an icicle*), the child must accept that the representation of the vehicle (e.g., *icicle*) has a general perceptual representation (i.e., a cold object) but acquires a new possible abstract representation in the context of the topic (i.e., someone with no feelings). This may explain, at least partially, why ToM is not linked to physical metaphor in this phase. During physical metaphor understanding children do not need to suppress the physical (apparent) representation of the vehicle to reach the metaphorical meaning, which conversely seems to help physical metaphoric meaning construction (Canal et al., 2022; Rubio Fernandez,

2007). At later stages, when the appearance-reality distinction is well-developed, more sophisticated aspects of ToM may come into play as facilitators for mental metaphor processing, such as emotion recognition (Canal et al., 2022).

A more puzzling finding is offered by the correlation/effects of working memory. While there was no significant working memory effect in predicting physical metaphors, higher levels of WM corresponded to a decrease in mental metaphor performance. Despite previous literature reporting a crucial role of executive functions during metaphor comprehension, these measures do not seem to support all metaphor types (Chiappe & Chiappe, 2007; Mashal, 2013). Recent evidence showed that fluid intelligence may not be a determinant in metaphor processing, especially for simpler metaphors (Stamenković et al., 2019). Also, another strand of literature reported that individuals with high working memory tend to capitalize on cognitively demanding skills even when those are suboptimal for task resolution (DeCaro et al., 2016; DeCaro & Van Stockum, 2018). Thus, it is possible that capitalizing on WM in a task that requires a combination of pragmatics and socio-cognitive skills may result in a drop in performance (See also Pompei et al., 2025).

Overall, our results offer some insights into the development of metaphor comprehension skills in early childhood, pointing to multiple steps that children need to overcome toward full-fledged metaphorical competence. By using a novel task that took into account literal and physical preferences, we were able to disentangle the multifaceted nature of metaphor in the acquisition. The age of six, when children transition to school, represents a key moment to develop from more literal to figurative interpretation, although sensory information still represents a source of processes, not yet abandoned for more abstract representations. At this stage, physical metaphors are fully grasped, but mental metaphors are still captured in a literal and physical phase. Yet these processes should not be considered in isolation, but as they develop in parallel with vocabulary and social cognition. Going beyond the *early-late* debate, we argue that metaphor comprehension abilities involve different aspects, from reasoning along the sensory-abstract dimension to

overcoming literal aspects of meaning, in turn interacting with structural linguistic knowledge and socio-cognitive skills.

## STUDY TWO

### HOW A WHITE DOG BECOMES A CLOUD:

#### A CROSS-SECTIONAL AND DISTRIBUTIONAL SEMANTICS INVESTIGATION OF METAPHOR

#### PRODUCTION SKILLS IN EARLY CHILDHOOD<sup>3</sup>

### 2.1. Introduction

According to some estimations, we produce one metaphor, broadly intended as a non-literal use of words, a minute (Pollio et al., 1997). Metaphors are produced to fulfill a broad range of functions, from shaping other thoughts, to persuading others or creating social connectedness (Bowes & Katz, 2015; Sopory & Dillard, 2002b; Thibodeau & Boroditsky, 2013). Metaphor production is attested also in children, that show a genuine use of metaphors especially to negotiate meaning during learning (Cameron, 2003; Deignan & Semino, 2022). However, despite a large body of literature on the acquisition of metaphor comprehension skills (Grigoroglou & Papafragou, 2019; Lecce et al., 2019; Pouscoulous & Tomasello, 2020; Winner et al., 1976), the study of metaphor production in development is very limited and discontinued, with very few and mostly early studies (Gottfried, 1997; Winner, 1979; Winner et al., 1980). These early studies either measured spontaneous (e.g., *red balloon* for an apple, Winner, 1979) or task-elicited metaphors (e.g., *zebra-shell* for the picture of a shell with white and black stripes, Gottfried, 1997) and supported the idea of an early development of metaphor production skills, already present in toddlers. Yet, the lack of a robust methodology for assessing and analyzing metaphor production data, alongside the scarce consideration of cognitive and socio-cognitive contribution, makes the level of knowledge about productive metaphor skills acquisition somewhat rudimental. This study aims to provide fresh evidence on metaphor production development in early childhood while proposing an

---

<sup>3</sup> This chapter is a manuscript in preparation for submission to a peer-review journal as “Pompei, C., Mangiaterra, V., Canal, P., Del Sette, P., Lecce, S., & Bambini, V., *How a white dog becomes a cloud: a cross-sectional and distributional semantics investigation of metaphor production skills in early childhood*”.

innovative methodological framework combining cross-sectional assessment, distributional semantics, and refined statistical approaches.

Early observational studies reported the use of metaphorical expressions in toddlers' communication, particularly during pretend play (Billow, 1981; Winner, 1979). In these studies, children's expressions - such as *snake* for a car toy twisted up - were categorized as metaphors when met one or both of the following criteria: the reference to an object using a different name from its conventional one, and the involvement of a comparison between two similar objects (Vosniadou, 1987). However, as suggested by Gottfried (1997), rather than simply establishing if a lexical innovation fall into a linguistic definition of metaphor, it is crucial to determine whether these expressions are *used metaphorically*, namely if children are engaged in a flexible language use apt to communicate a feature that is relevant to the context. To integrate findings on spontaneous metaphor production, Gottfried (1997) proposed a task in which children were trained to play a *look-like* game with a puppet that generates inappropriate compounds (e.g., *cherry-balloon* for a yellow football) and were asked to correct him with appropriate compounds (e.g., *lemon-balloon*). Preschool-aged children were able to produce appropriate compounds (e.g., *zebra-shell* for a shell with black and white stripes), and the performance increased until 5 years of age. In a sentence completion task, children were asked to complete simile sentences verbally (e.g., *Look at the boy standing over there. He looks as gigantic as...?*; Gardner et al., 1975). In this case, even if the number of appropriate metaphors was scarce until adolescence, preschoolers produced more appropriate metaphors than school-aged children and were outperformed only by adolescents (Gardner et al., 1975). Accordingly, metaphor production acquisition seems to delineate a U-shaped trajectory, with younger children being more able than school-aged children to generate metaphors, which were said to be stacked in a *literal stage* (Gardner & Winner, 1978).

Taking a deeper look at these studies might however cast doubts on the solidity of the idea of an early metaphor production competence. In metaphor compounds generation, children may rely on the communicative strategy of using familiar words to fill a gap, making an over-extension (i.e.,

*zebra*) based on perceived similarity that could vanish when children acquire the requisite words (i.e., *striped*), and in which the distinction between a literal and a figurative use is hard to trace (Clark, 2020). Additionally, in the sentence completion task by Gardner and colleagues (1975), while producing the highest (even if scarce) number of metaphors, preschoolers also produced the highest number of inappropriate responses (e.g. *gigantic as a clock*). Hence, preschoolers' performance might actually reflect the inability to reject those comparisons that would not make sense to other individuals, cutting across the boundaries of conventional language and experience (Gardner & Winner, 1978). More recently, some renewed interest in children's metaphor generation arose, with a special focus on atypical development (Kasirer & Mashal, 2016, 2018). Although involving older children, these studies pointed out that other skills, such as vocabulary and executive functions, may play a role in metaphor production development. This matches an extensive literature on metaphor comprehension, which highlighted the role of verbal and socio-cognitive skills development (Norbury, 2005b; Tonini et al., 2023). However, the paucity of studies and the methodological issues in investigating the productive dimension of metaphor competence make this domain an unsolved challenge for developmental pragmatics. In particular, to the best of our knowledge, there are no studies systematically investigating metaphor production during early childhood and the involvement of cognitive skills and socio-cognitive abilities.

Another consideration regards the conceptual processes and the lexical architecture involved in metaphor production, and how this might impact development. Generally speaking, producing a metaphor (e.g., *The white dog is a cloud*) requires articulating a vehicle (e.g., *cloud*) in association with a topic (e.g., *white dog*), based on operations that include selecting salient common features between concepts (Pérez-Hernández & Duvignau, 2016). The vehicle of a metaphor is expected to be *diagnostic*, that is, to make salient those semantic similarities and dissimilarities needed to clarify our understanding of the topic (Katz, 1989). Hence, vehicle selection constitutes a pivotal aspect of metaphor production and, as suggested by previous literature, how children move from one conceptual domain to another for metaphor generation purposes may provide crucial insights into

the development of this skills (Gottfried, 1997; Vosniadou, 1987; Winner et al., 1980). A promising strand of research to track down such conceptual *moving* from the topic to the vehicle and the selection of the latter is offered by distributional semantics (Kintsch & Bowles, 2002; Reid & Katz, 2018). This approach provides a quantitative measure of similarity between words: vector space models, exploiting the hypothesis that words occurring in the same context are similar in meaning (Harris, 1954), are trained on large corpora to represent words as vectors in multidimensional semantic spaces, based on their co-occurrences in texts (Lenci, 2008). Studies on metaphors in this field typically measured topic-vehicle semantic similarity, computing the cosine of the angle between the vectorial representations of the two terms (i.e., a measure of their proximity in the semantic space) and reported that individuals, given a topic and asked to choose between several vehicles, tend to select the semantically closest, or at least moderately close, to it (Clevenger & Edwards, 1988; Katz, 1992; Pierce & Chiappe, 2009). Conversely, greater semantic distance between the metaphor terms and the literal prompts is associated with higher creativity scores assigned to the metaphors by human raters (Beaty & Johnson, 2021; Skalicky & Crossley, 2018). The distributional semantics approach, thus, is useful to chart the breath of the conceptual and semantic space used when producing a metaphor. This approach, however, has never been applied to the study of metaphor production acquisition.

#### 2.1.1. The present study

This study aimed to investigate metaphor production and its cognitive and socio-cognitive correlates in early childhood, combining cross-sectional and distributional semantics methods. To address these aims, we adopted an elicitation task in which children were asked to produce metaphors given literal prompts (e.g., *a white dog*), administered to 171 children alongside tests of vocabulary, inhibitory control and Theory of Mind. In addition to accuracy, we also computed the Semantic Similarity between the produced metaphors (e.g., *the dog is a cloud*) and the provided prompt. Fine-grained statistical approaches were used to measure metaphor production abilities and to determine the role of other cognitive and socio-cognitive skills, combining more

conservative methods and Additive modeling. We hypothesize that: a) Metaphor production abilities increase during early childhood and that b) this improvement is supported by cognitive and socio-cognitive skills (i.e., vocabulary, inhibitory control, and Theory of Mind); c) Semantic Similarity between the literal prompt and the metaphors decrease with age thanks to greater vocabulary skills and semantic knowledge.

## **2.2. Method**

### **2.2.1. Participants**

The study involved 171 typically developing children aged 4 to 6 years ( $M_{age}=5.40$ ,  $SD=0.78$ ), divided by age into three groups: 4-year-old children (age range: 4;0, 4;11;  $M_{age} = 4;6$ ), 76 5-year-old children (age range: 5;0, 5;11;  $M_{age} = 5;5$ ), and 39 6-year-old children (age range: 6;0, 6;11;  $M_{age} = 6;6$ ), all Italian native speakers. Both parents signed written informed consent, and children were provided with age-appropriate information about their participation in the study. A cohort of 135 adults (Age range 18;0-59;0) was involved in the study and divided into two groups: Young Adults ( $n= 58$ , age range: 18;0-29;0) and Adults ( $n= 77$ , age range: 30;0-59;0) using the platform (<https://limesurvey.org>) configured not to collect any personal data. The study was approved by the Local Ethical Committee of the Department of Brain and Behavioral Sciences of the University of Pavia (n. protocol 029/2019) and followed the principles of the Declaration of Helsinki.

### **2.2.2. Procedure**

Each child took part in three individual sessions administered in a quiet room during school time. Each session lasted approximately 20 minutes. During the first session, children were assessed for their vocabulary and working memory skills, while the second session was dedicated to the assessment of Theory of Mind skills. The third session assessed metaphor production and metaphor comprehension. The adult group underwent a single online session assessing metaphor production skills only.

### 2.2.3. Measures

#### 2.2.3.1. *Vocabulary*

To assess vocabulary skills, we adopted the Italian version of the Peabody Picture Vocabulary Test - Revised (PPVT-R; Stella et al., 2000). This test has been shown to have excellent split-half reliability value ( $r = .88$ , Dunn & Dunn, 1981) and high validity against the McCarthy Scales of Children's Abilities ( $r = .79$ , Naglieri & Maxwell, 1981). The PPVT-R includes 175 verbal stimuli and measures receptive vocabulary with a picture selection task: children are asked to choose, among four images, the one that best describes the meaning of the word uttered by the experimenter. Following the standard procedure, for each child, the chronological age is used to set individual test starting points. Six consecutive errors in an 8-item block resulted in the interruption of the test. The total score consists of the number of correct answers (score range: 0-175).

#### 2.2.3.2. *Inhibitory Control*

To assess inhibitory control we used the Day-Night tasks (Gerstadt et al., 1994), which showed an excellent inter-rater reliability in previous studies (ICC= 99%, Cuevas et al., 2016). The task employs a set of laminated cards (10 x 15 cm). All children were presented with the control task version followed by the Stroop-like version. For the control task, children were instructed to say *night* when shown a moon card and to say *day* when shown a sun card. For the Stroop-like Day-Night task children were instructed to say *night* when shown a sun card and *day* when shown a moon card. The task consisted of 16 trials (half for each type). No feedback was given during test trials. The total administration time was approximately 3 minutes. The percentage of correct trials was calculated (total possible score range: 0-16).

#### 2.2.3.4. *Theory of Mind*

Theory of Mind was assessed using the Wellman & Liu scale (2004). The scale showed an average infit of -.1 (SD = 1.1) and average outfit of 0.0 (SD = .7), meeting the criteria for a Rasch model as well as a good convergence validity with the Saly and Ann task (Hiller et al., 2014). The scale

consist of five tasks: Diverse Desires (understanding another's likes may differ from your own), Diverse Beliefs (understanding another person may think differently about the same situation), Knowledge Access (ability to judge another's knowledge of a scenario), False Belief (judging another's false-belief about the content of a descriptive box), and Hidden Emotion (understanding a person's facial expression does not always match their emotion). For the first four tasks, children were able to respond verbally (requiring a single-word unambiguous response) or by pointing to a picture. Task 5 included an additional justification component, where the child was required to give a brief explanation for their response. Each task was administered using visual or material support (e.g., images depicting the main character) and presented both a control and a focal question, the latter coded as either correct (1 point) or incorrect (0 point). The raw score range was 0-5, then transformed into proportions, with a final total score range of 0-1.

#### 2.2.3.5. *Metaphor Comprehension*

To assess metaphor comprehension, we used a newly developed multiple-choice version of the Physical and Mental Metaphors task (Lecce, Pompei et al., in prep), originally employing the verbal explanation task format (Del Sette et al., 2020; Lecce et al., 2019). The task consisted of 10 metaphors: 5 physical (i.e., metaphors that capitalize on physical properties, such as *Dancers are feathers*) and 5 mental (i.e., metaphors that capitalize on mental or psychological aspects, such as *The teacher is an icicle*). Children are asked to select the best fitting explanation for each metaphor, choosing among a set of three options, presented both verbally and visually, corresponding to a figurative, a literal, and an unrelated interpretation (e.g., for the physical metaphor above, *They are light*, with an image representing dancers jumping almost weightlessly, *They are dressed in white*, with an image representing white dressed dancers; *They are short*, with an image representing short dancers; for the mental metaphor above, *She likes cold things*, with an image representing the teacher in the class with a frozen board; *She is strict*, with an image representing the teacher with a serious expression; *She likes singing*, with an image of the teacher singing). Answers are scored as correct (1) for figurative interpretations or incorrect (0) for literal and unrelated interpretations. For the

purposes of the study, we limited the analysis to the physical metaphor part, given that the metaphor production task includes only physical items (total possible score range: 0-5).

#### 2.2.3.6. *Metaphor Production*

To assess metaphor production abilities, we adapted the elicitation task by Cortés, Cobos, & Tarbox (2018). This task is composed of 4 items in which children are given literal prompts and asked to produce metaphors. Based on our sample, the internal consistency of the task was acceptable both in children (Cronbach's alpha = .78) and in adults (Cronbach's alpha = .8), the task also showed an excellent inter-rater reliability (ICC= 92%). For each item, children are provided with a brief description of an object (e.g., a very white dog) and then instructed to change the prompted literal word (i.e., white) with a metaphoric equivalent (e.g., *That dog is a cloud*). If the child does not provide any answer, it is possible to provide a prompt (e.g., think of something that is always white). Responses were scored 2 when expressing salient metaphors (e.g., *That dog is a cloud*), defined as those figurative associations where the produced *vehicle* (e.g., cloud) exhibit the prompt (e.g., white) as a salient feature, i.e., a highly typical, based on word knowledge; responses were scored 1 when using similes (e.g., *The dog is like a cloud*) and non-salient metaphors (e.g. *The dog is an ice-cream*), defined as metaphors built around a vehicle that highlight features only marginally salient with respect to the prompt; literal or no answers were scored 0 points (total score range: 0-8). The version for the adult sample was administered via the web-based platform LimeSurvey® (<https://limesurvey.org>). Participants were provided with a brief description of an object (e.g., a very white dog) and then instructed to change the prompted literal word (i.e., white) with a metaphoric equivalent (e.g., *That dog is a cloud*) in the written format. No prompt was provided during the task. The survey was built to show a single item per time. The scoring procedure followed the lines of the one used for children.

In addition to the above scoring procedure, we computed the semantic similarity between the literal prompts and the produced metaphors to explore the semantic relationships underlying vehicle selection. Specifically, following previous approaches, we calculated the cosine similarity between

the produced metaphors (computed as the mean vector of topic and vehicle, e.g., *dog* and *cloud*) and their relative literal prompts (computed as the mean vector of the two terms in the literal prompt, e.g., *dog* and *white*), using the pre-trained word embeddings made available by Gunther and colleagues (2015) for Italian, namely, 400-dimensional vectors trained on the itWac corpus (Baroni et al., 2009) using Continuous Bag of Words (CBOW) algorithm (Mikolov et al., 2013). To obtain the measures of semantic similarity, we used the *LSAfun* package (Günther et al., 2015) in the R programming language.

#### 2.2.4. Statistical analysis

As a preliminary step, we computed Pearson's correlations between all variables to determine the models' formulae. As for the main analysis, the purpose was to assess a) the development of metaphor production skills across ages; b) the role of linguistic, cognitive, and socio-cognitive skills in the development of metaphor production; c) the semantic similarity between the literal prompt and the produced metaphor varies across ages and depends on other cognitive skills and d) the relationship between Metaphor Production skills and Semantic Similarity across the lifespan, i.e., by comparing children and the adult cohort.

To meet aims a) and b), in line with the literature about pragmatic development (e.g., Köder & Falkum, 2020), developmental stages were studied by considering three age groups (i.e., 4-, 5-, and 6-year children): this strategy allows to straightforwardly capture non-monotonic effects across ages, without imposing any a priori assumptions about the functional form (e.g., linear, quadratic, cubic, etc.) of an age gradient (Stone et al., 2010). Specifically, we fitted a linear model using the mean Metaphor Production Accuracy of each subject as the dependent variable, testing for the difference between Age Groups (effect coded in a backward sequential way: 5 vs. 4; 6 vs. 5) in interaction with individual differences in linguistic cognitive and socio-cognitive variables (Vocabulary, Inhibitory Control, Theory of Mind, and Metaphor Comprehension), scaled for each level of Age groups. The model's formula was: Metaphor Production scores  $\sim$  Age group \* (Vocabulary + Inhibitory Control + Theory of Mind + Metaphor Comprehension).

As for aim c), we fitted a linear mixed-effect model using log-transformed Semantic Similarity (e.g., literal prompt-metaphor cosine) values as the dependent variable, testing for the difference between Age Groups (effect coded in a backward sequential way: 5 vs. 4; 6 vs. 5) in interaction with individual differences in linguistic cognitive and socio-cognitive variables (Theory of Mind and Metaphor Comprehension), scaled for each level of Age groups. For considering the variability at the individual (Subject) and Item level, we included the relevant random structures to account for participants' and materials' grouping factors. The model's formula was: Semantic Similarity values  $\sim$  Age group \* (Vocabulary + Theory of Mind + Metaphor Comprehension) + (1 | Subjects) + (1 | Items).

To account for the developmental endpoint of metaphor production and Semantic Similarity patterns we fitted two additional models, a linear model and a linear mixed-effect model using Age Groups (effect coded in a backward sequential way: 5 vs. 4; 6 vs. 5; 18-29 vs. 6; 30-59 vs. 18-29) as categorical predictor and Metaphor Production scores and Semantic Similarity values as dependent variables. The models' formulas were: Metaphor Production scores  $\sim$  Age group and, Semantic Similarity values  $\sim$  Age group + (1 | Subjects) + (1 | Items).

We acknowledge the risks of using cut-points to categorize continuous data (i.e., Age; Sauerbrei & Patrick, 2010), and we address this risk with the exploratory analysis described in the last paragraph of this section. An exploratory analysis was performed to follow up the results that emerged in the main analysis, in particular, to provide a more in-depth description of the contribution of cognitive predictors in Metaphor Production accuracy and Semantic Similarity values, during different developmental time windows. We used Generalized Additive Models (GAMs; Hastie & Tibshirani, 1986) to investigate the interaction between age, linguistic, cognitive, and socio-cognitive skills while respecting the continuous nature of the Age variable (for recent guidelines in using GAMs in language acquisition research see Wieling, 2018). GAMs are a family of models suitable for studying complex non-linear relationships between outcome and predictive variables, also allowing the investigation of interactions between continuous variables via tensor

products. A tensor product essentially models a non-linear interaction by allowing the coefficients underlying the smooth for one variable to vary non-linearly depending on the value of the other variable (Wood, 2017). This constitutes a great advantage in the developmental exploration of individual differences, which may have different effects depending on the developmental time window. We fitted two models, one using Metaphor Production Accuracy and one using Semantic Similarity values as dependent variables. The final model's formula for Metaphor Production Accuracy was:  $\text{Metaphor Production Accuracy} \sim s(\text{Age}) + s(\text{Vocabulary}) + s(\text{Inhibitory Control}) + s(\text{Theory of Mind}) + s(\text{Metaphor Comprehension}) + \text{ti}(\text{Age}, \text{Vocabulary}) + \text{ti}(\text{Age}, \text{Inhibitory Control}) + \text{ti}(\text{Age}, \text{Theory of Mind}) + \text{ti}(\text{Age}, \text{Metaphor Comprehension})$ . The final model's formula for Semantic Similarity was:  $\text{Semantic Similarity} \sim s(\text{Age}) + s(\text{Theory of Mind}) + s(\text{Metaphor Comprehension}) + \text{ti}(\text{Age}, \text{Theory of Mind}) + \text{ti}(\text{Age}, \text{Metaphor Comprehension}) + s(\text{Subjects}, \text{bs} = \text{"re"}) + s(\text{Items}, \text{bs} = \text{"re"})$ .

All statistical analyses were performed in R (v. 4.2.3; R Core Team, 2023), with the R Studio editor (v. 2023.09.1+494), using the *lme4* (v1.1-26; Bates et al., 2015), the *lmerTest* (v. 3.1-3; Kuznetsova, 2017), and the *emmeans* (v. 1.10.6-090001; Lenth, 2024) packages.

The package *mgcv* version 1.9-1 (Wood, 2023) was used to implement GAM models. We used the package *mgcViz* version 0.1.11 (Fasiolo et al., 2023) for data visualization. To inspect areas of significant interaction, we used *itsadug* package version 2.4.1 (van Rij et al., 2022). After fitting each mixed-effect model, we checked model assumptions with the diagnostic inspection tools included in the DHARMA package (v. 0.4.7; Hartig, 2024). In all models, the diagnostics were satisfactory.

### 2.3. Results

Average group performance for each task and Semantic Similarities values are reported for each age group in Table 2.1.

Table 2.1. Descriptive Statistics of children's performance in each age group.

	Min-Max score range	4 years	5 years	6 years	p
Vocabulary	0-175	52.2(18.6)	69.8(22.5)	91.7(19.9)	<0.001
Inhibition	0-16	11.2(4.89)	12.3(4.84)	13.8(3.83)	<0.001
Theory of Mind	0-1	0.66(0.19)	0.65(0.18)	0.72(0.18)	0.024
Metaphor Comprehension	0-5	2.50(1.08)	2.47(1.12)	3.21(1.00)	0.002
Metaphor Production	0-8	1.02(1.18)	2.07(1.59)	4.15(2.05)	<0.001

Correlations between the linguistic, cognitive, and distributional variables are reported in Fig. 2.1. Metaphor Production scores positively correlated with Age, Vocabulary, Theory of Mind, Inhibitory control, and Metaphor Comprehension.

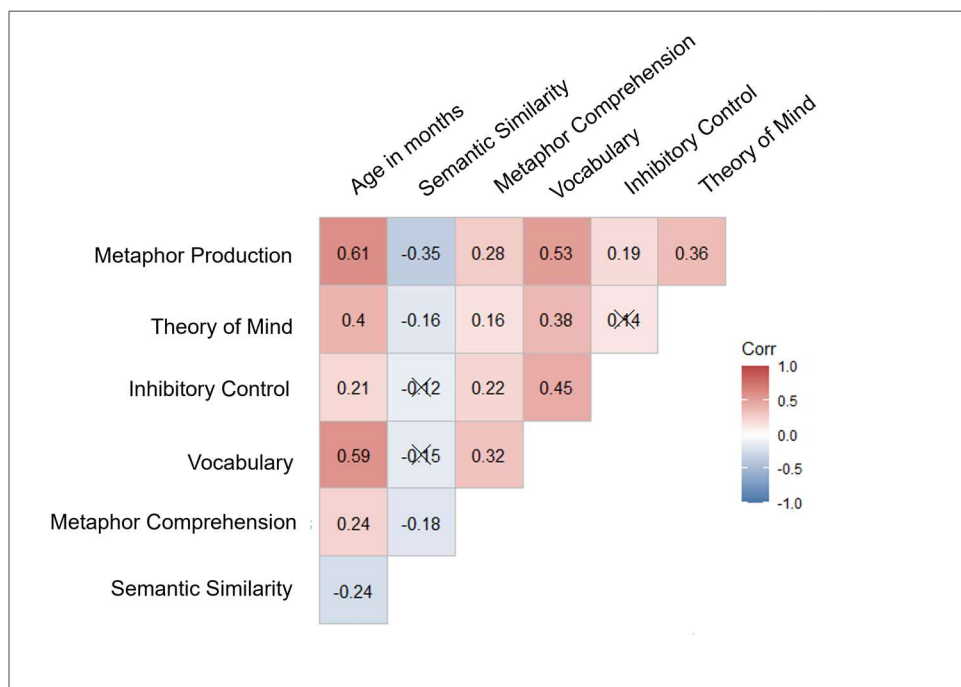
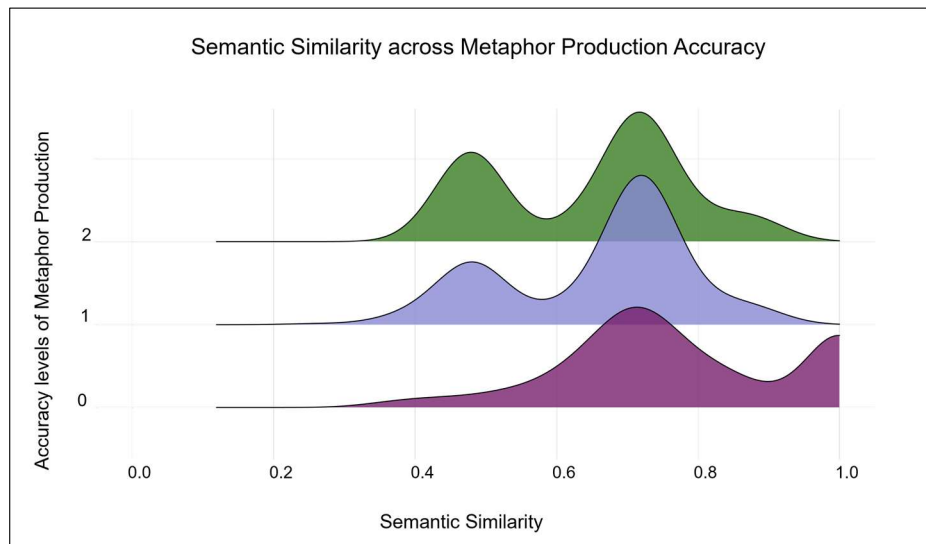


Figure 2.1. Correlogram between Metaphor Production scores, Semantic Similarity values, and linguistic, cognitive, and socio-cognitive variables. Positive correlations are displayed in red and negative correlations in blue. The color intensity is proportional to the magnitude of correlation coefficients. White cells indicate associations at p-value > .05. Age was transformed in months.

Semantic Similarity between the produced metaphors and literal prompts negatively correlated with Age, Metaphor Comprehension, and Theory of Mind. Semantic Similarity significantly correlated with Metaphor Production scores ( $r(165) = -.35, p < .001$ ), specifically, higher scores in Metaphor Production were associated with lower levels of Semantic Similarity (See also Fig. 2.2).



**Figure 2.2. Distribution of Semantic Similarity on Metaphor Production Accuracy levels.** Density plots of Semantic Similarity (y-axis) for each level of Metaphor Production Accuracy (0,1,2; x-axis) are reported.

### 2.3.1. Development of Metaphor Production skills and the role of individual differences

The output of the model on Metaphor Production Accuracy is shown in Tab.2.2.

The model on Accuracy showed a significant effect of age group (5 vs. 4; 6 vs. 5), with 5-year-old children performing significantly better than 4-year-old ones ( $\beta = 0.13, t = 3.85, p < .001$ ) and 6-year-old children performing significantly better than 5-year-old ones ( $\beta = 0.26, t = 7.09, p < .001$ ). A comprehensive visualization of Metaphor Production performance across Age Groups is reported in Figure 2.5a. The model also showed a main effect of Vocabulary ( $\beta = 0.06, t = 3.10, p = .002$ ), with higher vocabulary scores predicting better metaphor production performance (See Fig. 2.3a). Moreover, a main effect of ToM ( $\beta = 0.05, t = 2.89, p = .004$ ) was observed, with higher ToM abilities predicting higher metaphor production scores. A significant interaction between Age groups and ToM was observed ( $\beta = 0.09, t = 2.15, p = .033$ ), showing that the contribution of ToM in promoting metaphor production emerged particularly in the 5- to 6-year-old comparison (See Fig. 2.3b).

Table 2.2 Output of the Linear Model with Metaphor Production Accuracy as the dependent variable

Fixed Effects	Metaphor Production			
	Estimates	CI	t	p
(Intercept)	0.30	0.27 – 0.33	20.23	<0.001
Age [5vs4]	0.13	0.06 – 0.19	3.85	<0.001
Age [6vs5]	0.26	0.19 – 0.33	7.09	<0.001
Vocabulary	0.06	0.02 – 0.10	3.10	0.002
Inhibition	-0.03	-0.06 – 0.01	-1.53	0.128
Theory of Mind	0.05	0.02 – 0.08	2.89	0.004
Metaphor Comprehension	0.01	-0.02 – 0.04	0.56	0.573
Age [5vs4] × Vocabulary	-0.01	-0.09 – 0.07	-0.20	0.843
Age [6vs5] × Vocabulary	0.01	-0.08 – 0.11	0.26	0.797
Age [5vs4] × Inhibition	-0.01	-0.08 – 0.06	-0.24	0.812
Age [6vs5] × Inhibition	-0.06	-0.15 – 0.03	-1.29	0.201
Age [5vs4] × Theory of Mind	-0.03	-0.10 – 0.04	-0.90	0.372
Age [6vs5] × Theory of Mind	0.09	0.01 – 0.17	2.15	0.033
Age [5vs4] × Metaphor Comprehension	0.02	-0.04 – 0.09	0.73	0.468
Age [6vs5] × Metaphor Comprehension	-0.05	-0.12 – 0.03	-1.16	0.246
Subjects	171			
R <sup>2</sup> / R <sup>2</sup> adjusted	0.457 / 0.409			

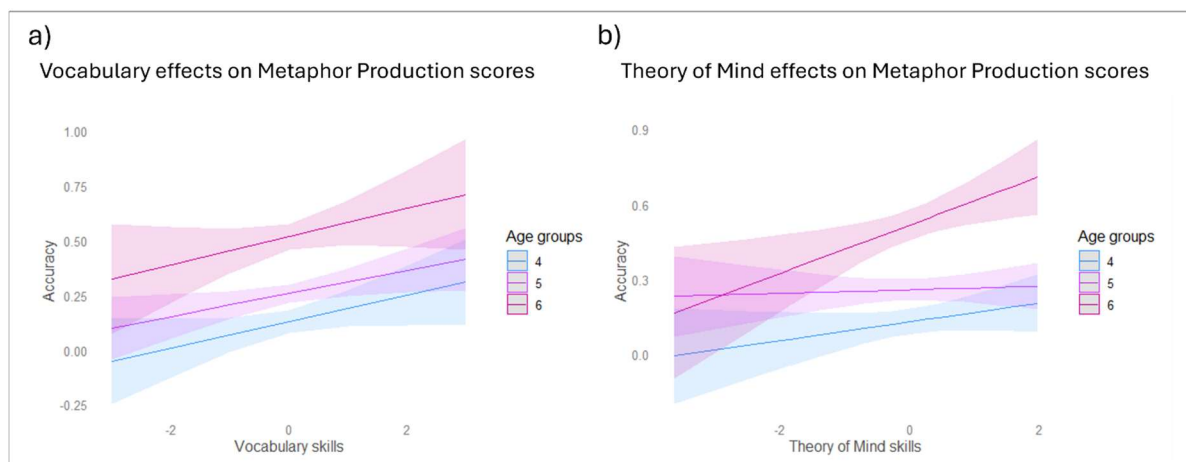


Figure 2.3. Association between Metaphor Production and linguistic and socio-cognitive predictors. The plot reports the relationships between Metaphor Production and Vocabulary (Panel a) and Theory of Mind (Panel b), across Age groups.

### 2.3.2. Effects of Age and Individual Differences on Semantic Similarity

The output of the model on Semantic Similarity is shown in Table 2.3.

The model showed a significant main effect of Age Groups [5vs.4] ( $\beta = -0.05$ ,  $t = -2.27$ ,  $p = .024$ ), with 5 years old children reporting lower values of Semantic Similarity than 4 years old ones (See Fig. 2.5b), and a main effect of Metaphor Comprehension ( $\beta = 0.05$ ,  $t = -2.28$ ,  $p = .023$ ), indicating that higher level of Metaphor Comprehension predicted lower Semantic Similarity. The model also revealed a significant Theory of Mind by Age group interaction for the 5vs.4 comparison ( $\beta = 0.05$ ,  $t = 2.70$ ,  $p = .007$ ).

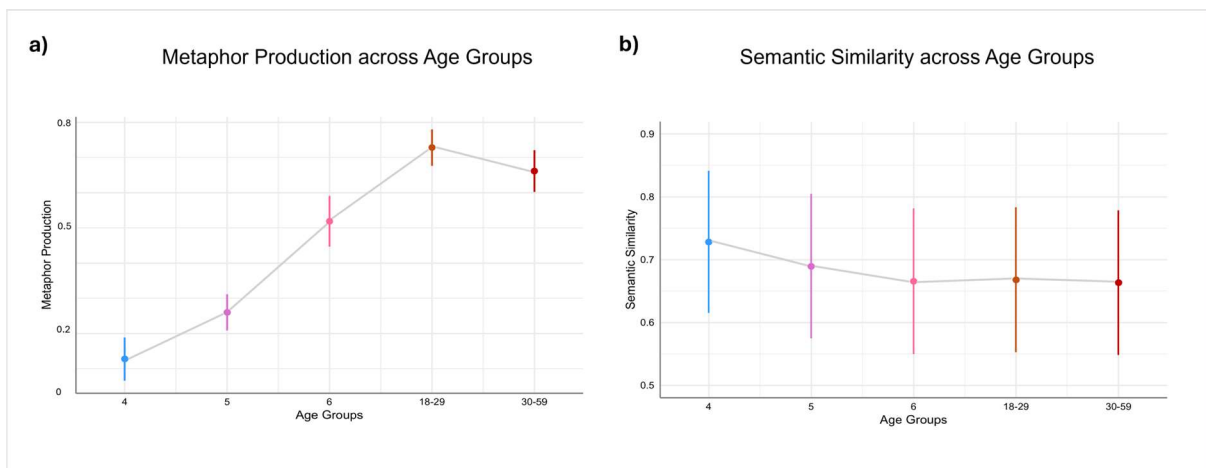
Table 2.3. Output of the Linear mixed-effects model with Semantic Similarity as the dependent variable.

Fixed Effects	Semantic Similarity			
	Estimates	CI	Statistic	p
Predictors				
(Intercept)	-0.39	-0.55 – -0.23	-4.75	<b>&lt;0.001</b>
Age [5vs4]	-0.05	-0.09 – -0.01	-2.27	<b>0.024</b>
Age [6vs5]	-0.03	-0.07 – 0.01	-1.47	0.141
Theory of Mind	-0.01	-0.03 – 0.01	-0.99	0.324
Metaphor Comprehension	-0.02	-0.04 – -0.00	-2.28	<b>0.023</b>
Age [5vs4] × Theory of Mind	0.05	0.01 – 0.09	2.70	<b>0.007</b>
Age [6vs5] × Theory of Mind	-0.02	-0.07 – 0.02	-1.02	0.309
Age [5vs4] × Metaphor Comprehension	0.03	-0.01 – 0.08	1.68	0.094
Age [6vs5] × Metaphor Comprehension	0.03	-0.01 – 0.08	1.44	0.151
<b>Random Effects</b>				
$\sigma^2$	0.02			
$\tau_{00 \text{ ID}}$	0.00			
$\tau_{00 \text{ IT}}$	0.03			
ICC	0.56			
$N_{\text{ID}}$	167			
$N_{\text{IT}}$	4			
Observations	528			
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.047 / 0.583			



### 2.3.3. Comparison with adults

The analysis comparing children and adults revealed that the Young Adults group performed significantly better than the 6 years group ( $\beta= 0.42, t= 4.63, p< .001$ ), while no significant effect of the Adults vs Young Adults comparison was observed. The model on Semantic Similarity confirmed a significant effect of Age Group for the 5vs.4 comparison ( $\beta= -0.04, t= -3.30, p< .001$ ) and revealed the 6vs.5 comparison approaching significance ( $\beta= -0.02, t= -1.93, p= .053$ ), with 6 years old group showing lower value of Semantic Similarity than the 5 years old group. No difference between the 6-year-old and the Young Adults group was observed ( $p>0.1$ ).



**Figure 2.5. Metaphor Production skills and Semantic Similarity values across Age groups.** Panel a) depicts model estimates of Metaphor Production across Age Groups. Panel b) depicts model estimates of Semantic Similarity values across Age Groups.

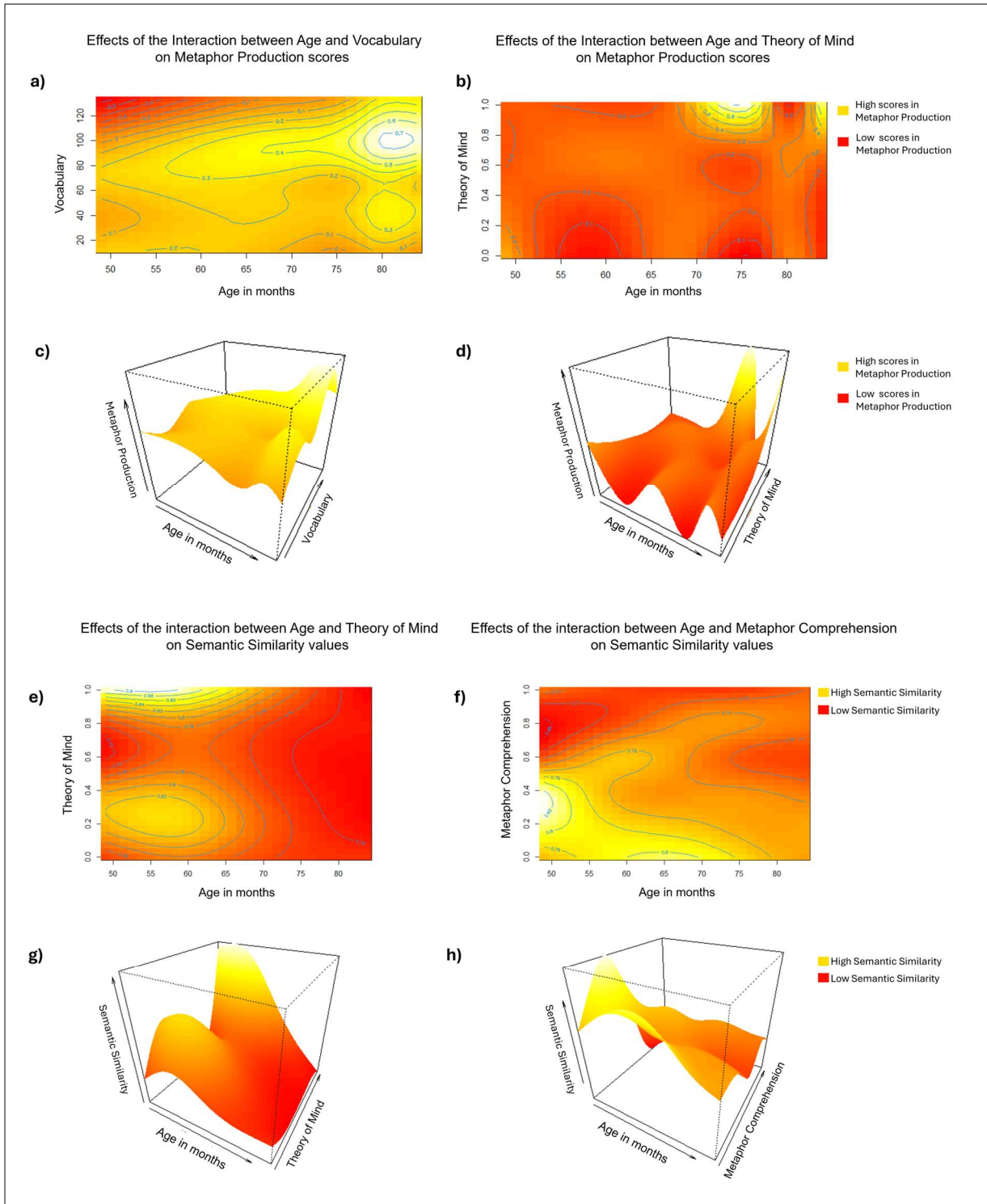
### 2.3.4. Exploratory analysis with Generalized Additive Models

The statistical approach followed so far confirms the hypothesis about a different role of cognitive and socio-cognitive variables in specific developmental periods. The present analysis further examined the role of cognitive and socio-cognitive skills in determining Accuracy and Semantic Similarity patterns of Metaphor Production during the early childhood time window. Models' outcomes are further explored by visual inspection, performed through tensor surfaces, which have to be interpreted as topographic contour maps, with yellow areas indicating high values and red areas indicating low values of Metaphor Production Accuracy and Semantic Similarity.

The GAM on Metaphor Production Accuracy, looking at the simple smooth terms, showed a significant linear smooth for Age ( $\text{edf} = 1.00$ ,  $F = 27.08$ ,  $p < .001$ ), and a significant non-linear smooth for Vocabulary ( $\text{edf} = 4.62$ ,  $F = 3.11$ ,  $p = .010$ ). Moreover, a significant smooth tensor interaction involving Age and Vocabulary ( $\text{edf} = 3.90$ ,  $F = 4.23$ ,  $p = .004$ ) emerged, pointing to differences in the effects of Vocabulary on Metaphor Production through Age increase. The model also revealed a significant smooth tensor interaction involving Age and Theory of Mind ( $\text{edf} = 8.42$ ,  $F = 0.78$ ,  $p = .025$ ). By visually inspecting the 2D and 3D tensor surfaces for the Age by Vocabulary interaction (Fig. 2.6a and 2.6c) is possible to observe a boosting effect of Vocabulary through Age, with a stronger effect in the interval between 75 and 85 months and Vocabulary scores between 80 and 120 (the yellow area with the presence of three isobars between 0.5 and 0.7 Metaphor Production Accuracy, Fig 2.6a). The 2D and 3D tensor surfaces for the Age by Theory of Mind interaction (Fig. 2.6c and 2.6d) display a larger effect of Theory of Mind in the interval between 70 and 80 Age months (the yellow area sees four isobars between 0.3 and 0.6 Metaphor Production accuracy, Fig 2.6b). The exploratory analysis of Metaphor Production scores confirmed and further enriched the main analysis, revealing a constant effect of Vocabulary in promoting Metaphor Production through Age, with a peak in older children and an effect of Theory of Mind at later stages only.

The GAM on Semantic Similarity showed a significant linear smooth for Age ( $\text{edf} = 1.00$ ,  $F = 5.86$ ,  $p = .016$ ), and a significant linear smooth for Metaphor Comprehension ( $\text{edf} = 1.00$ ,  $F = 6.34$ ,  $p = .012$ ). Additionally, a significant smooth tensor interaction involving Age and Theory of Mind ( $\text{edf} = 2.69$ ,  $F = 0.48$ ,  $p = .032$ ) emerged, pointing to differences in the effects of Theory of Mind on Semantic Similarity through Age. Importantly, the model also revealed a significant smooth tensor interaction involving Age and Metaphor Comprehension ( $\text{edf} = 6.05$ ,  $F = 1.08$ ,  $p = .020$ ), pointing to differences in the effects of Metaphor Comprehension on Metaphor Production through Age increase.

By visually inspecting the 2D and 3D tensor surfaces for the Age by Theory of Mind interaction (Fig. 2.6e and 2.6g) is possible to observe a large effect of Theory of Mind on Semantic Similarity values roughly between 50 and 55 age months, indicated by the dark red area with three isobars (between 0.76 and 0.72 cosine values; Fig 2.6e), suggesting a specific effect of Theory of Mind in promoting the use of more distant words at earlier stages. By visually inspecting the 2D and 3D tensor surfaces for the Age by Metaphor Comprehension interaction (Fig. 2.6h and 2.6f) is possible to appreciate an effect of Metaphor Comprehension through Age, with a larger effect at early stages: the yellow area extended horizontally with three isobars (0.82-0.78 cosine values) suggest an effect of lower levels of Metaphor Comprehension in determining higher values of Semantic Similarity. The dark red area sees three isobars between 0.72 and 0.68 cosine values (Fig 2.6h) highlighting a large effect of higher levels of Metaphor Comprehension abilities in determining lower values of Semantic Similarity.



**Figure 2.6. Visualization of the exploratory analysis.** The figure depicts eight panels. Panels (a) and (c) represent respectively the 2D and the 3D surface for the effect of Vocabulary (on the y-axis) through Age development (x-axis) on Metaphor Production accuracy (the color-scale). Panels (b) and (d) represent respectively the 2D and the 3D surface for the effect of Theory of Mind (on the y-axis) through Age development (x-axis) on Metaphor Production accuracy (the color-scale). Panels (e) and (g) represent respectively the 2D and the 3D surface for the effect of Theory of Mind (on the y-axis) through Age development (x-axis) on Semantic Similarity values (the color-scale).

Panels (f) and (b) represent respectively the 2D and the 3D surface for the effect of Metaphor Comprehension (on the y-axis) through Age development (x-axis) on Semantic Similarity values (the color-scale).

### 3. Discussion

Our study aimed to provide novel insights into metaphor production skills acquisition during early childhood, by using cross-sectional assessment and innovative distributional semantics approaches. To meet our aims we adopted an elicitation task asking children to produce metaphors given literal prompts. In addition to scoring answers based on metaphor accuracy, we computed Semantic Similarity between the literal prompts and the metaphor produced. We expected to observe a linear developmental trajectory during early childhood, supported by linguistic, cognitive, and socio-cognitive skills to different extents based on the developmental stage, and an increase of the semantic distance between the prompt and the metaphor words with age, driven by vocabulary skills.

Results on Metaphor Production Accuracy met the hypothesis, showing a progressive increase in metaphor production abilities from 4 to 6 years of age, with vocabulary skills supporting metaphor production across age groups and Theory of Mind boosting the development of these skills, particularly in the oldest group. Conversely, inhibitory control and metaphor comprehension skills didn't show any effect, pointing to the developmental segregation of receptive and expressive metaphor skills. Our predictions on semantic similarity were only partially confirmed: while we observed the expected decrease in the similarity between the literal prompt and the produced metaphor in 5 years old children (e.g., *That dog is a tooth*) vs. 4 years old ones (e.g., *That dog is a cat*), this effect was not predicted by vocabulary but rather by Theory of Mind, opening a window on what drives the conceptual operations underlying metaphor production. Adults outperformed children in Metaphor Production Accuracy but showed the same Semantic Similarity pattern as 6-year-old children.

As a first consideration, our work pointed out that early childhood is a period of crucial development for metaphorical competence: at 4 years of age, children rarely produce salient

metaphors, mainly creating obscure associations (e.g., *white dog-ball* or *white dog-elephant*); at 5 years of age, the performance increases and more salient metaphors and similes started to arise, continuing to increase from 5 to 6. Thus, children's ability to produce metaphors seems to develop linearly during early childhood (Vosniadou, 1987). There are a number of factors supporting this process throughout middle childhood. In particular, our analysis revealed a strong support of vocabulary skills during these years, indicating that children become more able to produce apt metaphor in parallel with the growth of their lexical knowledge. Notably, our finding is in contrast with evidence reported for older children, according to which receptive and expressive vocabulary skills play a minor role in metaphor production typical development (Kasirer & Mashal, 2016; Lampri et al., 2024). The explanation for this discrepancy might be purely developmental. While vocabulary is key in early phases, with children who have adequate semantic knowledge of the prompt being more able to activate a rich array of possible vehicles, which in turn increases the probability of producing salient metaphorical associations, at later stages, once vocabulary knowledge is more consolidated, other cognitive skills, such as working memory and Theory of Mind may become more relevant and support more complex productions. Consistently with this explanation, in L2 learners, the ability to generate metaphors is strongly linked to vocabulary (Gaskins & Rundblad, 2023; O'Reilly & Marsden, 2023), supporting the idea that a refined conventional semantic knowledge permits an aware violation of conceptual categories for metaphorical purposes (Gardner & Winner, 1978).

Another factor that drives the development of metaphor production is Theory of Mind. Interestingly, in our data, Theory of Mind had a role only in the improvement in the oldest group (6-year-olds). In these phases, children were described to deal with the *appearance-reality* distinction (i.e., the possibility of conceiving the co-existence of different representations of the same object; Gopnik & Astington, 1988) which may have a lot in common with the *literal-nonliteral* distinction (Vosniadou, 1987). Specifically, until 6-7 years of age, when presented with an object (e.g., a sponge) that occasionally *looks like* another object for its shape or color (e.g., a rock), children often

verbally describe the appearance (e.g., *the sponge looks like a rock*) as being the reality (e.g., *that object is a rock*) (Flavell et al., 1986; Sapp et al., 2000). Even if children are somehow aware of this distinction since earlier phases, the full-fledged ability to manage the co-existence of different realities may be a crucial acquisition for metaphor production development, and it may be the prerequisite to consider children's production as genuine metaphorical uses rather than reality overrides (Gardner & Winner, 1978). It is interesting to consider the effect of Theory of Mind together with the effect of vocabulary. The fact that the latter affects all age groups while the former plays a role only at later stages might suggest that they support different aspects of metaphor production: on the one hand, vocabulary provides the necessary knowledge to look for conceptual associates and retrieve the relative lexical elements, Theory of Mind might support a more sophisticated facets of metaphor production, i.e., the ability to genuinely reframe the prompted concepts with an alternative one, which is reached later in development. The additional exploratory analysis with GAMs provides further support to these considerations. Specifically, tensor surface representation confirmed that Theory of Mind has a larger effect in older children but also revealed that this larger effect is observed for older children who reached the highest score on the Theory of Mind scale. Notably, the task to assess Theory of Mind was the Wellman and Liu scale (Wellman & Liu, 2004), in which the higher ladder consists of the *Hidden Emotion* task. This task evaluates a child's ability to accept that people may feel an emotion (*reality*) but show another emotion (*appearance*) for several reasons (Wellman et al., 2011; Wellman & Liu, 2004). This additional evidence suggests that the ability to master several representations of reality, might be a prerequisite for full-fledged metaphor production skills. Concerning inhibition, we did not observe any effect on metaphor production in our sample. This is at odd with the largely reported involvement of inhibitory control in supporting metaphor competence in the adult population (Bambini et al., 2011; Chiappe & Chiappe, 2007; Sana et al., 2021). It is possible that inhibitory control comes into play at later stages, in line with evidence of its involvement at 15 years of age (Carriedo et al., 2016),

becoming particularly relevant in older adults (Bambini et al., 2021) or vulnerable groups (Frau et al., 2024), or that it may depend on task complexity and topic familiarity (Chen et al., 2025).

Another interesting non-significant effect is the one observed for metaphor comprehension. Recent insights on metaphor comprehension showed that at six years of age, children start to abandon literal interpretations of metaphorical expression, with a strong positive impact on metaphor comprehension (Lecce, Pompei et al., in preparation). However, albeit a weak positive correlation, metaphor comprehension abilities did not predict metaphor production accuracy, reinforcing the idea that these two dimensions of metaphorical competence may follow different developmental routes and capitalize to a different extent on cognitive skills. This dissociation is better observed in clinical conditions, such as schizophrenia, in which, while the receptive dimension of figurative language is severely impaired (Bambini et al., 2020, 2025; Mashal et al., 2018), metaphor production during spontaneous speech seems somewhat preserved (Despot et al., 2021; Elvevåg et al., 2011).

Moving to a deeper consideration of the conceptual processes underneath metaphor production development, the analysis of Semantic Similarity patterns showed that with age, children progressively use words that are more distant to the topic. The use of more distant words correlated also with higher accuracy scores, confirming that a certain degree of semantic distance is a prerequisite for metaphoricity (Katz, 1989; Winter & Strik-Lievers, 2023). Most of all, the observed correlation suggests that children become more accurate in metaphor by exploring the semantic network more broadly. This idea is also supported by the evidence that, at this stage, children's semantic network grows (Wojcik, 2018; Wojcik & Kandhadai, 2020), allowing for a series of high-level language tasks. In this light, we can explain the link between semantic similarity and metaphor comprehension, also confirmed by the exploratory analysis. While metaphor comprehension accuracy was not related to metaphor production accuracy, we observed a link between receptive metaphorical abilities and semantic distance of the terms used in the produced metaphor, which may reflect children's conceptual maturation. In particular, children who are able to produce more

distant associations in the metaphor production task are better at exploring the semantic network and this might go in hand with the ability to identify and explain the link between elements in a presented metaphor. The most surprising finding related to semantic distance is perhaps the significant association with Theory of Mind. In particular, Theory of Mind abilities predicted a greater distance between the literal prompt and the produced metaphors in the younger group. We can explain this finding by considering that the awareness that the interlocutor may have different knowledge or experience of certain words may guide the exploration of the semantic network in a social-oriented fashion (Doherty & Perner, 1998; F. Happé & Loth, 2002). This may result in avoiding egocentric answers that excessively capitalize on personal experience that does not allow a genuine conceptual shift (e.g., *that dog is like my puppet*) or lead to obscure associations, and ascertaining that the word used stands at an adequate distance from the prompt to evoke a shared experience (Cacciari, 2008). The exploratory analysis further revealed that the effect of Theory of Mind on semantic similarity across age is larger at early stages in the interval between the third and the fourth ladder of the Wellman and Liu scale, suggesting that passing the *knowledge access* and the *false belief* tasks (i.e., having the ability to understand that other may have different knowledge and beliefs than ones' own) is a necessary step to choose words with adequate distance from the literal prompt.

The adult group was here included for the purpose of providing evidence of a developmental endpoint of metaphor production (Gottfried, 1997). We observed that young adults and adults outperformed 6-year-old children, indicating that in early childhood metaphor production is not full-fledged. However, semantic similarity values observed in the adult groups are not different from the ones observed in 6-year-old children. At 6 years of age children already reached an optimal level of distance from the literal topic of their metaphorical expressions, however, other factors play a role in generating optimal metaphors. Besides, studies on the adult population suggest that metaphor production does not correspond to an asymptotic increase of the semantic distance between a target concept and the figurative word use, but rather it exists a middle-ground of good

metaphors (Clevenger & Edwards, 1988; Katz, 1989; Kintsch & Bowles, 2002). Hence, even if adults improve their metaphor production skills, correct insights on semantic network exploration gradient for metaphorical purposes are reached during early childhood. It is possible to assume that higher levels of semantic distance might be observed in the adult population when dealing with creative metaphors (Beaty & Johnson, 2021; Winter & Strik-Lievers, 2023).

Overall, by combining behavioral and distributional semantic approaches and refined statistical methods, this study provides the first robust insights into metaphor production development and its cognitive and socio-cognitive correlates in early childhood. Productive metaphor competence seems to have a linear developmental trajectory in this phase, supported by a complex interplay of other skills. Our data do not confirm the hypothesis of a U-shaped trajectory for metaphor production, at least for the first segment of the trajectory (Gardner et al., 1975). Previous studies may have observed different mechanisms from metaphor production, namely children's natural exploitation of the fictive framework of pretend play, thus observing magical expression rather than figurative uses. Conversely, in our task, children were asked to *communicate* and *deliver a message* about a specific feature (e.g., *being white*), using a word expected to be *diagnostic* in its (broadened or narrowed) denotation, resonating with the naturalistic process of metaphor generation (Carston, 2018; Katz, 1989). Hence, our task requires a figurative use of words, which is possible only by navigating the realm of pragmatics and conceptual associations, rather than magic.

Receptive vocabulary promotes metaphor production throughout development, suggesting that vocabulary offers the language platform for generating metaphors. Theory of Mind skills seem to play a role, particularly between the ages of 5 and 6 years, which is also the developmental phase in which children reach a good metaphor production accuracy. This suggests that Theory of Mind might play as a springboard for producing accurate metaphors, allowing children to consider others' mental states and the context shared with the interlocutor, dropping egocentric answers, common in earlier stages. Theory of Mind seems also to promote a broader exploration of the semantic network, possibly more directed to socially appropriate metaphors, based on shared

knowledge. This complex interplay of vocabulary, semantic knowledge and mindreading skill, plus a genuine skill in conceptual shifts, is what allows children to learn how to turn a white dog into a cloud, paving the way towards a more sophisticated use of language.

## STUDY THREE

### ‘BROCCOLI IS CANDY’: THE ROLE OF METAPHORS IN CHILDREN’S PERSUASIVE COMMUNICATION<sup>4</sup>

#### Abstract

Persuasion is a complex communicative process aiming at influencing others’ beliefs or behaviors. Imbuing everyday communication, persuasion is a crucial skill for children to manage social interactions successfully. While theoretically persuasion has been linked with the mastery of figurative language and with pragmatics more broadly, there is a scarcity of empirical evidence exploring the relationship between persuasion and metaphor skills. Here we tackle this issue in early childhood by examining whether individual differences in metaphor skills are related to those in persuasion. A sample of 167 children (age 4-6 years) was assessed for persuasive abilities alongside metaphor comprehension and production, in addition to vocabulary and working memory skills as control variables. Results showed an improvement in persuasive skills at 5 years of age. Across ages, children preferred to use positive persuasive strategies (i.e., offering rewards) over negative ones (i.e., punishments), while psychological strategies relying on mitigation and modeling were rarely used. Regardless of the type of strategy, persuasion correlated positively with vocabulary skills. Interestingly, greater use of psychological persuasive strategies was associated with better metaphor production skills (being conversely hampered by working memory), while no effect of metaphor comprehension was found. Overall, these findings suggest that some aspects of metaphor skills, within the broader set of pragmatic competencies, might be a driving factor in achieving a high-level persuasive style. Such aspects possibly deal with the functions of metaphors to favor flexible conceptualization and social use of language.

---

<sup>4</sup> This chapter is a manuscript currently under review for publication as “Pompei, C., Lecce, S., Del Sette, P., Didoni, E., Bischetti, L., & Bambini, V. ‘Broccoli is candy: the role of metaphors in children persuasive communication’” in the *Language Development and Research* peer-reviewed journal.

### 3.1. Introduction

Be it a child who wants to spend more time in front of the TV, a charismatic leader who needs to win the loyalty of his constituents, or a lawyer who has to convince the jury of his client's innocence, making information appear convincing or persuasive towards one's interests is a key aspect of human communicative competence. The ability to persuade others is a challenge that individuals undertake early in development and then need to adapt to a wide range of situations, drawing on multiple linguistic and psychological resources. In this study, we investigated the blooming of persuasion through early childhood, while considering its cognitive and linguistic correlates and, in particular, the role of receptive and expressive metaphor skills.

The conceptualization of persuasion can be traced back to classical rhetorical studies, particularly to Aristotle, who conceived persuasion as the intrinsic purpose of the art of oratory (Montanari, 1996). Integrating more recent considerations, persuasion is defined as a complex communicative process aiming at influencing others' beliefs or behaviors (Freeley & Steinberg, 2009). Furthermore, persuasion also refers to the set of cognitive processes responding to the communication of a message whereby interlocutors change their "attitudes or behavior regarding an issue", via the use of linguistic tools (Perloff, 1993) with the intention of modifying the cognitive environment of the audience (Sperber & Wilson, 1986). In this respect, persuasion relies both on expressive aspects (dealing with the verbal form in which it is delivered) and psychological elements used to achieve a perlocutionary goal, i.e., the consequential effects on the receiver. In other words, persuasion is successful not only by delivering a proposal or suggestion but depends also on the receiver adopting a certain psychological attitude (Sbisà, 2013).

Focusing on the linguistic level, accounts emphasize the pragmatic nature of persuasion, pointing out that nothing like the ability to engage and persuade others to think, say, or do what we would like them to do allows us to participate in social communication (Dillard, 2010). Indeed, to achieve different persuasive goals, speakers need a variety of strategies, which include structural linguistic mechanisms such as the use of requests, referential, expressive, and phatic utterances (Nord, 2008)

but mostly exploit pragmatics-based tools (Khafaga et al., 2023). Among the wide ensemble of pragmatic devices, some are said to contribute more (or more effectively) than others to achieving strong persuasive effects (Baldi, 2020): being able to evoke a common cognitive ground, a sort of “shared territory” favoring possible joint visions and solutions (Bülow-Møller, 2005). These include implicatures and presuppositions (Lombardi Vallauri, 2022), and in particular the clever use of figurative language, with the most prominent role played by metaphor (Sopory & Dillard, 2002; Van Patten, 2013).

The link between metaphor and persuasion can be best understood by inspecting the definition and scope of metaphor itself. Just like persuasion, metaphor has a two-fold nature, including specific linguistic processes and psychological effects. According to account in the pragmatics of language, metaphors are loose uses of words (Carston, 2010a), by which concepts conveyed by lexical entries are broadened or narrowed towards the creation of novel and *ad hoc* concepts. For instance, upon hearing an utterance such as *Get the nugget of ice cream in the refrigerator*, the listener needs to use the context to first expand the lexical denotation of the metaphorically used word *nugget* (dropping the feature of being made out of gold, while promoting the aspects related to being precious or a small quantity) and then derive the intended meaning of the sentences as referring to a small amount of very good ice cream (for a more detailed description, see, e.g., Sperber & Wilson, 2008). Mastering these processes relies on a complex interplay of cognitive and neural mechanisms (Bischetti et al., 2024) that reflect the pinnacle of our linguistic and communicative skills and mature throughout development going through different stages (Falkum, 2018; Lecce, et al., 2019; Pouscoulous, 2014).

At a more psychological level, metaphors are deeply intertwined with our conceptualization of various phenomena, particularly complex ones (Lakoff & Johnson, 1980). The list of cognitive and social functions of metaphor is long and includes favoring the processing of complex concepts from different points of view. Metaphors can make technical concepts accessible to non-experts and provide a deep conceptualization of phenomena by eliciting thoughts about a topic and

increasing the memorability of concepts. Moreover, from the emotional point of view, metaphors can help in expressing feelings. (Christidou et al., 2004; Fainsilber & Ortony, 1987; Katz, 1992). For these reasons, metaphorical devices in different modalities (e.g., verbal and visual) are largely used in politics and advertising to persuade people to buy certain products and modify certain behaviors or ideas (Burgers et al., 2016; McQuarrie & Phillips, 2005). Prominent psychological accounts identify different ways in which metaphors achieve conceptualization and, in turn, persuasion (Holyoak & Stamenković, 2018; Sopory & Dillard, 2002a). According to cognitive linguistics, metaphor promotes conceptualization via embodied cognitive mappings and such processes of mental simulation are key for the comprehension of the persuasive message (Sopory & Dillard, 2002a). Differently, for the analogy account, metaphor stimulates the identification of similarities between concepts (Holyoak & Stamenković, 2018) and, in turn, focusing on similar target-base relations evokes a richer set of associations in semantic memory which ultimately leads to greater elaboration and persuasion compared to literal language.

Across studies, what emerges is especially the flexibility of metaphors in conveying messages of various kinds, promoting either negative or positive inferences. A recent review pointed out that metaphors are often employed in political discourse to emphasize negative consequences while downplaying positive ones, (e.g., on immigration policies), with large-scale social consequences (Boeynaems, et al., 2017). Studies on visual metaphors in advertising reported that metaphors suddenly convey aspects related to rewards or positive outcomes of certain actions. For example, a detergent was advertised with a picture of a bomb on the side; participants reported that they immediately perceived aspects related to the effectiveness of the product, in the absence of consideration of negative features related to the bomb (McQuarrie & Phillips, 2005).

Overall, metaphor seems to reach persuasive effects via multiple psychological processes (Ottati & Renstrom, 2010), thus constituting a true high-level persuasive strategy, capable of changing the interlocutor's mental state and attitudes toward the topic. These effects seem to be attested also in children (Pompei et al., 2024). Despite the numerous connections between persuasive strategies

and metaphors highlighted above, a clear understanding of whether and how the two abilities are related to one another is still missing, especially in young children.

The study of persuasion is a relatively underexplored area within the field of developmental social psychology. Investigations on children at-home conversations have reported that children as young as 5 years exhibit simple persuasive tactics in their family interactions (Bartsch et al., 2011) or with their play partners (Köymen et al., 2016). During development, persuasion skills are also articulated in different strategies that begin with the acquisition of the simpler ones such as positive (i.e., based on the reward system or emphasizing positive outcomes) and negative (i.e., based on the punishment and threat system, or focused on negative outcomes), and that peak with the elaboration of higher-level strategies, involving mitigation (e.g. using trade-offs, providing alternative strategies for approaching the phenomenon) and modeling (e.g. lead by example) (Peterson, Slaughter, & Wellman, 2018; Lonigro et al., 2017). Studies have also shown that persuasive skills mature together with other skills, including Theory of Mind, building on the idea that being able to self-represent others' beliefs and mental states is necessary to produce strategies useful to effectively influence others' beliefs, opinions, and behaviors (Barajas, Linero, & Alarcón, 2022; Lonigro et al., 2017). Other studies focused on the role of language and showed the key role of high-level linguistic skills in the transition toward the most complex forms of persuasive attempts. Nippold, Ward-Lonergan, and Fanning (2005) reported that, while children as early as age 11 showed adult-like performance for syntactic and semantic features of the persuasive production, adults outperformed children and adolescents in discourse-pragmatic dimensions, also providing different persuasive argumentations and using different types of strategies. Consistently, a pioneering work by Crowhurst & Piche (1979) reported that young adolescents, when asked to direct persuasive attempts to different target audiences (e.g., teachers vs. peers) *via* essay, still struggle in modulating their linguistic repertoire taking into account the interlocutor. More recent studies have suggested that persuasive performance in adolescents may be related to a large variety of discourse features, also depending on working memory skills (Heilmann, Malone, & Westerveld,

2020). This sparse evidence on the importance of high-level language skills suggests that pragmatics and metaphor in particular, as a fundamental aspect of pragmatic competence (Domaneschi & Bambini, 2020), might play a role in the maturation of persuasive skills.

### ***3.1.1. The present study***

In the present study, we investigated the developmental pathway of persuasive communication in early childhood, focusing on the role of metaphorical competence while accounting for more general linguistic and cognitive abilities. Specifically, we considered the role of vocabulary, working memory, and metaphor expressive and receptive skills in determining children's use of different persuasive strategies, namely differentiating between those utterances focusing on positive or negative outcomes or applying psychological mitigation and modeling strategies. Our hypotheses were that: (a) persuasion skills begin to develop at around 5 years of age, in line with previous observational studies (Bartsch et al., 2011); (b) metaphor skills scaffold high-level persuasive strategies over and above general linguistic and cognitive skills. We based the latter prediction on two main findings emerging from the literature described above: first, the role that metaphor use plays in modulating psychological processes during high-level persuasive argumentation (Ottati & Renstrom, 2010), also in children (Pompei et al., 2024), and second, the role of high-level language aspects in the development of persuasive skills (Nippold et al., 2005). These hypotheses were explored with a cross-sectional design study in which we employed an elicitation task for persuasion assessment and both receptive and expressive tasks for metaphor skills.

## **3.2. Methods**

### **3.2.1. Participants**

A sample of 246 children ranging in age from four to six years was enrolled in the present study. Children were recruited from local schools and kindergartens located in Lombardy, Italy. Before running the data analysis, we excluded children who met one or more of the following criteria: being diagnosed with intellectual disabilities or neurodevelopmental disorders, hearing deficits, and not having acquired the Italian language before the age of 3. The final sample included 167 children

including 52 4-year-old children (age range: 4;0, 4;11;  $M_{age} = 4;6$ ), 76 5-year-old children (age range: 5;0, 5;11;  $M_{age} = 5;5$ ), and 39 6-year-old children (age range: 6;0, 6;11;  $M_{age} = 6;6$ ). We ran an after-the-fact power analysis (O'Keefe, 2007) focusing on the main research aim, namely the relationship between persuasion and metaphor skills. We found that with 167 participants and  $\alpha = .05$ , we achieved a high statistical power ( $1 - \beta = .99$ ) to detect a moderate effect (in line with the overall effect  $r = .42$  reported in the metaanalysis by Sopory & Dillard, 2002b). Both parents signed written informed consent, and children were provided with age-appropriate information about their participation in the study. The study was approved by the Local Ethical Committee of the Department of Brain and Behavioral Sciences of the University of Pavia (n. protocol 029/2019) and followed the principles of the Declaration of Helsinki.

### 3.2.2. Procedure

Each child took part in two individual sessions administered in a silent place during school time. Each session lasted approximately 20 minutes. During the first session, children were assessed for their vocabulary and working memory, while the second session was dedicated to the assessment of metaphor skills (both receptive and expressive) and persuasive skills.

### 3.2.3. Measures

#### 3.2.3.1. Vocabulary

The Italian version of the Peabody Picture Vocabulary Test - Revised (PPVT-R; Stella, Pizzoli, & Tressoldi, 2000) was used to assess children's vocabulary skills, which has been shown to have excellent split-half reliability value ( $r = .88$ , Dunn & Dunn, 1981) and high validity against the McCarthy Scales of Children's Abilities ( $r = .79$ , Naglieri & Maxwell, 1981). The PPVT-R includes 175 verbal stimuli and measures receptive vocabulary with a picture selection task: children are asked to choose, among four images, the one that best describes the meaning of the word uttered by the experimenter. Following the standard procedure, for each child, the chronological age is used to set individual test starting points. An incorrect response to any of the first 8 benchmark items results in a retraction of the starting point. Six consecutive errors in an 8-item block result in

the interruption of the test. The total score consists of the number of correct answers (score range: 0-175).

### 3.2.3.2. *Working Memory*

Working Memory was assessed using the backward word recall task (Lanfranchi et al., 2004), a largely used task included in a battery measuring verbal working memory globally quite reliable (overall Cronbach's alpha = .56). In this task, children are presented with lists of two to five words and asked to repeat the list immediately, in reverse order. Two items for each of the four levels (2-word; 3-word; 4-word; 5-word trials) of difficulty are administered; failure in both items corresponded to the interruption of the test. The total score consists of the number of correctly achieved levels (score range: 0–8).

### 3.2.3.3. *Metaphor Comprehension*

To assess metaphor comprehension, we used the newly developed multiple-choice version of the Physical and Mental Metaphor task (Lecce, Pompei, in prep), originally created in the verbal explanation task format (Del Sette et al., 2020; Lecce et al., 2019). The task was adapted by creating multiple choice options for 6 verbal items extracted from the verbal explanation version of the task, plus 4 novel metaphors, which show acceptable item-total correlation (range of Pearson's correlation  $.37 < r_s < .57$ ). The test includes 10 metaphors: 5 physical (i.e., metaphors that capitalize on physical properties, such as *Dancers are feathers*) and 5 mental (i.e., metaphors that capitalize on mental or psychological aspects, such as *The teacher is an icicle*). Children are asked to select the best fitting explanation for each metaphor, choosing among a set of three options, presented both verbally and visually (correct: physical, e.g., *They are light*, with an image representing dancers jumping almost weightless; mental, e.g., *She is strict*, with an image representing a teacher scolding her pupils; incorrect literal: physical, e.g., *They are dressed in white*, with an image representing white dressed dancers; mental, e.g., *She likes cold things*, with an image representing the teacher eating an icicle; incorrect unrelated: physical, e.g., *They are short*, with an image representing short dancers; mental, e.g., *She likes to sing*, with an image representing the teacher singing). Answers are scored

as correct (1) for appropriate interpretation or incorrect (0) for literal and unrelated (score range: 0-10).

#### 3.2.3.4. *Metaphor Production*

To assess metaphor production abilities, we adapted the elicitation task by Cortés, Cobos, & Tarbox (2018). This task is composed of 4 items in which children are given literal prompts and asked to produce metaphors. This task is composed of 4 items in which children are given literal prompts and asked to produce metaphors. Based on our sample, the task showed acceptable internal consistency (Cronbach's alpha = .78) and excellent inter-rater reliability (92% agreement between coders; Pompei et al., 2023). More in detail, they are provided with a brief description of an object and then instructed to change the prompted literal word with a metaphoric equivalent (e.g., to describe a dog being white without using the word 'white'). Salient metaphors (e.g., *The dog is a cloud*) are defined as those figurative associations for which the implicated meaning matches – on a common ground based on shared word knowledge – with the prompt word (e.g., *white*), and are scored 2 points; similes (e.g., *The dog is like a cloud*) and non-salient metaphors (e.g. *The dog is an ice-cream*) defined as metaphors built around a vehicle that highlight features only marginally matching with the prompt word 1 point, and literal or no answers were scored 0 points (score range: 0-8).

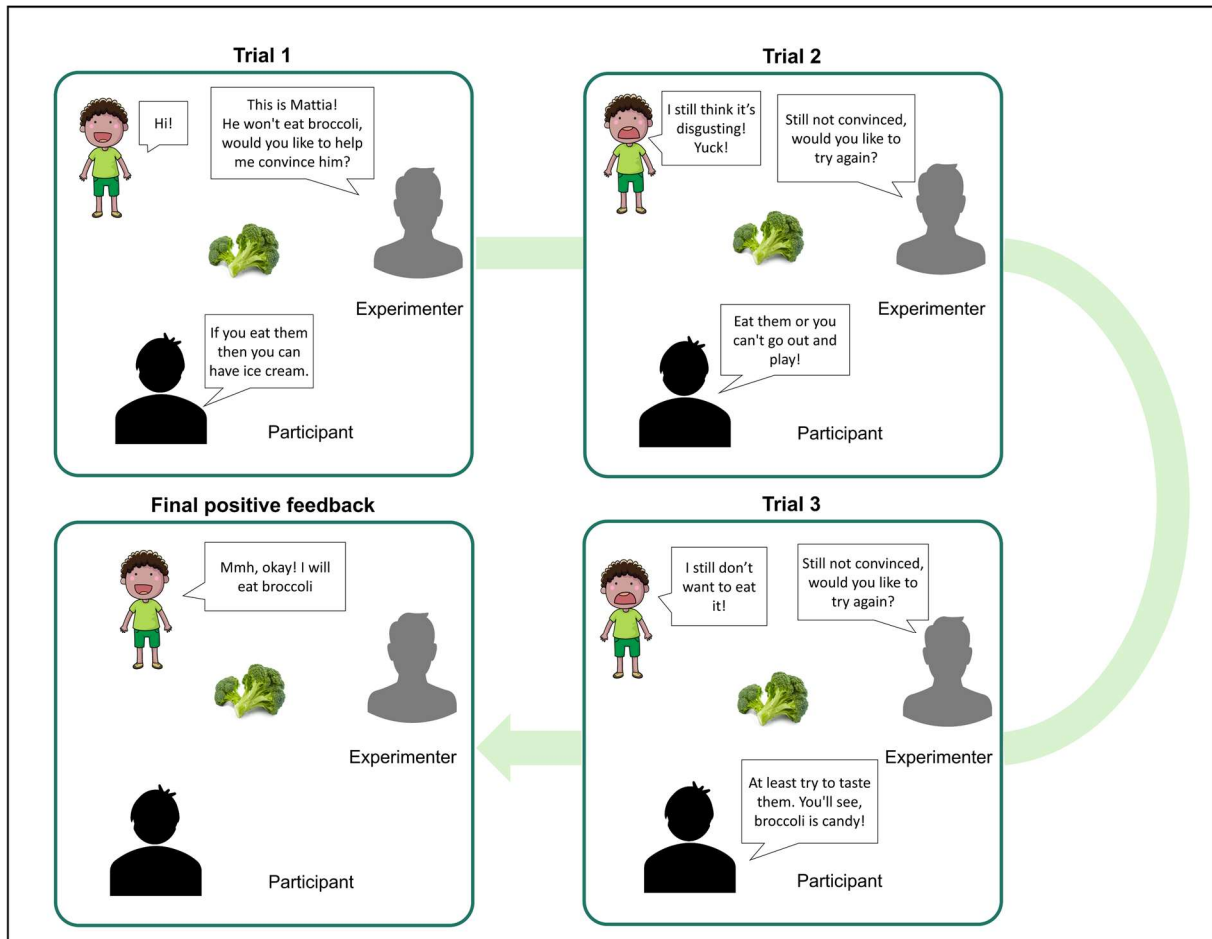
#### 3.2.3.5. *Persuasion*

To measure persuasive skills, we adapted the interactive task by Peterson and colleagues (2018). In this task, children sit at the table with the experimenter interacting with a puppet called Mattia. The puppet is introduced as a peer to them. The task consists of two items that require the child to persuade the puppet to perform an action, i.e., to eat broccoli and to brush his teeth, using only words. In the original version of the task, two authors independently coded a random set of 33 children's complete transcripts (i.e., 198 responses representing 33% of the 594 total responses produced across six episodes per child) showing an excellent inter-rater reliability (94% agreement between coders). Each item includes three trials: in the first two attempts, the puppet provides

negative feedback to persuasive attempts; after the negative feedback, the experimenter challenges the child to try again. Regardless of the quality of the child's persuasive production, the third trial is always successful (e.g., *Great! You have convinced him!*), and the puppet eventually agrees to perform the action (See Fig.3.1. for the schematic representation of the task). Following the recommendations proposed by Peterson and colleagues (2018), unrelated answers or no answers are scored as 0; occurrences of persuasive arguments are counted as 1 and summed up: for each child, the final score consists of the total number of persuasive statements uttered by the child. Repeated arguments within the same item are not considered. Persuasive arguments are further categorized as Positive, Negative, or Psychological. The Positive label refers to sentences that use good consequences and rewards to achieve persuasion (e.g., *If you brush your teeth, then they become strong, If you eat it, we'll give you a gift*). The Negative label refers to sentences that exploit threats and punishment to reach the aim (e.g., *If you don't brush your teeth, you get cavities, If you don't eat broccoli, you stay small*). The Psychological strategies are statements introducing a compromise or a modulation of the request as an incentive to achieve persuasion (e.g. *Mash the broccoli and try putting lemon on it, or Try putting on a toothpaste you like*) or those statements that exploit the use of one's experience as role model (e.g. *Do you know that I always eat broccoli?*, or *My friend Michele also washes them. When I wash my teeth, he also washes them*). In addition to Peterson and colleagues (2018) coding, we annotated the presence of metaphors in the answers (e.g., *Broccoli is candy*), which resulted in a metaphor count measure.

To confirm the inter-rater reliability of the persuasion test adapted in Italian two authors (CP, ED) independently coded 20% of data and the agreement between raters was determined with the interclass correlation coefficient (ICC). The ICC was calculated on a mean-rating ( $k = 2$ ), average-agreement, 2-way mixed-effects model, and values were reported along standard guidelines (Koo & Li, 2016).

The average agreement for each strategy of the coding procedure was overall excellent: Negative, ICC = 1, 95% CI [1, 1],  $p < .001$ ; Positive, ICC = .99, 95% CI [.99, 1],  $p < .001$ ; Psychological, ICC = .99, 95% CI [.99 - .99],  $p < .001$ .



*Figure 3.1. Schematic representation of the Persuasion task adapted from Peterson et al., (2018). Examples are offered in English translation from Italian*

### 3.2.4. Statistical Analysis

As a preliminary step, after computing Pearson's correlations between all variables, we checked the developmental effects on all linguistic and cognitive covariates via linear models (with dependent variables being Vocabulary, Working Memory, and Metaphor Comprehension and Production).

As for the main analysis, its purpose was to assess: a) the development of persuasive abilities considering the use of different persuasive strategies (Negative, Positive, and Psychological); and b) the role of metaphor skills in persuasion, also controlling for linguistic and cognitive abilities. In line with the literature about

pragmatic development (e.g., Köder & Falkum, 2020), developmental stages were studied by considering three age groups (i.e., 4-, 5-, and 6-year children): this strategy allowed us to straightforwardly capture non-monotonic effects across ages, without imposing any a priori assumptions about the functional form (e.g., linear, quadratic, cubic, etc.) of an age gradient (Stone et al., 2010). To meet our goals, and in light of our research questions and hypotheses, we fitted a Generalized Linear Mixed-effect Model (GLMM) testing for the difference between Age Groups (effect coded in a backward sequential way: 5 vs. 4; 6 vs. 5) in interaction with the Type of persuasive strategy (effect coded in a backward sequential way: positive vs. negative; psychological vs. positive) and scaled individual differences in linguistic and cognitive variables (Vocabulary, Working Memory, and Metaphor Comprehension and Production). We fixed model parameters to follow a Poisson distribution, with a log link function. For considering the variability at the individual (Subject) and Item level, we included the relevant random structures to account for participants' and material's grouping factors. Moreover, we added the term accounting for repeated assessment within items (Trial) in the random part. The model's formula was: Persuasion scores  $\sim$  Age group \* Type \* (Vocabulary + Working Memory + Metaphor Comprehension + Metaphor Production) + (1|Subject) + (1|Items) + (1|Trial). We also measured the proportion of Metaphors in children's persuasive statements via a separate linear regression model, which formula was: Metaphor count  $\sim$  Age group.

After fitting each mixed-effect model, we checked model assumptions with the diagnostic inspection tools included in the DHARMA package (v. 0.4.7; Hartig, 2024). In both models, the diagnostics were satisfactory (see also the additional materials provided in the OSF repository). Differences between factors, simple effects (i.e., slopes) of individual measures, and differences between slopes (i.e., interactions) are reported as Odd Ratios (OR; with their 95% Confidence Interval). The significance of fixed effects (i.e., the associated  $p$ -values) in each generalized mixed model was calculated with the approximation based approach (for further details see Kuznetsova, Brockhoff, & Christensen, 2017). To detail differences between levels of the categorical predictors in a pairwise fashion, we conducted post-hoc analysis on age-related and type-related effects, and the statistical significance of differences was Tukey-adjusted. All statistical analyses were performed in R (v. 4.2.3; R Core Team, 2023), with the R Studio editor (v. 2023.09.1+494), using the *lme4* (v1.1-26; Bates et al., 2015), the *lmerTest* (v. 3.1-3; Kuznetsova, 2017), and the *emmeans* (v. 1.10.6-090001; Lenth, 2024) packages.

### 3.3. Results

Children’s performance in each variable, across Age Groups, is shown in Table 3.1.

Table 3.1. Descriptive Statistics of children’s performance in each age group.

Measures	Range of possible scores	Whole sample N = 167	4 years N = 52	5 years N = 76	6 years N = 39
Vocabulary	0-175	69.4 (24.8)	51.6 (17.7)	70.0 (21.5)	91.7 (19.9)
Working Memory	0-8	2.06 (1.75)	1.06 (1.49)	2.04 (1.69)	3.44 (1.19)
Metaphor Comprehension	0-10	4.38 (1.56)	4.33 (1.42)	4.04 (1.60)	5.15 (1.50)
Metaphor Production	0-8	2.23 (1.97)	1.02 (1.18)	2.07 (1.59)	4.15 (2.05)
Persuasion Total score	0-∞	4.08 (2.63)	2.64 (2.23)	4.33 (2.66)	5.46 (2.21)
Persuasion Positive score	0-∞	1.57 (1.70)	0.72 (1.07)	1.80 (1.60)	2.26 (2.12)
Persuasion Negative score	0-∞	1.10 (1.58)	0.66 (1.20)	1.21 (1.61)	1.44 (1.87)
Persuasion Psychological score	0-∞	1.40 (1.64)	1.25 (1.48)	1.32 (1.59)	1.74 (1.89)
Metaphors count	0-∞	0.14 (0.49)	0.08 (0.27)	0.09 (0.29)	0.33 (0.87)

Note: Cells report the average value along with the standard deviation (in parentheses) for each measure for the whole sample of 167 children and separated for each age group

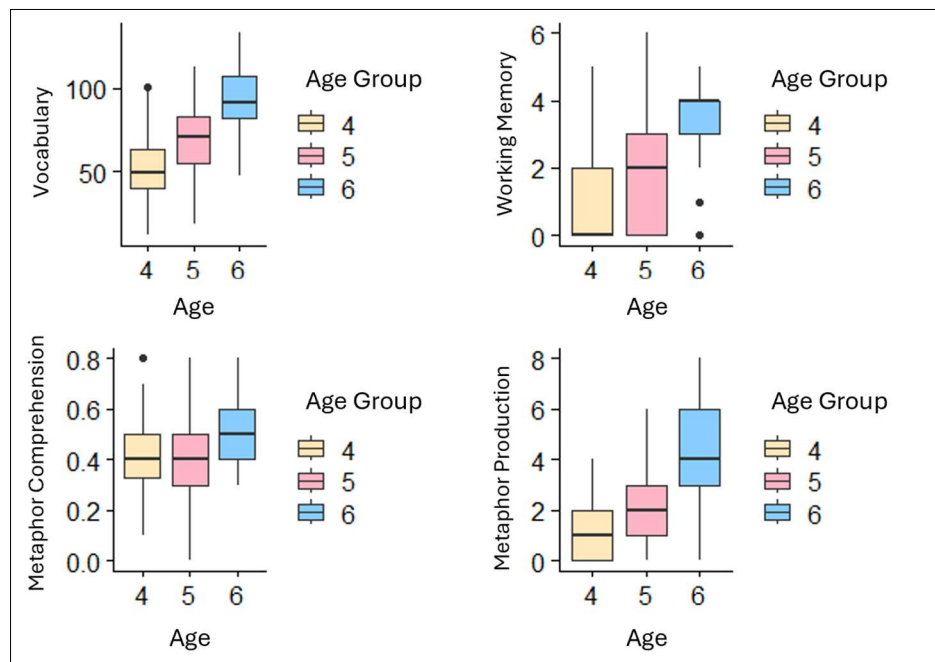
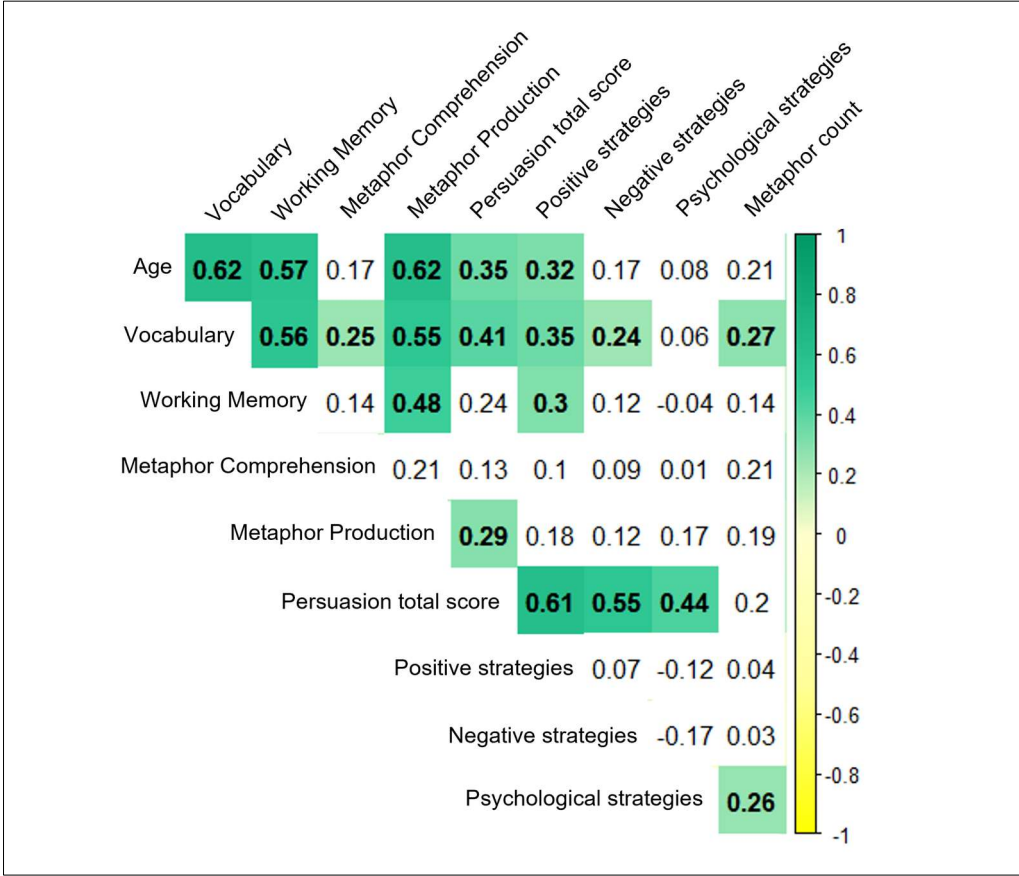


Figure 3.2. Box plot of the linguistic, cognitive, and socio-cognitive measures for each age group. The central mark indicates the median, the bottom edge indicates the 25th percentile and the top edge indicates the 75th percentile of data. The whiskers indicate 1.5 times the interquartile range.

Pearson’s correlation coefficients are plotted in the correlogram in Figure 3.3.

The Total Persuasion score was positively associated with age, as well as with Vocabulary and Metaphor Production. Vocabulary further correlated positively with Positive and Negative persuasion strategies while Working Memory correlated with Positive strategies only. Moreover, Metaphor count was positively correlated with Vocabulary and Psychological persuasion strategies.



**Figure 3.3. Correlogram between Persuasion scores and linguistic, and pragmatic variables.** The plot shows correlations between Persuasion scores and Vocabulary, Working Memory and Metaphor Comprehension and Production and, Metaphors count. Positive correlations are displayed in green and negative correlations in yellow. The color intensity is proportional to the magnitude of correlation coefficients. White cells indicate associations at  $p$ -value  $> .05$ . Age was transformed in months.

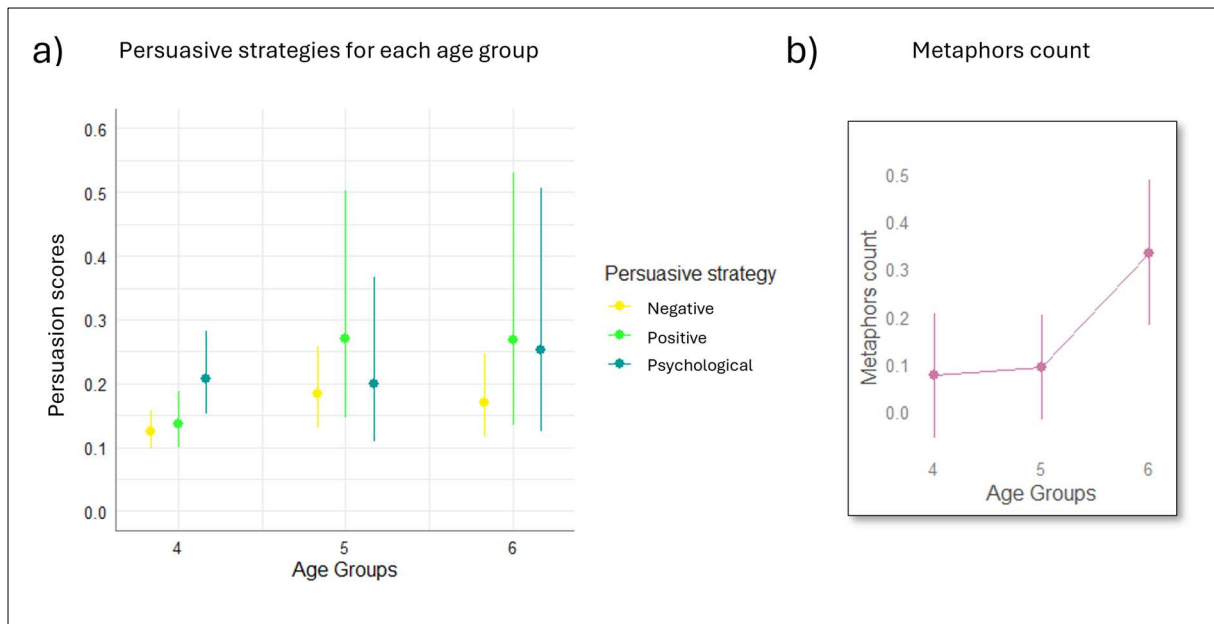
### 3.3.1. Developmental effects of linguistic and cognitive covariates

The model on Vocabulary showed a significant effect of Age Group for both the 5 vs. 4 and the 6 vs. 5 comparisons (5 vs. 4:  $\beta = 16.67$ , CI = [9.58, 23.76],  $t = 4.64$ ,  $p < .001$ ; 6 vs. 5:  $\beta = 22.83$ , CI = [14.82, 30.85],  $t = 5.62$ ,  $p < .001$ ), showing an increase in verbal abilities both between the ages of 4 and 5 and between the ages of 5 and 6. The model on Working Memory showed a significant effect of Age Group for both the 5 vs. 4 and the 6 vs. 5 comparisons (5 vs. 4:  $\beta = 0.90$ , CI = [0.37,

1.43],  $t = 3.37, p = .001$ ; 6 vs. 5:  $\beta = 1.45, CI = [0.85, 2.05], t = 4.79, p < .001$ ), showing an increase in working memory skills both between the ages of 4 and 5 and between the ages of 5 and 6. The model on Metaphor Comprehension showed a significant effect of Age Group for the 6 vs. 5 comparison (6 vs. 5:  $\beta = 0.11, CI = [0.05, 0.17], t = 3.76, p < .001$ ), showing an increase in metaphor comprehension skills only between 5 and 6 years of age. The model on Metaphor Production showed a significant effect of Age Group for both the 5 vs. 4 and the 6 vs. 5 comparisons (5 vs. 4:  $\beta = 1.05, CI = [0.49, 1.60], t = 3.72, p < .001$ ; 6 vs. 5:  $\beta = 2.09, CI = [1.48, 2.70], t = 6.72, p < .001$ ), showing an increase in metaphor production skills both between the ages of 4 and 5 and between the ages of 5 and 6.

### 3.3.2. Developmental trajectories of persuasive skills and association with cognitive and linguistic covariates

The model on Persuasion scores (see Table 2) showed a significant difference between 4 and 5 years in Persuasion scores (5 vs. 4:  $OR = 1.42, CI = [1.10, 1.83], z = 2.68, p = .007$ ), with 5-year-old children using more persuasive arguments than 4 years children, and a significant difference between Positive and Negative arguments, (Positive vs. Negative arguments:  $OR = 1.36, CI = [1.09, 1.69], z = 2.75, p = .006$ ), indicating that children in general used more positive than negative arguments. Moreover, a significant Age Group (5 vs. 4) by Type (Psychological vs. Positive arguments) interaction was observed ( $OR = 0.49, CI = [0.29, 0.81], z = -2.79, p = .005$ ), with an increase of positive rather than psychological arguments between 4 and 5 years of age (see Figure 3.4a). When inspecting the pairwise differences between persuasion strategies within age groups, we confirmed that 5 years old children used to a greater extent positive over negative ( $\Delta\beta = 0.39, z = 2.79, p = .010$ ) and psychological over positive strategies ( $\Delta\beta = -0.30, z = -2.23, p = .049$ ). Differently, no comparison in the other age groups reached statistical significance (4:  $\Delta\beta < 0.42, p > .096$ ; 6:  $\Delta\beta < 0.46; p > .092$ ).



**Figure 3.4. Development of persuasive strategies and focus on the use of metaphors.** Panel a) depicts model estimates of the use of different persuasive strategies across Age Groups. Panel b) displays the use of metaphors across different Age Groups. In both panels, error bars represent the standard deviation of model estimates.

Regarding the scaled continuous predictors, the model showed a main effect of Vocabulary ( $OR = 1.21$ ,  $CI = [1.06, 1.37]$ ,  $\chi = 2.88$ ,  $p = .004$ ): greater vocabulary skills were associated with the use of more persuasive arguments, regardless of the type of strategy (see Figure 3.5a).

The model highlighted also a significant Working Memory by Type (Psychological vs. Positive) interaction ( $OR = 0.69$ ,  $CI = [0.55, 0.86]$ ,  $\chi = -3.23$ ,  $p = .001$ ): the post-hoc analysis identified that higher levels of working memory were associated with a higher use of positive strategies ( $\beta = 0.10$ ,  $CI = [0.00, 0.19]$ ,  $\chi = 2.01$ ,  $p = .045$ ) and a lower use of psychological arguments ( $\beta = -0.12$ ,  $CI = [-0.22, -0.01]$ ,  $\chi = -2.24$ ,  $p = .025$ ; see Figure 3.5b).

In addition, a significant Metaphor Production by Type (Psychological vs. Positive arguments:  $OR = 1.39$ ,  $CI = [1.12, 1.73]$ ,  $\chi = 3.00$ ,  $p = .003$ ) interaction was observed: specifically, the post-hoc analysis revealed that higher abilities to produce metaphors were associated with the use of more psychological arguments compared to positive ones, due to a significant simple effect with

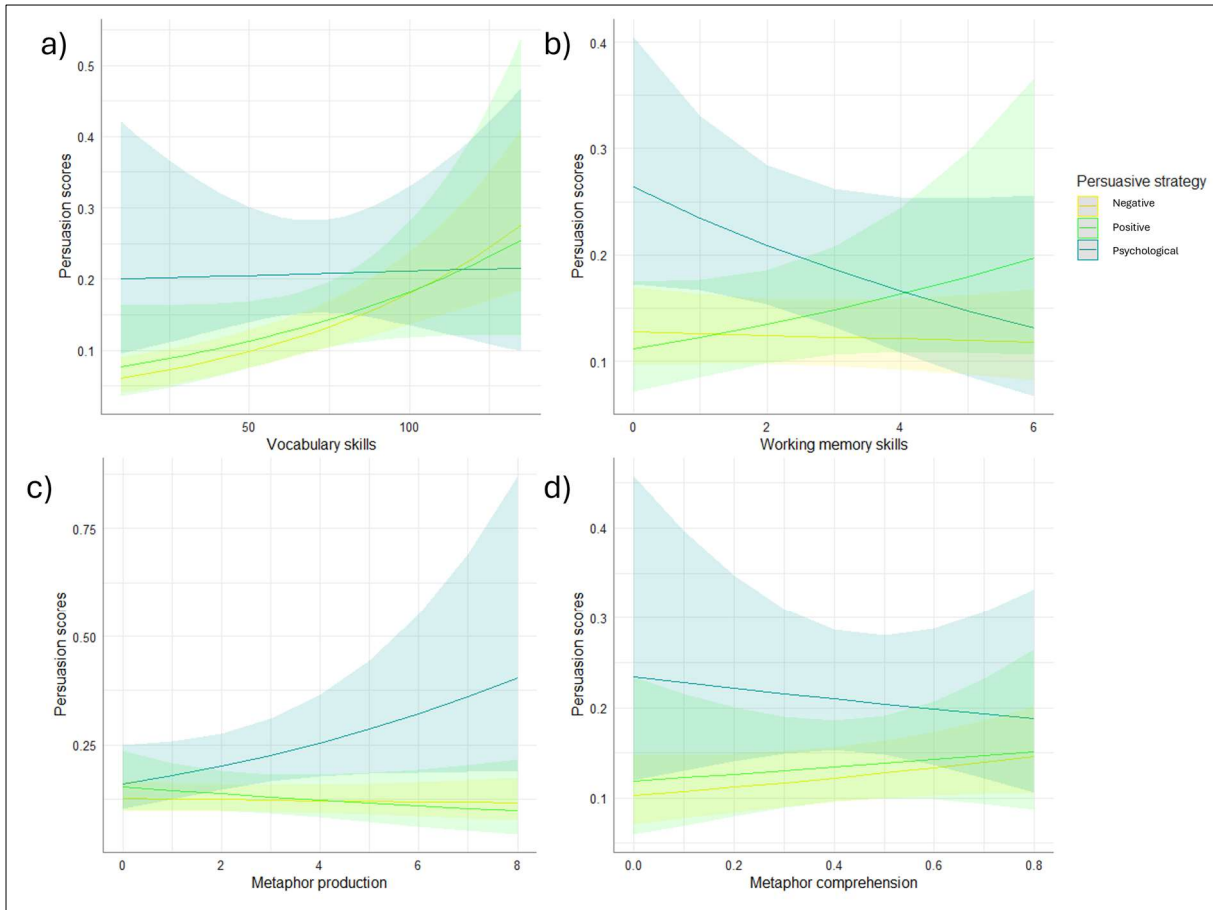
psychological strategies ( $\beta = 0.12$ ,  $CI = [0.03, 0.20]$ ,  $z = 2.57$ ,  $p = .010$ ; see Figure 3.5c). No effect of Metaphor Comprehension on Persuasion scores was found (see Figure 3.5d).<sup>5</sup>

---

<sup>5</sup> We acknowledge the debate about the treatment of the age variable. To allow drawing comparisons with the current state of the art, we kept in the main text the analysis with the age groups. For the sake of completeness, we replicated our analyses using age as a continuous variable, and, due to potential collinearity with vocabulary measures (which is highly correlated –  $r = .62$  – in our sample, we cautiously fit a model with age only. The model with continuous age replicated the effects of the original (i.e., with age groups) for all predictors. Furthermore, it coherently reported a significant effect of scaled age in predicting a greater use of persuasive arguments.

Table 3.2. Output of the Generalized Linear mixed-effects model with the Persuasion scores as the dependent variable.

Fixed effects	Incidence Rate Ratios	Response		
		CI	z-value	P value
(Intercept)	0.19	0.15 – 0.25	-13.20	< .001
Age 5vs4	1.42	1.10 – 1.83	2.68	.007
Age 6vs5	1.05	0.81 – 1.36	0.35	.728
Positive vs Negative	1.36	1.09 – 1.69	2.75	.006
Psychological vs Positive	1.04	0.85 – 1.27	0.39	.697
Vocabulary	1.21	1.06 – 1.37	2.88	.004
Working Memory	0.98	0.87 – 1.10	-0.35	.726
Metaphor Comprehension	1.02	0.93 – 1.13	0.51	.612
Metaphor Production	1.03	0.92 – 1.16	0.55	.583
Age 5vs4 × Positive vs Negative	1.35	0.76 – 2.37	1.03	.303
Age 6vs5 × Positive vs Negative	1.08	0.64 – 1.80	0.28	.783
Age 5vs4 × Psychological vs Positive	0.49	0.29 – 0.81	-2.79	.005
Age 6vs5 × Psychological vs Positive	1.27	0.77 – 2.09	0.95	.345
Positive vs Negative × Vocabulary	0.94	0.72 – 1.22	-0.48	.630
Psychological vs Positive × Vocabulary	0.80	0.62 – 1.02	-1.82	.069
Positive vs Negative × Working Memory	1.21	0.96 – 1.53	1.60	.110
Psychological vs Positive × Working Memory	0.69	0.55 – 0.86	-3.23	.001
Positive vs Negative × Metaphor Comprehension	0.98	0.81 – 1.19	-0.21	.834
Psychological vs Positive × Metaphor Comprehension	0.91	0.76 – 1.10	-0.97	.333
Positive vs Negative × Metaphor Production	0.92	0.73 – 1.16	-0.73	.466
Psychological vs Positive × Metaphor Production	1.39	1.12 – 1.73	3.00	.003
<b>Random Effects</b>				
	<b>Variance</b>	<b>SD</b>		
Intercept <sub>Subject</sub>	0.10	0.32		
Intercept <sub>Trial</sub>	0.01	0.10		
Intercept <sub>Item</sub>	0.02	0.14		
ICC <sub>SubjectTrialItem</sub>	0.07			
Model fit	Marginal	Conditional		
R <sup>2</sup>	.094	.159		



**Figure 3.5. Association between persuasive strategies and covariates.** The plot reports the relationships between persuasive strategies (Negative, Positive, Psychological) and Vocabulary (Panel a), Working Memory (Panel b), Metaphor production (Panel c), and Metaphor comprehension (Panel d).

The model on Metaphor count (See Table 3.2.) showed a significant effect of Age Group for the 6 vs. 5 comparison ( $\beta = 0.24$ ,  $CI = [0.05, 0.43]$ ,  $t = 2.53$ ,  $p = .012$ ), showing that children’s use of metaphors increased between 5 and 6 years of age (See Figure 3.4b).

*Table 3.3. Output of the Linear model with the Metaphor count as the dependent variable.*

<b>Metaphor count</b>				
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>z-value</i>	<i>p-value</i>
(Intercept)	0.17	0.09 – 0.24	4.31	<b>&lt; .001</b>
5 vs. 4	0.02	-0.15 – 0.19	0.19	.848
6 vs. 5	0.24	0.05 – 0.43	2.53	<b>.012</b>
Model Fit	Marginal		Adjusted	
R <sup>2</sup>	.046		.034	

### 3.4. Discussion

The main goal of this study was to explore the long-standing yet empirically under-investigated hypothesis that metaphor and persuasion skills are intertwined, with a focus on typically developing children in early childhood. The results confirmed our hypothesis, showing an effect of expressive metaphoric skills, even controlling for general language skills, in promoting specifically the use of high-level psychological persuasive strategies, involving mitigation and modeling, in early childhood. Before discussing this issue, we will comment on the data regarding the age-related changes in persuasive skills.

Our results showed a significant increase in persuasion performance between 4 and 5 years of age. This finding fits with previously reported data on the development of persuasive skills using observational measures (Bartsch et al., 2011) and offers more stringent evidence based on a controlled experimental task. Moreover, while previous studies described a linear pattern of development from age 3 to adolescence (Slaughter, Peterson, & Moore, 2013; Peterson et al., 2018), our data showed that the age of 5 years constitutes a turning point for persuasive skills. Interestingly, in this developmental phase, children develop one for the key components of critical thinking, namely they become able to distinguish between strong and weak reasons and to revise their beliefs when they learn that the underlying reasons were invalid (Schleihauf et al., 2022), suggesting that the capacity to evaluate how well reasons speak for or against a given belief might go hand in hand with the ability to build persuasive arguments.

What drives the improvement in persuasion performance seems to be in particular the use of positive persuasive strategies, which are cornerstones of persuasive arguments even at six years of age. In line with previous studies (Slaughter, Peterson, & Moore, 2013), we observed that children preferred positive strategies (i.e., proposing a reward) over negative ones (i.e., implying penalties to avoid noncompliance), regardless of age. We speculate that this type of preference may depend on parenting style, which was shown to play a strong role in children's communicative skills. At present, the authoritative style (i.e., controlling and demanding style, encouraging verbal give and

take, and sharing with the child the reasoning behind parents' policy, see Lavrič & Naterer, 2020) is the most widely used in Western societies (Yaffe, 2023). The use of more positive strategies in our sample might reflect the modern transition from the authoritarian style (i.e., relying on control and sanctions) to the authoritative one, encouraging communicative exchanges that are based on reciprocity, positively connoted, and less based on punishment (Estlein, 2021; Wilson et al., 2012). For what concerns psychological strategies, we do not observe a general increase in the use of these strategies in our sample. Indeed, a more varied use of persuasive strategies, including psychological ones, is observed later in development, particularly in middle childhood (Lonigro et al., 2017). Although high-level persuasive strategies do not significantly increase in early childhood, our study highlights individual differences in their use. Specifically, metaphor production skills predicted the use of psychological strategies, with children better at metaphor production also being better at using psychological tools to build persuasive arguments. At a general level, this finding is compatible with previous evidence that high-level language skills favor the development of persuasive skills in adolescents (Nippold et al., 2005). More specifically, the effect that metaphor production exerts on psychological strategies can be explained in light of its role in shaping cognitive and conceptual operations. Either via embodied simulation (as argued by cognitive linguistics theories) or via semantic associations (as suggested by the analogy account), metaphor promotes thinking processes (Holyoak & Stamenković, 2018). Such thinking activity, either propositional or not, is deemed key to understanding and mastering persuasive argumentation. In particular, a metaphor might promote a mental simulation of the concept described metaphorically (Canal et al., 2022), engaging image-based processes that enhance the persuasive capacity (Sopory & Dillard, 2002a). Also, compared to a literal expression, a metaphor increases the connections and inferences drawn for a given concept (Zhu & Gopnik, 2023), generating also a greater number of thoughts connected with the message advocacy and ultimately amplifying the persuasive power (Sopory & Dillard, 2002a). In this light, higher expressive metaphorical skills might provide children with a sophisticated toolkit for a deeper conceptualization of the topic and the

implementation of effective argumentation. While our data do not allow us to discriminate between embodied and semantic mechanisms of metaphorical thinking, it is possible that the two dimensions coexist and further strengthen the link between expressive metaphor skills and persuasion. For instance, children might transfer a rich set of inferences about a concept (Zhu & Gopnik, 2023) to shape informative persuasive arguments (Mazzarella & Vaccargiu, 2024; Rossi & Macagno, 2021) and exploit for the same purpose also the images and sensorimotor experience evoked by the metaphor.

Another possible link between metaphor expressive skills and persuasive skills, partly connected to the cognitive elaboration views proposed above, deals with the emotional evaluation of concepts. In promoting a deeper conceptualization, metaphor also leads to a greater evaluation of the concepts, in terms of valence associated to its attributes (Fainsilber & Ortony, 1987). The greater number of valenced thoughts might, when they are in the appropriate direction, in turn, promote persuasion. Recently, a training study on both metaphors and climate change, revealed that children not only improved in their knowledge on the topic, but the training also modulated their psychological and emotional attitudes (Hope and Despair), thus enhancing their pro-environmental behavior compared to their peers who undergo the same training on climate change without the use of metaphors (Pompei et al., 2024). This evidence suggests a deep connection between metaphors and emotional, psychological, and behavioral transitions, which are core dimensions of persuasion implementation (Sbisà, 2013).

Finally, the interplay between metaphor expressive skills and persuasion might be related to the variety of social functions served by metaphors, even in children. Expressive metaphorical skills allow children to negotiate meaning within the communicative exchange, as it occurs during learning when pupils dynamically modify their metaphorical productions during collaborative meaning-making (Deignan & Semino, 2022). Moreover, metaphorical communication is capable of increasing the level of intimacy between interlocutors, as Bowes and Katz showed (2015), and these positive effects on social relationships are manifest already in children (Del Sette et al., 2021).

As soon as children start to build their expressive metaphor competence, they may draw on these abilities, ensuring reciprocity in communication, greater intimacy with the interlocutor, and the creation of a shared communicative background.

The analysis of the use of metaphorical expressions during children's persuasive attempts makes the link between expressive metaphorical skills and persuasion skills even more striking. Older children were able to actively use metaphorical expressions to implement sophisticated psychological persuasive attempts. Specifically, older children used expressions such as *Broccoli is candy*, or *Brushing your teeth makes you shiny*, in 2.7% of the total persuasive attempts. Whereas for positive statements, such as *Broccoli makes you grow up*, the child must retrieve their semantic knowledge about the topic (i.e., broccoli's well-known positive effects on growth) through close associations (i.e., *broccoli-growth*), in the case of metaphorical statements, the child needs to exploit more distant associations (i.e., *broccoli-candy*) (Kintsch, 2000; Wojcik & Kandhadai, 2020). In this view, the paradigmatic association *broccoli-candy* may modify the psychological attitude of the persuadee, bridging the semantic distance between the two terms by capitalizing on the common experience of the sweetness of candies. Moreover, following embodied account (Gibbs, 2006), the use of a metaphor may activate multimodal semantic processing, exploiting the imagery involved in metaphor understanding (Canal et al., 2022), making for instance the persuadee experience the positively valenced sweetness of a candy. This may directly increase the sensorial experience and the valence related to the argument, thus boosting the persuasive effect via cognitive and affective processes (Dillard & Seo, 2012; Seo et al., 2013). Also, the use of metaphors may increase the level of connection between the two interlocutors (Colston & Rasse, 2022), realizing higher levels of social connectedness that can enhance persuasive effects (Gass & Seiter, 2022). Through the use of metaphor, hence, children generate a cognitive multiplication of interpretive spaces (Baldi, 2020), maximizing persuasive effects by activating cognitive, affective, and social channels for access to the interlocutor's representation of the topic.

While metaphor production skills were key, we didn't observe a significant effect of receptive metaphorical skills on high-level persuasion strategies or any other type. A possible explanation for this result might have to do with the slow maturation of metaphor comprehension skills, at odds with expressive abilities. While there is evidence that children are good producers of metaphors very early on (Gardner et al., 1975; Vosniadou, 1987), for comprehension early children are able to perform certain metaphor tasks (Pouscoulous & Tomasello, 2020), but their ability to fully articulate metaphorical meaning is still fragile till late childhood (Del Sette et al., 2020; Lecce et al., 2019; Winner et al., 1976). According to this perspective, metaphor comprehension and production abilities may follow diverse developmental trajectories, thus enriching communicative competence in different moments and to different extents. We argue that receptive metaphor skills might equip the child with additional social competencies for persuasive purposes at later stages, possibly in middle childhood, in parallel with the flourishing of sophisticated mind-reading skills and the effect of metaphor on fostering peer relationships (Del Sette et al., 2021). Besides metaphor skills, our data confirmed the role of general linguistic skills in supporting persuasion abilities across strategies, in line with Nippold et al. (2005). However, by looking at the correlations, positive and negative strategies were positively associated with vocabulary skills (as suggested also by the inspection of Figure 3.4), while psychological ones were not. This suggests that, while children might rely on their linguistic repertoire to produce positive and negative strategies, this is not sufficient to achieve complex psychological arguments. Working memory skills, conversely, hinder the production of psychological strategies. Previous studies showed that individuals with high working memory tend to persist in using complex, attention-demanding approaches, even when those are suboptimal for task demand (DeCaro et al., 2016). Specifically, generating psychological persuasive strategies involves high-level cognitive and communicative tasks, which cannot be supported by executive functions alone, and capitalizing excessively on working memory in a pragmatic task might result in a drop in performance.

While this study provides information on the unexplored link between metaphors and persuasion in development, it has some limitations. Firstly, we did not account for possible mediating effects of ToM, which is strongly involved in persuasive skills in childhood (Lonigro et al., 2017; Peterson et al., 2018) and has a crucial role in the development of pragmatic skills more broadly (Del Sette et al., 2020; Lecce et al., 2019; Petit et al., 2024; Tonini et al., 2023). Secondly, our task consisted of only two items, which besides being limited in number, involved the conversation with a peer (while children may use different strategies when persuading adults Lonigro et al., 2017) and were based on requests that did not involve an actual desire of the child. Future studies may include ToM measures and map socio-communicative abilities in a wider range of tasks, also modulating variables such as the age of the interlocutor and the child's motivation within the persuasive context. Furthermore, although we believe that the metaphor production effects on psychological persuasion are genuine, being observed for psychological strategies only, we cannot rule out that this association is due to the compatibility in modality between the two tasks, in both cases expressive. Further studies involving other non-metaphorical productive tasks would help assess whether modality is the only factor that counts or there is a true association between the two abilities at stake.

The link between persuasion and metaphor operates on multiple levels: on the pragmatic level, metaphor constitutes a tool that enables sharing common ground in terms of meanings and experiences; on the cognitive level, metaphor strengthens the ability to find persuasive arguments; in addition, metaphor operates on social connections, enabling increased levels of intimacy between conversational partners. Both phenomena operate on these three levels; at the same time, their connection cannot but hold a multifaceted nature.

Albeit not conclusive, our results offer first insights into the relevance of metaphorical skills, particularly expressive ones, during early childhood, for communicative and social purposes. Children develop their persuasive skills early on, to begin actively signing their *contract* with the world. At the age of 5, children begin to formulate their first complex requests and do so using

strategies based on reward or focusing on the positive outcomes of their proposals, preferring them over strategies based on punishment or threats of negative outcomes. Only a few children, however, attempt a psychological approach to the interlocutor during early childhood, creating a channel for changing attitudes, not just behavior, toward the topic. Children who adopt these high-level strategies are those who start enriching their communicative repertoire, pivoting around metaphors as powerful tools for thinking and navigating the social world.

## STUDY FOUR

### METAPHORS AS THE EARTH'S PARACHUTE: THE EFFICACY OF THE METACOM-GREEN

#### PROGRAM IN FOSTERING ENVIRONMENTAL AWARENESS IN CHILDREN<sup>6</sup>

##### Abstract

Communicating the environmental crisis is one of the main challenges of our time. Media, scholars, and educational services frequently convey Climate Change-related information through metaphor, an effective tool for making complex issues available to the public and shaping opinions during social crises. However, little evidence exists on the effectiveness of metaphors in communicating environmental issues, particularly to children. A critical matter is represented by the not-yet-fully mature children's metaphor comprehension abilities, which might in turn limit their engagement in the debate and reduce their opportunities for active participation. To address these issues, we built upon the proven-effective MetaCom training to develop a novel intervention program – the MetaCom-Green – designed to enhance both metaphor comprehension and Climate Change awareness. Each training session was built around one metaphor targeting an environmental issue (e.g., *The ocean is a soup* or *Animals are eating bites of fire*). Sixty-three fourth graders were semi-randomly assigned to the MetaCom-Green or a control group, with the latter receiving an equivalent but metaphor-free environmental awareness program. The MetaCom-Green confirmed its efficacy in promoting metaphor understanding. Furthermore, while both groups improved in Climate Change-related knowledge, only the MetaCom-Green group consistently changed their attitudes (with higher Hope and lower Despair for the future, and greater pro-environmental behavior). These findings suggest that strengthening children's metaphorical comprehension skills not only enhances the understanding of complex topics but also positively influences psychological

---

<sup>6</sup> This chapter is a manuscript in preparation for submission to a peer-review journal as “Pompei, C., Mangiaterra, V., Frau, F., Luciani, F., Bischetti, L., Lecce, S., & Bambini, V., *Metaphors as the earth's parachute: the efficacy of the MetaCom-Green program in fostering environmental awareness in children*”.

substrates, increasing sense of agency, decreasing resignation, and promoting action. The MetaCom-Green training offers a promising approach to including children in environmental discussion and green policies, also supporting their language development and social functioning.

#### 4.1. Introduction

*I don't want you to be hopeful. I want you to panic...I want you to act as if our house is on fire. Because it is.* In this famous statement uttered during her speech at the World Economic Forum, Greta Thunberg explicitly uses a metaphor to frame global warming as an urgent crisis, aiming to elicit immediate action (i.e., *I want you to act*) by modulating emotional and psychological responses (i.e., *I want you to panic*). The effectiveness of metaphors for guiding thoughts and shaping knowledge and attitudes toward a topic is largely attested in the literature (Thibodeau & Boroditsky, 2013). First, metaphors were shown to be effective in making complex issues, such as social or scientific debates, more accessible to the general public (Deignan & Semino, 2019). Moreover, metaphor frames were reported to influence attitudes and opinions in uncertain situations (Scherer et al., 2015; Thibodeau, Hendricks, et al., 2017). Consistently, in the last decades, metaphors stood aside the greatest challenges of our society: justifying the war, informing about spreading diseases, driving specific opinions towards migration flows, or guiding action during the pandemics (Cisneros, 2008; Hanne & Hawken, 2007; Lakoff, 2012; Semino, 2021). Similarly, climate change communication increasingly adopted figurative language to deliver climate change-related concepts to the citizens (Shaw & Nerlich, 2015). However, the use of metaphors is not free of challenges. While presenting several advantages for communication purposes, they are reported to be costly in terms of processing resources, both at the behavioral and neural level (Bambini et al., 2013, 2024). This should be taken into account when using metaphors to deliver information to populations with fragile pragmatics skills. Educational materials, for instance, largely extend and explore metaphors for environmental communication (Deignan et al., 2019). Children, however, do not possess fully mature abilities to understand figurative meanings (Lecce et al., 2019), which might have a negative impact on the possibility of observing the effectiveness of these expressions. This study aims to test whether training metaphor competence in middle childhood can overcome these challenging aspects to fully exploit the advantages of using metaphors during social crises, in particular for climate change communication purposes.

The literature on climate change communication highlighted that metaphors are effective for environmental purposes both to shape concepts and to guide behavior. For instance, some studies explored how metaphorical language scaffolds knowledge on climate change (Niebert & Gropengiesser, 2013), reinforcing the idea that metaphors enhance accessibility (Deignan & Semino, 2019). Others examined the role of *green metaphors* in triggering urgency, via frames such as *war* or *enemy* metaphors, which can amplify the perceived severity of the crisis (Flusberg et al., 2017). However, these expressions showed different effects on climate change perception, which varied among experts and non-experts (Deignan et al., 2019; Flusberg & Thibodeau, 2023; Thibodeau, Frantz, et al., 2017). Moreover, experts often warned about the risks of some scientific metaphors being too simplistic, misleading, or generating unexpected outcomes depending on several aspects (e.g., for the *Genes as a blueprint* metaphor, an uncontrolled promotion of deterministic beliefs, see Condit & Condit, 2001; Nelkin, 2001). Consistently, some green metaphors turned out to be thorny, especially for the young population (Deignan & Semino, 2022), which was the most affected by the adverse outcomes of these metaphors. For instance, a metaphor such as *The ticking clock*, (Zurru, 2024) was shown to spawned the so-called *climate doomism*, namely a high sense of concern paired with a low sense of agency (Johnstone & Stickles, 2024), rather than promoting action. Also, a consistent strand of research suggested that efforts to build excessive concern can spill over into denial and paralysis in younger citizens (Chamila Roshani Perera & Rathnasiri Hewege, 2013).

The literature on metaphor development and educational psychology provides insights to understand the side effects of using green metaphors. To start with, studies on metaphors in education suggest that when using metaphors such as *Pollution is a blanket*, if the knowledge of the topic is too scarce (e.g., *pollution*), the knowledge of the vehicle (e.g., *blanket*) cannot supply the missing information necessary to build the conceptualization of the phenomenon (Cameron, 2003; Littlemore, 2016). Moreover, in young populations, metaphor understanding may be challenging, as these expressions might load on not-yet-mature pragmatic inferential abilities, with children still

struggling to derive especially more abstract meanings (Grigoroglou & Papafragou, 2019; Lecce et al., 2019). As a result, the most influential population for climate change resolution (i.e., children) risk not having enough basic knowledge, nor inferential abilities to make sense of the phenomenon through metaphors, with detrimental effects on learning. Moreover, the miscomprehension may lead to negative affection towards CC such as *doomism* and despair, which inhibits, rather than promote, action.

Recently, a conspicuous body of literature tried to develop intervention programs to modulate attitudes toward climate change in students exploiting educational techniques spanning from focus group to virtual reality (Markowitz et al., 2018; Monroe et al., 2019). However, a recent systematic review of educational programs on climate change revealed that while some positive effects were observed on knowledge, the evidence of a genuine influence of such interventions on children's pro-environmental attitudes is scattered (Rousell & Cutter-Mackenzie-Knowles, 2020). Surprisingly, even if climate change concepts are often conveyed metaphorically (e.g., *greenhouse gasses*, Morote & Hernández, 2022) to young studies, to the best of our knowledge, none of these intervention programs explicitly controlled for the presence of metaphors nor children metaphor comprehension skills. Importantly, recent evidence suggested that metaphor comprehension can be trained in middle childhood via ad hoc intervention programs (e.g., the MetaCom training, Tonini et al., 2022; Del Sette et al., 2024). This evidence opens up the possibility that training children's metaphorical abilities while introducing them to climate change may help in building environmental awareness by avoiding misleading and exploiting metaphors' cognitive and psychological power.

#### 4.1.1. The present study

The present study stemmed from the idea that, when jointly promoting metaphorical competence, it is possible to fully exploit metaphor efficacy in delivering knowledge and shaping thoughts and actions about climate change. Capitalizing on previous research on metaphor training, we implemented a green and enriched version of the MetaCom training (i.e., the *MetaCom-Green*) to

promote metaphor comprehension while fostering environmental awareness. The MetaCom is an intervention based on pragmatic theory of metaphor comprehension, which promotes metaphor skills via training attention to context and inferential reasoning about context-dependent modulations of meaning (Tonini et al., 2022). It was shown to be effective, compared to a control trainings, in improving metaphor comprehension, with also positive far transfer effects on reading comprehension as well as Theory of Mind (Del Sette et al., 2024). The MetaCom-Green inherits these foundational tenets while incorporating an informative context about climate change phenomena as well as an extended enterprise of tasks employing metaphorical thinking to modulate attitudes towards climate change.

We designed the study to test the efficacy of the MetaCom-Green in promoting not only metaphor comprehension but also environmental awareness. Hence, fourth-grader children were assessed pre- and post- training for metaphor skills as well as for a battery of reliable tools to evaluate environmental knowledge and attitudes (i.e., Hope, Despair, Ecoempathy, and Pro-active behavior), which we adapted in a computerized version. As a control group, we implemented an equivalent version (i.e., the *Green training*) consisting of the same informative passages used for the MetaCom-Green, without the use of metaphors and tasks aimed to strengthen knowledge and proactive awareness.

We hypothesize that: a) the MetaCom-Green training promotes metaphor comprehension, possibly extending to other linguistic skills and theory of Mind, replicating the results of Tonini and colleagues (2022) and Del Sette and colleagues (2024); b) both the MetaCom-Green and the control groups improve in climate change-related knowledge; c) the MetaCom-Green training additionally has specific effects in modulating psychological attitudes towards climate change and foster pro-environmental behavior compared to the control group.

## **4.2. Methods**

### **4.2.1. Participants**

To calculate the necessary sample size, following Tonini and colleagues (2022), we ran a power analysis with G\*Power. Results showed that a sample of 54 participants was necessary to reach an 80% power given a medium effect size within a factorial design ( $f = 0.25$ , or equivalently  $d = 0.50$ ). Anticipating some dropouts and exclusions, we selected four classes to exceed the minimum number by some margin. Children were recruited from four classes in Lombardy, in northern Italy. Both parents of each child signed written informed consent. Children for whom no informed consent was obtained were engaged in different activities under the supervision of one of their teachers. Ethical approval for this study was granted by the Ethical Committee of the Department of Brain and Behavioral Sciences of the University of Pavia (n. protocol: 120/2023).

A final total sample of sixty-eight four-grader children ( $M_{\text{age}} = 9;4$   $SD_{\text{age}} = 0;36$ ) participated to the study. After the pre-training assessment, children in each class were semi-randomly divided into groups, balancing cases of learning disorder and bilingualism, when possible. Each training group consisted of eight children on average. Groups were randomly assigned either to the MetaCom-Green (34 children) or the control training condition (34 children). Both training programs was meant to be accessible for children with special needs following the Guidelines for the intervention with persons with disabilities (American Psychological Association, 2022).

Participants were then excluded from the analysis if they met one or more of the following criteria: (i) diagnosis of learning disability and/or neurodevelopmental disorder, (ii) not speaking Italian from three years of age, (iii) missing either pre- or post-training assessment, and (iv) missing more than one session of the training. The final samples consisted of 32 children ( $M_{\text{age}} = 9;4$ , age range = 8;8–10;2) in the MetaCom group and 31 children ( $M_{\text{age}} = 9;4$ , age range = 8;5–10;3) in the control group. Check for participants at the ceiling in the near and far transfer measures at pre-training (> 95%) were performed but no exclusions were necessary.

#### 4.2.2. Design and procedure

The study was conducted in three main phases: a pre-training assessment (T0), a training phase (the MetaCom-Green and the control training), and a post-training assessment (T1). Two

experimenters - blind to children's training assigned condition - collected the data at T0 and T1. The assessment at T0 was conducted during two collective and one individual session. The collective sessions included paper and pencil assessment of the linguistic and socio-cognitive profile while the individual session investigated the children's pragmatic and cognitive profile and a computerized assessment (i.e., via tablet) of climate change-related knowledge, attitudes, and behavior.

The MetaCom-Green training was administered by the leading experimenter of the research team. Two research assistants semi-blind to the research questions were instructed to train children on climate change-related topics without the use of metaphors and administered the control training. The two intervention programs consisted of two sessions per week, for two consecutive weeks, plus a session of guided play for a total of five sessions of 60 minutes each. The two training programs were balanced for duration, frequency, and intensity, and were equal in their content: the four control training sessions included the same prompts and a comparable number and structure of the tasks, crucially not including metaphor items or tasks. Experimenters aimed at eliciting the participation of the entire group during the intervention sessions, providing positive feedback or corrections when necessary. The post-training assessment included two sessions, one individual and one collective. In this phase, only outcome measures were assessed. The study took 8 weeks, and the two groups were tested in parallel. Once the experimental paradigm was concluded, following the good practice of randomized clinical trials, children of the control group were involved in two sessions of the original MetaCom.

#### 4.2.3. Assessment

##### 4.2.3.1. *MacArthur Scale of Subjective Social Status*

To measure Socioeconomic status, we adapted in Italian the MacArthur Scale of Subjective Social Status (N. Adler, 2007), a measure of perceived Socioeconomic and Sociometric status, presented in a ladder format with 10 steps. The MacArthur scale showed good construct validity correlating with other measures of subjective socioeconomic status and health measures (Cundiff et al., 2013)

and Social Comparison Index in children (Amir et al., 2019), as well as substantial test-retest reliability (ICC: 75%; Giatti et al., 2012). Two versions of the ladder have distinct references to which individuals can compare themselves. The society ladder is a global measure of subjective social status and is related to the individual's place in the social hierarchy. References used in this scale are objective indicators such as income, education, and occupation. The community ladder, on the other hand, assesses how children see themselves in the ladder considering the community, namely the school, where they live in; those who are positively considered by peers stand at the top of the ladder and those who have poor peer relationships are at the bottom. Children are asked to rank their position in the society and in the school choosing one of the 10 ladders.

Possible score range: 1-10.

#### 4.2.3.2. *Vocabulary*

Receptive vocabulary skills were assessed using the Italian version of the Peabody Picture Vocabulary Test -Revised (PPVT-R; Stella et al., 2000). This test demonstrated an excellent split-half reliability value ( $r = .88$ , Dunn & Dunn, 1981) and high validity strong validity in relation to the McCarthy Scales of Children's Abilities ( $r = .79$ , Naglieri, 1981). The PPVT-R consists of 175 verbal stimuli and evaluates receptive vocabulary through a picture-selection task: children are presented with four images and asked to choose the one that best represents the meaning of the word spoken by the experimenter. Following prior research (Del Sette et al., 2024), the test was administered collectively by starting 20 items before the standard starting point for nine-year-olds (i.e., item 70). Children matched each word to the corresponding picture on a printed booklet. The final score was calculated according to the standard scoring procedure, with possible total scores ranging from 50 to 175.

#### 4.2.3.3. *Working Memory*

Working memory was assessed using the Digit Span Backward Test (WISC-R; Orsini, 1997; Wechsler, 1974). In this task, children were asked to recall sequences of two to eight numbers in reverse order. Each trial consisted of two sequences of the same length and was considered

successful (i.e., awarded 1 point) if at least one sequence was recalled correctly. Failing both sequences in a trial resulted in test termination. Previous studies have demonstrated that the Digit Span Backward Test had good reliability (Cronbach's  $\alpha = .80$ ), high test-retest reliability ( $r_s \geq .74$ ; Williams et al., 2003), and moderate construct validity, as indicated by its correlation with fluid intelligence ( $r = .64$ ; Conway & Kovacs, 2013). Possible total scores range from 0 to 7.

#### 4.2.4. Focus variables

##### 4.2.4.1. *Theory of Mind*

We selected seven mental stories from the Strange Stories task (Mazzola & Camaioni, 2002; Happé, 1994), in which children are asked to infer an actor's intentions after the presentation of a brief passage describing a social scenario, followed by an open question. This task assesses the ability to attribute mental states (e.g., desires, beliefs, or intentions), including higher-order mental states. The selected set included two stories involving double bluffs, two on misunderstanding, two on persuasion, and one on a white lie. After reading the stories, children were asked to explain a character's statement in a written format, with no time constraints. The Strange Stories task has demonstrated good psychometric properties, including acceptable internal consistency and high test-retest reliability in the middle-childhood population (Devine & Hughes, 2016). Following established scoring guidelines (White et al., 2009), responses were rated on a 3-point scale: 0 for an incorrect answer, 1 for a partially correct and implicit answer, and 2 for a fully correct and explicit answer. Possible total scores range from 0 to 14. Two authors (CP; VM) independently coded 30% of the answers, and the Intraclass Correlation Coefficient showed high inter-rater reliability at both time points (T0=94%; T1=96%).

##### 4.2.4.2. *Reading Comprehension*

Reading comprehension was assessed using the Memory and Transfer standardized Reading Comprehension battery (MT Reading Comprehension; Cornoldi & Colpo, 1998), a validated tool designed for primary school children. The task involves reading short narratives followed by comprehension questions. Performance in this battery has been shown to correlate significantly

with related skills such as prior knowledge, reading fluency, and verbal abilities (e.g. Florit et al., 2020; Lecce et al., 2021). Following Tonini and colleagues (2022), we selected the *The Ancestor's cave* story, followed by 10 multiple-choice questions. Completing the task required deep text understanding and the use of inferential reasoning skills, and we considered reading comprehension abilities as a potential far-transfer measure. Possible total scores range from 0 to 10.

#### 4.2.4.3. *Physical and Mental Metaphors task*

Metaphor comprehension was assessed using a modified version of the 12-item Physical and Mental Metaphors (PMM) task (Lecce et al., 2019; Del Sette et al., 2020), a tool for evaluating the understanding of nominal metaphors in middle childhood, validated for 9 years-old children (Del Sette et al., 2021). This task presents children with metaphors embedded in minimal contexts (e.g., Dancers are butterflies; The teacher is an icicle) and has been shown to be a sensitive measure for detecting both individual differences and training-induced improvements in metaphor understanding (Tonini et al., 2022). Accuracy in the PMM task was considered the primary outcome measure of the MetaCom training. The PMM task exhibits strong psychometric properties, including good internal consistency (composite reliability =.76), strong test-retest reliability ( $r$  =.68), and adequate construct validity (Del Sette et al., 2021). Children were asked to verbally explain the meaning of each metaphor, and responses were scored based on their accuracy in articulating the relationship between the topic and the vehicle of the metaphor. Following established guidelines (Lecce et al., 2019; Del Sette et al., 2020), responses were rated with: 0 for incorrect, literal, or "Don't know" answers; 1 for partially correct responses that referenced non-salient features; and 2 for fully correct responses that captured salient aspects of the metaphor's meaning. For example, for the metaphor *Daddy is a volcano*, an answer like *Daddy eats ketchup* would receive a score of 0, *Daddy is tall and strong* would receive 1 point, and *Daddy is very angry* would be rated as 2. To ensure reliability, two independent raters (CP, VM) coded 25% of the responses, and the Intraclass Correlation Coefficient showed high inter-rater reliability at both time points (T0=97%; T1=98%). Given the unidimensional factor structure of the PMM task (Del Sette et al.,

2021), a global accuracy score was computed by aggregating responses across all items. As a result, possible total scores ranged from 0 to 24, then rescaled from 0 to 1.

#### 4.2.4.4. *Climate Change-related Attitudes*

To the best of our knowledge, no scales to assess climate change-related attitudes in children were developed in Italian. For this reason, we adapted three existing scales - Hope, Despair, and Pro-environmental behavior – originally validated for Australian adolescents (Stevenson & Peterson, 2016) - into Italian and in an age-appropriate, computerized version. Each scale was presented by a Martian character, who explained that it arrived on Earth to understand what was happening to the planet. The Martian then presented each item as a sentence-summary of what he learned. For the Hope and Despair scales, building on previous research on valid child-friendly Likert-scale alternatives (Katsos & Bishop, 2011; Schaeken et al., 2018), children were instructed to prize the Martian with a strawberry depending on the correctness of the statement: from a very small strawberry if the Martian didn't understand the sentence to a very big one if he understood very well (See Fig. 4.1.). Responses were recorded using a slider with 21 anchor points (possible score range from 0 to 20 per item), with the strawberry dynamically increasing in size as the score increased.

##### 4.2.4.4.1. *Hope scale*

The *Hope scale* comprised 8 items assessing the dimension of Hope as a combination of agency (e.g., *I know that there are many things that I can do to help solve problems caused by climate change*) and positive thinking (e.g., *Even when some people give up, I know there will be people who will continue to try to solve problems caused by climate change*). The Italian adaptation based on our sample demonstrated good internal consistency (Cronbach's  $\alpha = .80$ ), consistent with the original scale (Cronbach's  $\alpha = .80$ ). Possible total scores ranged from 0 to 160.

##### 4.2.4.4.2. *Despair scale*

The *Despair scale* included 4 items about helplessness, hopelessness, and resignation towards climate change (e.g., *Climate change-related problems are out of my control*). The scale exhibited lower but

still acceptable internal consistency (Cronbach's  $\alpha = .6$ ), in line with the original version (Cronbach's  $\alpha = .59$ ) and with other measures of climate concern (Stevenson et al., 2014). Possible total scores ranged from 0 to 80.

4.2.4.4.3. *Pro-environmental behavior scale*

For the *Pro-environmental* scale, a slightly different procedure was used. Since this scale measured the frequency of children's pro-environmental actions, children reported how often they engaged in specific behaviors (e.g., *Turn off the lights at home when there are not in use*) by selecting from five smiley faces ranging from *sad* to *happy* and corresponding to the range from *never* (0 points) to *always* (4 points). The scale consisted of 10 items covering three dimensions: Household Behavior, Information Seeking, and Transportation. The Italian adaptation based on our sample demonstrated good reliability (Cronbach's  $\alpha = .77$ ), in line with the original version (Cronbach's  $\alpha = .74$ ). Possible total scores range from 0 to 40.

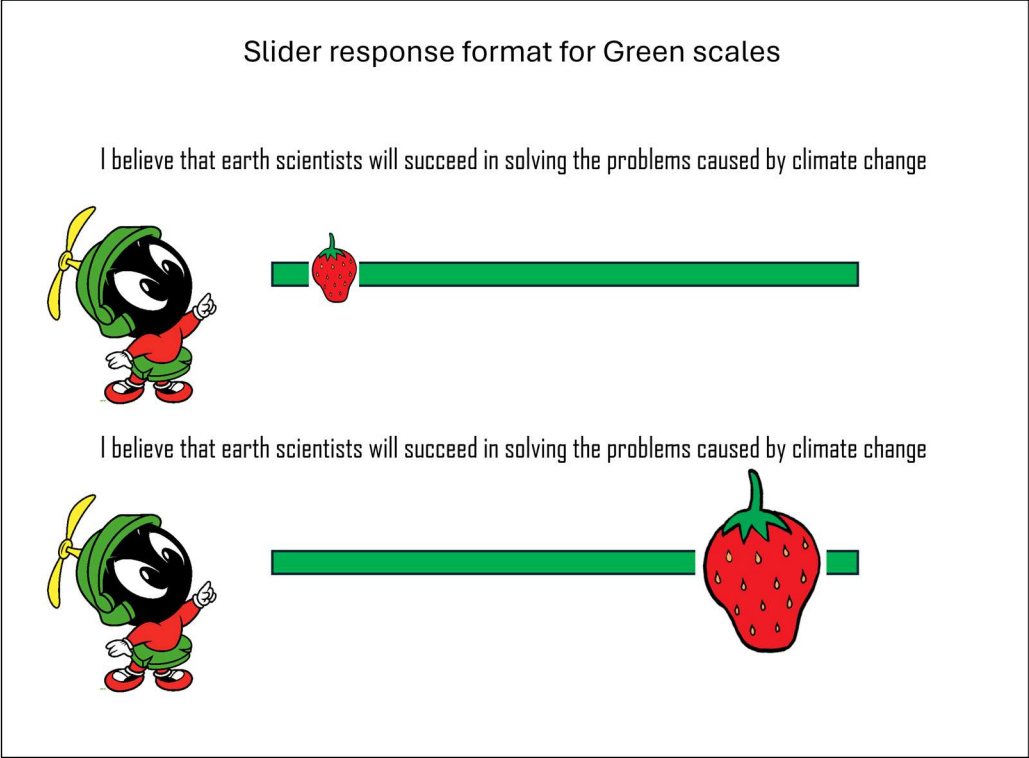


Figure 4.1. Computerized procedure for Climate Change-related attitudes assessment. The panel represents an example from the Hope scale. Children were presented with the slider and the strawberry set in the middle of the slider. Children can tap

*the strawberry and slide it in both directions: left sliding resulted in a decrease of the strawberry size, and sliding right increased the strawberry size.*

#### 4.2.4.5. *Climate Change Knowledge scale*

To verify the increase in the understanding of climate change-related concepts and phenomena, we adapted an existing survey into Italian. This survey was originally developed to assess the effectiveness of training program on CC-related knowledge among Australian primary school children (Taber & Taylor, 2009). The original scale included 25 multiple-choice questions on climate change-related knowledge with four response options (i.e., True, False, I don't know, I don't understand). For our study, we selected and adapted a subset of items based on the following criteria: the item should not be country- or region-specific (e.g., related to Australia); the item should align with the content covered in the training (e.g., excluding topics like the Ozone's Hole); the item should avoid strong emotional trigger (e.g., cancer); and the item should demonstrate a significant increase in correct responses after training in the original study. The test was implemented via tablet. The final adapted scale consisted of 15 items, with a possible total score ranging from 0 to 15.

#### 4.2.4.6. *Connection to Nature Index*

To assess the emotional disposition of children towards the environment, i.e., Ecoempathy, we used the Connection to Nature Index, a valid and reliable tool developed by Cheng & Monroe (2012) for middle childhood. The Index showed good construct validity, as it correlates with other measures contributing to the development of affective attitudes toward nature as well as good reliability (Cronbach's  $\alpha = 0.87$ ). This index consisted of 15 items covering four different factors: a) enjoyment of nature, b) empathy for creatures, c) sense of oneness, and d) sense of responsibility. Children were asked to rate how they agreed with the statements (e.g., I like to hear different sounds *in nature*) using a 5-point response format from 1 (*strongly disagree*) to 5 (*strongly agree*). The test was implemented via tablet. Total possible score range: 15-75.

#### 4.2.5. Training programs

##### 4.2.5.1. *The MetaCom-Green*

The MetaCom-Green training program is shaped after the MetaCom training (Tonini et al., 2022), with several improvements and specifics. While it maintains some features of the original version (i.e., the general structure and the rationale of the exercises), changes were implemented to boost receptive productive and metalinguistic metaphor skills while promoting climate change awareness. The MetaCom-Green was composed of four standard sessions plus a guided-play session. Each standard session was built around a metaphor about one of four phenomena related to the environmental crisis (e.g., *The ocean is a soup* for the topic of water pollution). The metaphorical items were built by consulting more than 50 entertaining and educational Italian textbooks on climate change for school-aged children and adapting three of the most commonly used nominal metaphors and a more novel one. Each session was of increasing complexity in terms of the novelty of the metaphor used (e.g., first session: *The ocean is a soup*; third session: *Animals eat bites of fire*), and of the amount of information provided in the passage and was composed of 8 tasks (Tasks 1,3,7 and 8 were novel integrations for the MetaCom-Green):

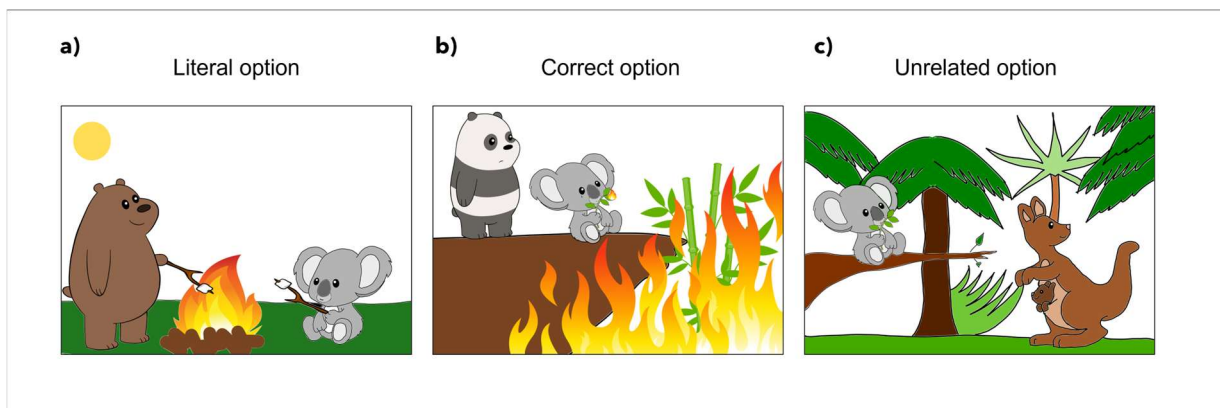
- 1) *Meaning analysis (multiple-choice)*: Once the trainer presented the text about the climate change-related phenomenon children were asked to answer a multiple-choice question about the meaning of the metaphor by choosing among a correct, an incorrect literal, and an incorrect unrelated alternative. The trainer started the discussion by identifying the encyclopedic features of the concept encoded by the metaphor's vehicle (e.g., for *bites of fire*, being very hot, not feeding, etc.) and analyzing the context to select the relevant properties for the metaphorical meaning-making;
- 2) *Meaning analysis (verbal explanation)*; After the multiple-choice task, children were asked to answer a second question about the meaning of the metaphor, in a verbal explanation modality. In this case, the trainer exploited different answers to fine-tune the metaphor's interpretation, explicitly addressing the differences between salient and non-salient

properties. This task aimed at strengthening the adjustment of the concepts to the context and to understanding the implicated meaning of the metaphor through pragmatic inference.

- 3) *Metacognitive considerations*: In this task, children were asked to answer a question about the effectiveness and utility of the metaphor in delivering the message, explaining why the metaphor was effective and useful (or not) in communicating the phenomenon compared to a literal counterpart. This task aimed at eliciting metacognitive considerations about metaphor use for communicating scientific concepts. It was inspired by the Goal-directed, Interactive Think-Aloud (GITA) technique, which allowed the students to examine the processes used to interpret and learn from metaphors (Cameron, 2003).
- 4) *Sentence-picture matching*: Children were asked to select the picture that better represented the meaning of the metaphor (e.g. *animals eat bites of fire*) from three alternatives representing three different interpretations: one illustrated the literal meaning of the vehicle (e.g. animals eating small bites of hot food, Figure 4.2(a)), one depicted the metaphorical meaning of the vehicle (e.g., a forest in flame, Figure 4.2(b)), and one was a non-related representation of the metaphor's topic (e.g., the animals in the forest, Figure 4.2(c)). This task aimed at consolidating the knowledge about the meaning of the metaphor in a multimodal fashion.
- 5) *Association task*: This task, inspired by Willinger and colleagues (2019), consisted of providing children with a set of four isolated words (e.g., bites of fire - chili pepper - reproach – anger), which included the term used as vehicle in the metaphor under consideration (e.g. *bites of fire*), and asking them to find metaphorical associations between the four words. There were both literal (e.g., reproach-anger) and metaphorical (e.g., anger-chili pepper) possible associations. This task aimed at strengthening the learned metaphorical meaning and promoting its generalization also for mental content;
- 6) *Metaphor production*: Children were asked to recall or invent another metaphor for animals suffering due to non-optimal environmental conditions. Creative metaphor production was

encouraged during the discussion. This task was meant to strengthen also the productive dimension of metaphorical skills;

- 7) *Knowledge task*: This task consisted in a reading comprehension task in which children were asked to answer both multiple-choice and open questions about the content of the text. This task aimed at strengthening the knowledge of the information about the climate change-related phenomena presented in the passage.
- 8) *Proactivity task*: The last task required children to combine the knowledge acquired and the metaphorical conceptualization for pro-environmental purposes. Specifically, children were asked to imagine an object they could turn into to solve the situation (e.g., being a hydrant to stop the fire) and explain how they could metaphorically be that object in daily life situations (e.g., close the tap while washing teeth).



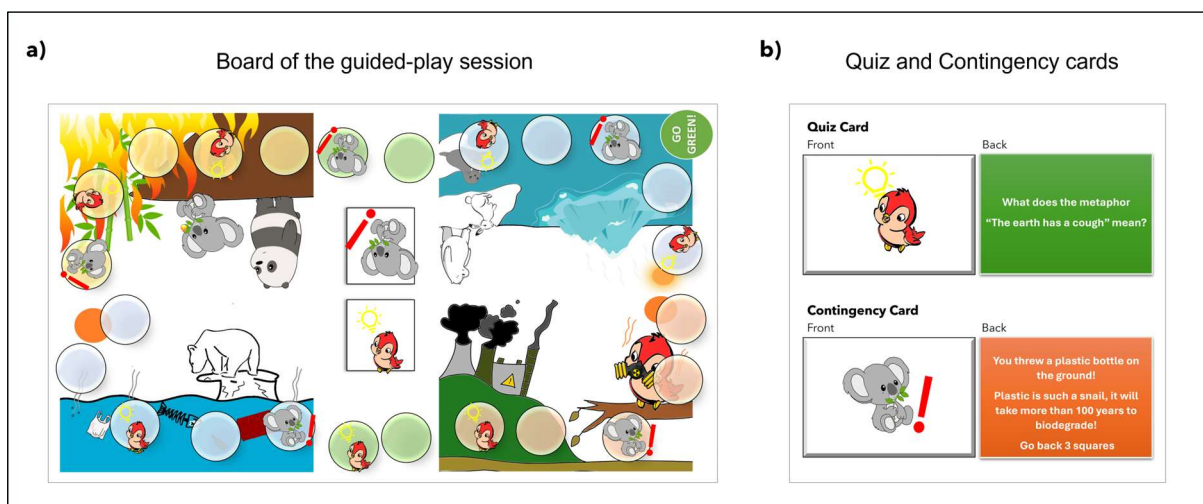
**Figure 4.2. Picture matching task.** Images used as options for the picture-matching task of the metaphor “Animals eat bites of fire”. Panel a) Literal option; Panel b) Correct option; Panel c) Unrelated option

In the MetaCom-Green a single item was presented for each session. During the first session, the experimenter presented the training introducing the concept of metaphor and the concept of climate change. Children were told that metaphors are often used to introduce complex and scientific concepts because they may help elucidate important aspects of the topic. A total of four items were presented. Sessions were held in groups but were articulated in different moments: First, the trainer read the question loudly and asked children to individually reason about the question and write down their answers in the booklet. Then, the trainer started the group discussion after

each item. The trainer gave feedback for each item both positively reinforcing correct answers and redirecting incorrect ones. As in the Metacom original version, at the end of each session, children were provided with a card related to the content of the session. In the MetaCom-Green the card contained the correct picture used in the picture matching task (see Fig. 3b) and the metaphor learned, additionally, on the other side of the card there were six pro-environmental actions children can do, always metaphorically presented (e.g., *Animals are not dolls*: avoid buying products that are tested on animals. Ask your parents how to recognize them). Cards were presented to the children as a thank-you gift and as a useful tool for the final game of the last session.

#### 4.2.5.1.1. Guided-play session

For the final session of the training, we decided to adopt a guided-play strategy following recent insights from pedagogy and educational psychology. For these purposes, we implemented the MetaGame-Green. The session consisted of a board game, with the main scope of saving the ecosystem. The board was made in fabric and measured 84,1 cm x118, 9 cm. It was structured in four parts (Ocean, Glaciers, Forest, and Atmosphere), reflecting the four topics covered during the training (See Fig. 4.3a).



**Figure 4.3. Material for the guided-play session.** Panel a) represents a stylized version of the board of the MetaGame-Green. Panel b) reports two examples of a Quiz card and a Contingency card of the MetaGame-Green

Children were divided into four teams; each team had as a first aim to become responsible for one of the elements of the ecosystem and restore it. To do so, children have to collect *clean-air* cards

that allow them to acquire both the Element (e.g., The ocean) and the items to restore it (e.g., clean water, healthy fishes, etc.). To collect *cleaned-air* cards children had to answer correctly to *quiz* cards or deal with *contingency* cards. Quiz cards were built to contain 25% of climate change-only questions (e.g., *Which are the main sources of pollution?*), 25% of metaphor-only related questions (e.g., *What does it mean the metaphor “My classmate is a fox”?*) and 50% of questions related to *green metaphors* (e.g., *What does it mean that “turtles have collars”?*). climate change-only quizzes addressed the topic covered during the four canonical sessions but did not contain questions from the Knowledge scale used in the assessment, metaphors-only quizzes resembled the structure of the training exercises (i.e., association tasks, metaphor comprehension and production tasks, etc.) but do not contain metaphors from the PMM task. Contingency cards contained both negative and positive contingencies related to anti- or pro-environmental behaviors; immediate consequences related to climate change or role-taking towards animals, always metaphorically framed. The trainer supervised the game following the principles and practice of guided play (See Weisberg et al., 2016).

#### 4.2.5.2. Control training

To match the MetaCom-Green training, we used active control training. The control training consisted of a parallel training program on climate change phenomena. The training covered the same topics as the MetaCom-Green, using the same passages without the use of metaphors. The control training included four sessions plus a guided play session. During the canonical sessions, children were presented with a passage on a CC-related phenomenon followed by 6 tasks comparable to the ones of the MetaCom-Green, aimed at controlling for each possible confounding variable, thus maintaining constant among the two trainings the levels of knowledge strengthening, multimodality, interactional and metalinguistic load and motivational aspects (i.e., two tasks of meaning analysis; a picture-matching task; a role-playing and metalinguistics task; a drawing task; a proactivity task). As for the Metacom-Green, at the end of each session, children were provided with a card related to the content of the session, containing the correct picture for the picture-matching task (see Fig. 4.2b) and the title of the passage. On the back side of the card,

there were 6 bullet points on notions about the phenomena (e.g., Excessive heat in nature is anything but a mere discomfort to animal species and leads to serious consequences such as droughts and fires). Cards were presented to the children as a thank-you gift and as a useful tool for the final game of the last session. For the control training, we did not use a board game because of the high metaphoricity of its structure, being a visual (or *act-out*) metaphor for children (Fleer, 2013) The guided play session was performed in small groups with the format of quiz game. The quizzes consisted of exercises that focused on the analysis of explicit content and that involved only low-level inferential abilities of topic on climate change (75%) and on general knowledge (history, geography, etc.; 25%). As for the MetaCom-Green, the trainer followed the practice and principles of guided play (See Weisberg et al., 2016).

#### 4.2.6. Statistical Analysis

All analyses were performed using the integrated environment for R, R studio (R Core Team, 2023). As a preliminary step, we computed Pearson's correlations among all variables. Moreover, we performed independent sample-t tests between the two groups at pre-training, to detect possible pre-training differences. To meet our first aim, namely to verify the effectiveness of metaphor comprehension, we fitted a Cumulative link mixed model (CLMM) with the *clmm()* function from the ordinal package, v. 2019.12-10 (Christensen, 2015), using the accuracy in the PMM task as the dependent variable assessing the proportional odds of the scoring (0, 1, 2). All other variables were analyzed by means of Linear Mixed Models (Bates et al., 2024), with the *lmer()* function. In all models, Time (Post-training vs. Pre-training) and Group (MetaCom-Green vs. Control) were included as sum-coded within and between factors, respectively. Scaled scores for the Digit-span Backward and Peabody Picture Vocabulary Test-Revised were added as fixed effects, in interaction with Time and Group. Random intercepts for individual-level (Subjects) and material-level (Items) variability were included in each model. Post-hoc analysis were performed using the *emmeans* package (Lenth, 2024).

### 4.3. Results

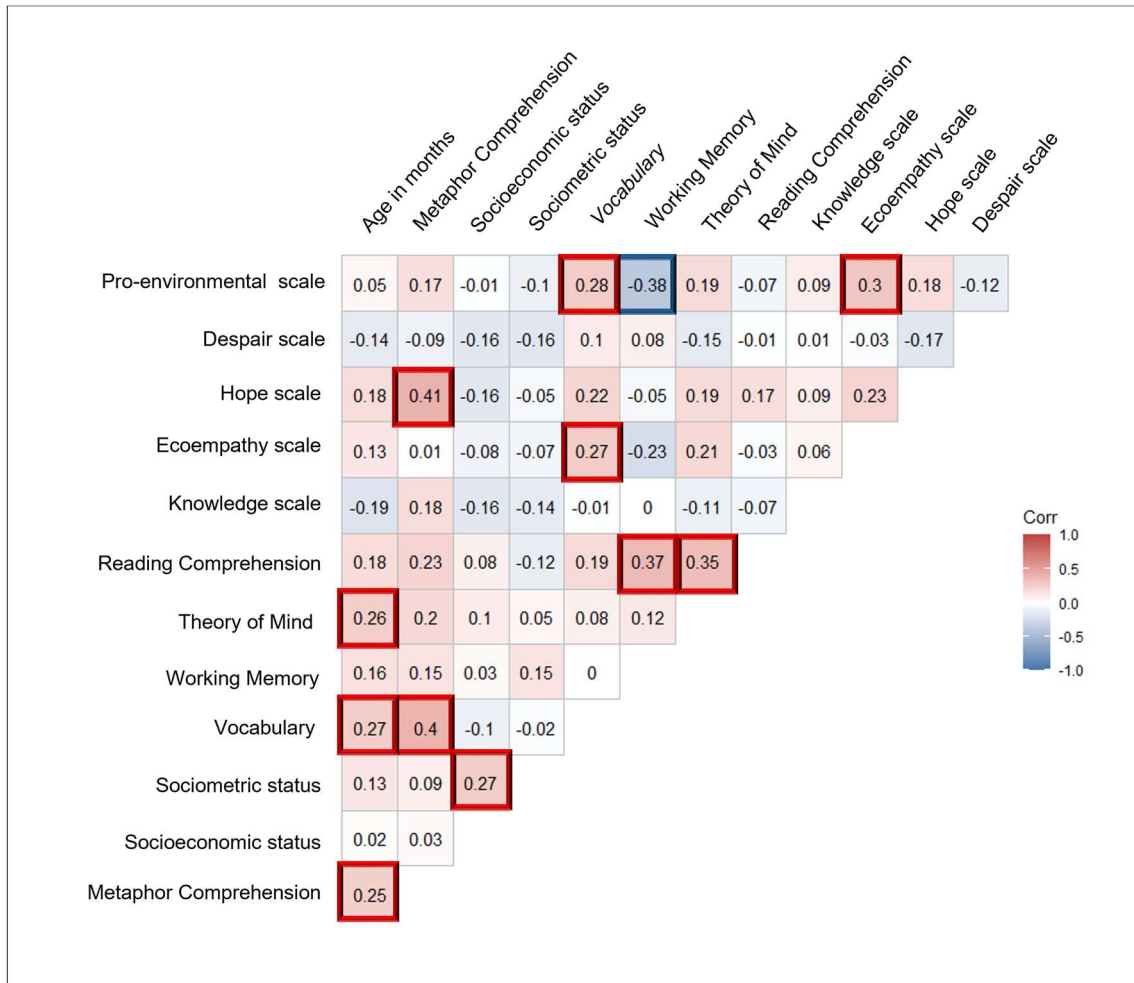
#### 4.3.1. Descriptive statistics

Descriptive statistics for each variable are reported for pre- and post-training in Table 4.1.

Table 4.1. Descriptive statistics of the two training groups at pre- and post-training on control and dependent variables

Measures	Metacom_Green		Control	
	Pre-training mean(SD)	Post-training mean(SD)	Pre-training mean(SD)	Post-training mean(SD)
Vocabulary <i>Min to max score: 50-175</i>	143 (13.8)	-	143 (12.5)	-
Working memory <i>Min to max score: 0-7</i>	3.86 (0.81)	-	4.03 (0.95)	-
Socioeconomic status <i>Min to max score: 0-10</i>	4.32 (1.09)	4.16 (1.30)	4.39 (1.20)	4.25 (1.24)
Sociometric status <i>Min to max score: 0-10</i>	3.29 (1.75)	3.53 (1.68)	3.67 (1.69)	3.78 (1.90)
Theory of mind <i>Min to max score: 0-14</i>	6.47 (2.85)	7.41 (2.72)	6.97 (2.96)	7.22 (2.78)
Reading comprehension <i>Min to max score: 0-10</i>	6.06 (1.72)	7.28 (1.71)	5.61 (2.40)	6.00 (2.54)
Metaphor comprehension <i>Min to max score: 0-24</i>	14.5 (3.62)	17.0 (3.08)	13.5 (3.10)	13.8 (2.76)
Knowledge scale <i>Min to max score: 0-10</i>	5.85 (1.76)	7.24 (1.64)	5.82 (2.28)	6.36 (1.98)
Hope scale <i>Min to max score: 0-160</i>	115 (26.4)	123 (29.2)	96.7 (33.4)	92.3 (34.1)
Despair scale <i>Min to max score: 0-80</i>	26.7 (18.4)	19.1 (14.9)	25.1 (17.0)	24.0 (15.6)
Ecoempathy scale <i>Min to max score: 15-75</i>	69.2 (4.52)	70.7 (4.94)	68.3 (6.17)	67.6 (7.37)
Pro-environmental scale <i>Min to max score: 0-40</i>	24.3 (5.69)	25.7 (6.14)	25.8 (6.02)	24.6 (6.94)

Correlations among all variables of interest are reported in Fig. 4.4.



**Figure 4.4 Correlation between outcomes and covariates at pre-training.** The plot shows correlations between linguistic, cognitive, and socio-cognitive variables and green scales. Positive correlations are displayed in red and negative correlations in blue. The color intensity is proportional to the magnitude of correlation coefficients. Highlighted cells indicate associations at  $p$ -value  $< .05$ .

Table 4.2. Independent sample *t*-tests comparing the two training groups on control and dependent variables at pre-training.

	<b>t</b>	<b>df</b>	<b>p</b>	<b>Mean difference</b>	<b>SD difference</b>	<b>Choen's d</b>
<i>Mc Arthur Socioeconomic Status</i>	0.59	51	.56	0.02	0	0.15
<i>Mc Arthur Sociometric Status</i>	1.08	61	.28	0.04	0	0.27
<i>Peabody Picture Vocabulary Test-R</i>	0.21	60	.89	0	- 0.02	0.07
<i>Digit-span Backwards</i>	0.29	57	.78	0.01	0.03	0.07
<i>Strange Stories</i>	0.34	60	.74	0.02	0.02	0.09
<i>MT Reading comprehension task</i>	-0.66	53	.51	-0.04	0.08	0.17
<i>Physical and Mental metaphors task</i>	-1.02	60	.31	-0.03	-0.02	0.26
<i>Knowledge</i>	0.30	58	.77	0.02	-0.04	0.08
<i>Hope</i>	-2.23	57	.02	-0.11	0.04	0.56
<i>Despair</i>	-0.57	61	.57	-0.03	-0.01	0.14
<i>Pro-environmental scale</i>	0.87	61	.40	0.03	0.01	0.21
<i>Ecoempathy</i>	-0.76	55	.45	-0.01	0.02	0.19

#### 4.3.2. Main Analysis

The CLMM on Physical and Mental Metaphors scores showed a main effect of Time (OR = 1.41, 95% CI [1.15, 1.73],  $z = 3.35$ ;  $p = .001$ ), Group (OR = 0.58, 95% CI [-0.36, 0.11],  $z = 1.46$ ;  $p < .001$ ), and a significant Time by Group interaction (OR = 0.57, 95% CI [0.38, 0.86],  $z = 2.71$ ;  $p = .007$ ). A deeper look into the cumulative probabilities for each Accuracy level between groups at T1 revealed that compared to the control group, the Metacom-Green group showed more fully corrected answers (i.e., 2 points;  $\Delta\text{prob} = 0.09$ , CI [0.10, 0.17],  $z = 3.22$ ,  $p = .017$ ), less partially corrected answers (1 point;  $\Delta\text{prob} = 0.11$ , CI [0.19, 0.21],  $z = 3.43$ ,  $p = .008$ ), and less wrong answers (0 point;  $\Delta\text{prob} = -0.20$ , CI [-0.32, -0.08],  $z = 4.77$ ,  $p < .001$ ).



**Figure 4.5. Effect plot of changes in Accuracy levels for the PMM task.** Each band represents the proportion of Inaccurate (lower band), Partially Correct (medium band), and Accurate (higher band) responses pre- (T0) and post- training (T1) across groups.

The model on Theory of Mind reported a significant effect of Time ( $\beta = 0.04$ ,  $SE = 0.02$ ,  $t = 2.41$ ,  $p = .018$ ), with both groups performing better at post-assessment. The effect of time was also explored within groups. The significant effect of Time seems guided by the improvement of the MetaCom-Green at T1 ( $\beta = -0.06$ , 95% CI [-0.11, -0.01],  $t = -2.25$ ,  $p = .028$ ), statistically weaker in controls ( $\beta = -0.03$ , 95% CI [-0.08, 0.24],  $t = -1.11$ ,  $p = .272$ ). The model on Reading comprehension skills showed a main effect of Time ( $\beta = 0.08$ ,  $SE = 0.02$ ,  $t = 3.70$ ,  $p < .001$ ), a main effect of Group ( $\beta = -0.09$ ,  $SE = 0.04$ ,  $t = -2.11$ ,  $p = .037$ ), as well a significant Time by Group interaction ( $\beta = 0.10$ ,  $SE = 0.04$ ,  $t = 2.36$ ,  $p = .020$ ), with a significant improvement between T0 and T1 in the MetaCom-Green group only ( $\beta = -0.13$ , 95% CI [-0.18, -0.07],  $t = -4.25$ ,  $p < .001$ ). The model on Knowledge showed a main effect of Time ( $\beta = 0.08$ ,  $SE = 0.02$ ,  $t = 2.01$ ,  $p = .046$ ) and a significant Time by Group interaction ( $\beta = 0.08$ ,  $SE = 0.04$ ,  $t = 2.01$ ,  $p = .046$ ). By exploring this effect within groups, comparisons indicate that the MetaCom-Green group only performed differently between T0 and T1 ( $\beta = 0.12$ , 95% CI [-0.18, -0.06],  $t = 4.12$ ,  $p < .001$ ).

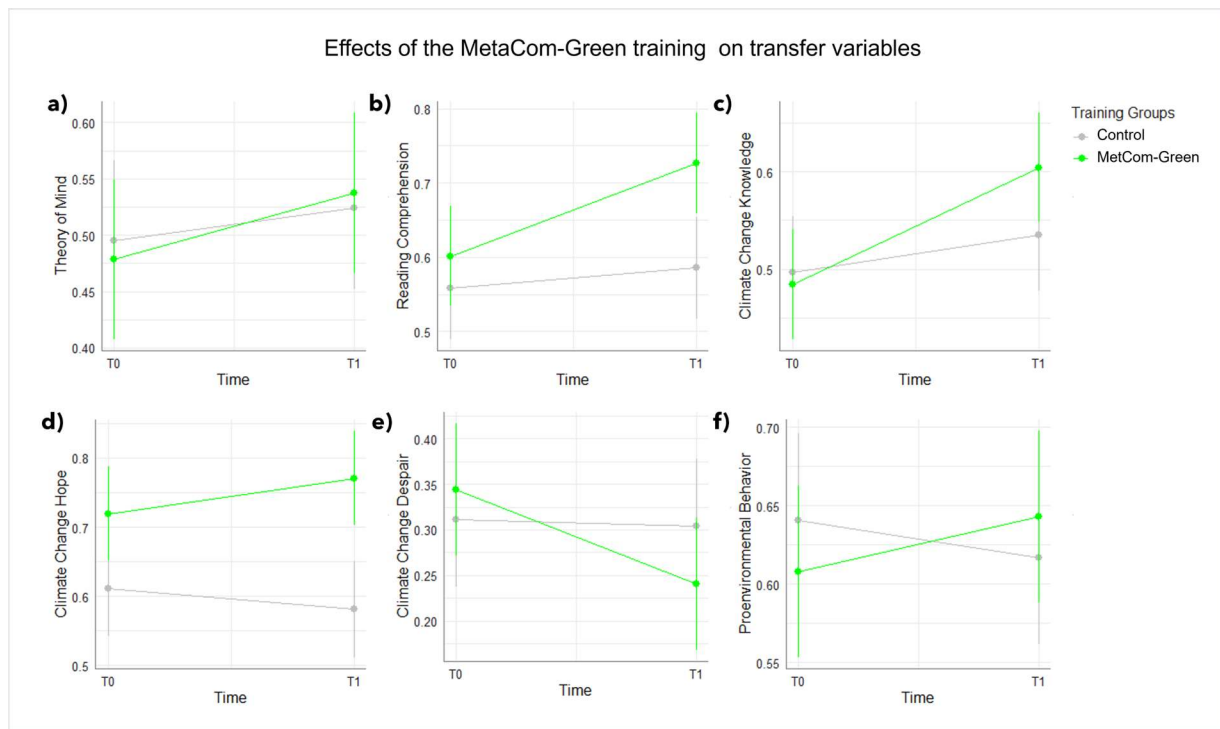
The model on Hope revealed a main effect of Group ( $\beta = 0.15$ ,  $SE = 0.05$ ,  $t = -3.23$ ,  $p = .002$ ) and a significant interaction Time by Group ( $\beta = 0.08$ ,  $SE = 0.03$ ,  $t = 2.51$ ,  $p = .013$ ). Despite the difference between groups at T0 ( $\beta = -0.12$ , 95% CI [-0.21, -0.01],  $t = -2.17$ ,  $p < .001$ ), the

significant difference at T1 ( $\beta = -0.19$ ,  $SE = 0.05$ ,  $t = -2.82$ ,  $p < .001$ ) was only guided by the significant increase between assessments for the Metacom green group only ( $\beta = 0.51$ ,  $SE = 0.22$ ,  $t = 2.54$ ,  $p = .014$ ).

The model of Despair showed a main effect of Time ( $\beta = -0.06$ ,  $SE = 0.04$ ,  $t = -2.20$ ,  $p = .030$ ) and a significant interaction Time by Group ( $\beta = -0.10$ ,  $SE = 0.04$ ,  $t = -2.20$ ,  $p = .030$ ). Post-hoc analysis confirmed that only the MetaCom-Green group performed differently at pre- and post-assessment ( $\beta = 0.10$ ,  $SE = 0.31$ ,  $t = 3.32$ ,  $p = .002$ ), showing lower levels of Despair compared to the Control group.

The model on Pro-environmental behavior showed a significant Time by Group interaction ( $\beta = 0.06$ ,  $SE = 0.26$ ,  $t = 2.26$ ,  $p = .026$ ), with the MetaCom-Green group showing higher levels of Pro-environmental behavior at T1 compared to the Control group.

No significant effects were found for the model on Ecoempathy ( $ps > .05$ ).



**Figure 4.6. Effects of the MetaCom-Green training on far transfer and green variables.** Panel a) depicts the effect of the training on Theory of Mind for the control and the experimental groups. Panel b) depicts the effect of the training on Reading Comprehension abilities for the control and the experimental groups. Panel c) depicts the effect of the training on climate change-related Knowledge for the control and the experimental groups. Panel d) depicts the effect of the training on climate change-related Hope for the

*control and the experimental groups. Panel e) depicts the effect of the training on climate change-related Despair for the control and the experimental groups. Panel f) depicts the effect of the training on Pro-environmental behavior for the control and the experimental groups.*

#### **4.4. Discussion**

This study tested the hypothesis that training metaphor skills in middle childhood may help in exploiting the full potential of metaphors to deliver climate change-related information avoiding misunderstanding and despair, while promoting sustainable behavior. We implemented the MetaCom-Green program, as an enriched version of a training program with proven efficacy, as well as an equivalent control training on climate change without the use of metaphors. We expected to replicate the results of Tonini and colleagues (2022) and Del Sette and Colleagues (2024) for both near (Metaphor Comprehension) and far (Reading Comprehension and theory of Mind) variables. As for the green variables, we hypothesized to observe an improvement of both groups in climate change-related knowledge, with the MetaCom-Green group only showing a modification of the psychological and behavioral attitudes towards climate change. The analysis confirmed most of our hypotheses: the MetaCom-Green group outperformed the control group in metaphor and reading comprehension skills, although both groups improved in Theory of Mind. Notably, while both groups showed a similar trend in Knowledge improvement, only the MetaCom-Green group showed a statistically significant increase in performance. Additionally, only the children in the MetaCom-Green group increased their Hope levels, decreased their level of Despair, and, crucially, showed more sustainable behaviors. No effects on the Eco-empathy were observed for either group.

Starting with the effects on linguistic and socio-cognitive abilities, we replicated the findings in Tonini and colleagues (2022) showing the efficacy of the MetaCom, also in its Green version, in boosting metaphor comprehension skills in middle childhood, also extending its effects to far transfer measure of reading comprehension (Tonini et al., 2022). This was indeed expected since the MetaCom-Green maintains the theoretical foundation of the MetaCom, potentiating attention to context and inferential processes. However, contrary to recent findings showing a causal effect

of metaphor training on Theory of Mind (Del Sette et al., 2024), we didn't find a specific effect of the MetaCom-Green on children's socio-cognitive skills, with both groups showing higher Theory of Mind performance at post-training. While the original MetaCom consisted of both physical and mental metaphor items, the MetaCom-green covered topics related to environmental phenomena, capitalizing more on physical than mental content. Recent evidence suggested that the relationship between Metaphors and Theory of Mind may be stronger for mental, rather than physical metaphors (Canal et al., 2022; Lecce et al., 2019; Lecce, Pompei et al., in prep). In this view, it may be the case that training in environmental metaphors, which mainly refer to physical characteristics (e.g., being warm or being polluted), doesn't generalize to mentalizing skills. Moreover, the control training also consisted of a role-playing task in which children were asked to explain the phenomenon of the training session (e.g., wildfires) to a fictive classmate who ignored the phenomenon. These factors may have led to a slight increase of both groups in Theory of Mind.

Moving to the main focus of interest, that is, the effects on the green variables, the first result to comment is the performance on the Knowledge scale, which numerically increased in both groups, but significantly improved only the MetaCom-Green group. This result is consistent with previous attempts at training climate change knowledge in middle childhood without a specific focus on the communicative tools employed. These trainings showed only slight improvement in climate change-related knowledge (Taber & Taylor, 2009) and did not account for the presence of possibly misunderstood metaphors (Cameron, 2003). Interestingly, in a recent study, authors asked about the causes and effects of climate change to a large sample of primary and secondary schooled children, observing a discrete knowledge of the phenomenon, except for questions related precisely to a metaphorically expressed concept, i.e., *Greenhouse* gases (Morote & Hernández, 2022), which studies considered a possible source of misunderstanding (Forgács & Pléh, 2022). Our findings suggest that presenting climate concepts via metaphors, while at the same time accompanying children in figurative language understanding, promotes knowledge more effectively than traditional environmental training. This evidence lends concrete support to the largely shared idea

that metaphors constitute a powerful learning mechanism starting from early childhood, especially because they promote inferences about a phenomenon, thus enriching its conceptualization (Zhu & Gopnik, 2023). Even if metaphors do not provide additional or new information compared to their literal counterparts, they constitute a way of *learning by thinking* through which new knowledge is generated without supplying additional data on the topic (Lombrozo, 2020). The MetaCom-Green training, while minimizing the risk of misunderstanding metaphors, exploits the metaphor function of conceptual co-builder and ensures knowledge to a greater extent than a literal educational program.

While knowledge constitutes the fundamental basis for awareness, being informed about climate change may lead to different levels of willingness to act (Boyes et al., 2007; Skamp et al., 2013). Perhaps the most important finding of this work is that the MetaCom-Green efficacy is not limited to knowledge but showed positive effects also on hope and sustainable behavior and decreased despair toward the environmental crisis. Insights from the literature on the clinical populations, particularly in patients affected by depressive disorders, suggest that metaphors may foster hope by opening new possibilities, bypassing skepticism, and drawing near personal experience (Burns, 2024). Along these lines, metaphors may fill a gap between the awareness of the risk and the possibility of thinking about the future. Specifically, thinking metaphorically (e.g., task 8, to imagine becoming an object to save the planet) might offer a new way of thinking about the future, fulfilling a similar function to *utopian thinking* (i.e., imagining a future society in which climate change has been addressed) which have been shown to promote hope and diminish fear (Daysh et al., 2024). Conversely, literal *to-do lists* do not seem to offer these possibilities and may not affect behavior, resulting rather overwhelming (Bird et al., 2024). Another consideration regards the strong link between metaphors and embodied experience (Littlemore, 2019). Accounts in cognitive linguistics suggest that the processing of metaphors may promote mental simulations and embodied experience of the concept (Canal et al., 2022; Gibbs, 2006), in terms of eliciting *sense-making imaginative elaboration around bodily functions* (Winnicott et al., 1989). The direct experience of nature

has a pivotal role in raising the basis for environmental awareness already in early childhood (Cameron-Faulkner et al., 2017). Consistently, at later stages environmental education studies showed that tools simulating direct experience (e.g., virtual reality) seem to constitute the most effective ways to promote engagement in sustainability (Chen & Martin, 2015; Markowitz et al., 2018; Rooney-Varga et al., 2020), despite not always improving learning (Makransky et al., 2019) and potentially generating strong negative emotions (Lavoie et al., 2021). Metaphors may be a viable simulation tool that sustains embodied experience by keeping children at a *safe distance*, which is crucial to avoid despair. The last consideration regards the more affective aspects of climate change-related attitudes. While modulating other attitudes, the MetaCom-Green didn't display any effect on eco-empathy. Affective Theory of Mind seems to have an impact on climate change education, in terms of creating and strengthening the connections between the phenomenon and the consequences on plants, animals, and people (Satchwell et al., 2025). Previous training on climate change based on affective Theory of Mind showed promising results on eco-empathy and proactive behavior (Gopinath & Kumar, 2024). Ecoempathy is a very specific socio-cognitive and affective aspect that probably might benefit from a specific training. This is in line with our result on Theory of Mind, with the MetaCom-green being not specific enough to foster Theory of Mind to a greater extent compared to the control one. Future intervention programs may disentangle cognitive and affective routes to sustainability, possibly combining the effects of these two tools. Overall, our results suggest that scaffolding children's metaphorical competence makes it possible to restore the efficacy of metaphors in communicating scientific concepts. Accompanying children toward a full comprehension of green metaphors may prevent doomism, and foster hope and sustainable behavior, gently – but effectively- including children in the management of social crises.

## CONCLUSION

The main purpose of this work was to provide novel empirical evidence on the development of metaphor skills, toward a novel developmental pragmatics account of metaphor acquisition.

The need for a theoretical and methodological turn in the study of metaphor development is corroborated by the fact that so far studies on metaphor development have returned a disagreement on the acquisition timing of this ability of almost 10 years. The heterogeneous results of decades of research in developmental pragmatics also brought to light the lack of a systematic investigation on the role of specific cognitive and socio-cognitive substrates in metaphor receptive and expressive abilities. Rather than making room for the *early* vs. *late* debate, the driving force behind this work was to open the way to the study of metaphor acquisition broadening the developmental perspective. Firmly rejecting an *on-off* view of metaphorical ability, I have approached the study of this phenomenon trying to observe it in its first manifestations, respecting its nuances and appreciating the full flowering in its applications.

In accompanying into a brief journey through the ontogenesis of metaphor, these four studies have opened a small glimpse into the possibility of redesigning the study of metaphor in a developmental key, as a pragmatic competence that will be able to settle in the literal vs. non-literal continuum as part of the communicative enterprise, only after passing through complex developmental phases. I will briefly review the main contributions offered by the empirical effort of this work, concluding with a theoretical proposal for the study of metaphors in acquisition.

### **1. The development of metaphors in early childhood: *the chrysalis years***

Taken altogether, the four studies describe a multistage development toward a full-fledged metaphorical competence. Children start to produce their first accurate metaphors at 5 years of age, while in the same period are facing a literal step in metaphor comprehension. This asymmetry is confirmed and emphasized when investigating the ability to generalize metaphorical skills to the real world. In particular, **Study 1** suggests the presence of two different stages toward metaphor comprehension: the literal and the physical ones. Between 4 and 5 years of age, children seem to

approach metaphors trying to solve them as literal sentences. While previous studies considered this moment or a deflection in the developmental trajectory or a mere preference maintained until adulthood (Falkum, 2022; Gardner et al., 1975), I propose another interpretation, merging insights from developmental psychology and Relevance Theory (Sperber & Wilson, 2008). Out of the pretend-play context, during preschool years, children start having expectations of *relevance* toward the interlocutor. In interpreting a sentence like *The teacher is an icicle*, children between 4 and 5 years of age might activate the decoding of the utterance and in turn recognize the communicative intention of the interlocutor. At that point, given the rudimental level of their cognitive, socio-cognitive, and pragmatic enterprise, are not able to activate assumptions that bring away from the propositional content of the utterance. In this view, the lack of full-fledged inferential skills may result in literal interpretations meeting children's expectations of relevance. Hence, far from being a preference that holds toward adulthood, the literal one is an ontogenetic stage reflecting that children are employing the available low-level pragmatic toolkit, which will be replaced by a more sophisticated version of the same mechanism as far as requisite cognitive, socio-cognitive and inferential skills are developed. Complementary, **Study 1** also signaled the presence of a *physical step*, during which children prefer physical interpretations over psychological ones for mental metaphors, suggesting that managing metaphors involving mental reality may require an additional load in terms of both inferential effort and socio-cognitive cost. Similarly to what we argued for the literal stage, physical interpretations are the results of the activation of the inferential chain reaching a step beyond the literal one. In interpreting *The teacher is an icicle*, children in the physical step might prefer interpretations such as *The teacher is freezing*, showing that to reach high-level inferential processes the pragmatic system needs a specific involvement of Theory of Mind.

Hence, interpreting metaphors literally (and then physically) is not a preference, but rather a necessary step, pragmatic in its nature. The relation between the literal world and its figurative denotation is further explored by **Study 2**, which analyzed how children turn literal concepts into figurative meanings proving two pieces of evidence. First, metaphor comprehension and metaphor

production develop at different time rates: at 5 years of age children have already improved in their ability to generate some metaphors, and no contribution of metaphor comprehension is observed. However, some relation is observed in terms of the exploration of the semantic network for figurative purposes. An adequate semantic network exploration for choosing appropriate metaphor vehicles seems to be guided by both Theory of Mind and rudimental metaphor comprehension skills and, during early childhood, reaches an adult-like level. We can speculate that once they climbed the literal step in physical metaphors described in **Study 1**, children learn to create associations among distant words originating a coherent network of meanings to explore in a social-oriented fashion. However, this does not mean that metaphor comprehension supports metaphor production, but rather that the receptive system may share some features with the expressive one in terms of lexical-semantic strategies. The asymmetry between comprehension and production, with production emerging earlier, is not the only case of asymmetry in language acquisition (Hendriks & Koster, 2010) and also resonates with dissociations in clinical populations, in which metaphor production seems to be preserved in spite of metaphor comprehension deficits (Despot et al., 2021). Explaining this asymmetry may be more than challenging. However, I think that this work provides some evidence that metaphor production and metaphor comprehension may constitute two separate systems of the same module, capitalizing on different skills to different extents, with the expressive system reaching a better performance slightly earlier. Consistently, in metaphor production, Theory of Mind comes into play at later stages to fine-tune the productions in a less egocentric perspective but is not determinant for the improvement observed between 4 and 5 years. Conversely, to interpret metaphors children must master connections of meanings taking into account the interlocutor's communicative intention, hence, relying on personal experience is not sufficient. Along these lines, I have experienced several spontaneous metaphor productions by young children who, even in the absence of a full comprehension of physical and mental metaphors, were able to produce metaphors, even about complex abstract concepts. I had the impression that children used metaphors to make sense of their experience, as they produced

metaphors in two main circumstances: during the presentation of task instructions or during highly emotionally connotated situations. As for the first case, I was presenting the first Theory of Mind task of the Wellman and Liu scale (Wellman & Liu, 2004b), where a character is described as loving to eat carrots. As soon as I provided this feature, the child told me, with an *eureka*-like attitude: *So he is a rabbit!* (See Appendix). A bunch of these cases occurred during task instructions (e.g., *being a broken machine* for the Digit Span Backward). The second strand of cases would require a more delicate and broad discussion I cannot go through for the purposes of this dissertation. However, I will provide an example to substantiate my speculations. During an assessment session, one of the children referred a series of domestic abuses, concluding her speech saying: *Sadness is what happens to me but happiness...happiness is running free as a dog* (See Appendix). Also in this case, the child used metaphors to make sense of her experience and, eventually, share this experience. What I argue is that the expressive system of metaphor abilities plays a role in bonding the inner and the social world, helping in generalizing the experiential knowledge acquired, and thus being more ontogenetically urgent than the receptive one. Consistently, metaphor production constitutes, in turn, a foundational skill for socio-communicative intents, as shown in **Study 3** for persuasive communication. During early childhood, those children who already master some expressive metaphorical abilities are able to employ high-level persuasive strategies, independently of the actual use of metaphors in the persuasive statement. Importantly, metaphors are not mere linguistic trappings, and their potential is not limited to the cognitive effects they have in the receptive modality. Rather, metaphor production skills provide, in turn, a cognitive-grounded pragmatic springboard for high-level communicative purposes. Being *metaphorical thinkers* opens a privileged gateway and maximizes children's effectiveness in the social world. However, during early childhood, children have a long way to go before reaching a full-fledged mastery of metaphorical abilities, particularly for what concern the receptive modality. Only in middle childhood children would possess a more high-level receptive system with an adult-like structure, albeit not fine-tuned yet. This developmental phase seems to be characterized by a more elaborate receptive system that

conserves a certain degree of plasticity, enabling children to particularly benefit from training, as it happens for other abilities in the same period (Rueda et al., 2021). **Study 4** constitutes a window onto the most complex level of metaphorical competence: its generalization in the conceptual domain. The MetaCom-Green training plays the role of a protective factor against metaphorical misunderstanding, as well as showing that an adequate metaphorical receptive system is required to benefit from metaphors' cognitive and psychological effects. Particularly, metaphors act on humans' motivational force, entering the psychological world via multimodal channels and fostering a deep conceptualization and simulated experience of concepts. In this context, supporting children in refining their metaphorical competence ensures their entry into society as individuals who can understand the complex scientific debate and actively participate into it, also having a decisive role. Capitalizing on this evidence, if early childhood constitutes a sort of *chrysalis phase*, in which children strengthen and learn to generalize their rudimentary competence, during middle childhood we can observe the first butterfly's flaps.

## **2. Milk pragmatics: toward a novel developmental account of metaphorical meanings**

In this dissertation, I have sometimes mentioned the phrasing of *metaphorical competence*. What I mean by metaphorical competence is *the owning of flexible expressive and receptive pragmatic systems, in which concepts can be broadened and narrowed to enable metaphorical meaning interpretation and generation for communicative, social, conceptual, and psychological purposes*. These systems are built upon vocabulary, which constitutes the lexical scaffolding, Executive Functions, which hold the cognitive engineering, and Theory of Mind, which drives the continuous social and mental-oriented linkage among concepts, also constituting the socio-cognitive bridge toward the interlocutor. Inferences are the high-level operations of the receptive system. Producing statements able to activate the inferential chain in the interlocutor is the high-level outcome of the expressive system.

These systems allow the navigation of the literal/non-literal continuum; however, while low-level activations of the receptive system are sufficient to grasp literal meanings, a high machinery of the same system is required for metaphors. The study of metaphorical competence concerns how

children acquire the elements of these systems and when children become ready to use these systems for socio-communicative and conceptual purposes.

During early childhood, the linguistic, cognitive, and socio-cognitive architecture involved in the high-level functioning of these systems is not yet developed, and this is self-evident in the pragmatic behavior of the child. The literal step constitutes a *relevance-guided* mechanism, which is the result of the application of a rudimentary version of the system, rather than the expression of a lifelong preference for the literal end of the continuum. Conversely, human beings are meant for metaphors, as proved by the comprehension-production asymmetry, with an earlier emergence of metaphor production as an invaluable social and psychological tool. With the fine-tuning of metaphor comprehension abilities, metaphors consolidate their leading role in the child's conceptual system. From then on, metaphorical competence will stand aside the greatest communicative challenges of our adult life.

Keeping this work as due and respectful arm's length from any evolutionary consideration, but being deeply inspired by an allegory used by Andrea Moro (2011) to describe phylogenetic processes of grammar, I conclude my dissertation with an evocative hint, which is the hypothesis of a *Milk Pragmatics*. The exploration of children's errors toward a fine-tuned metaphorical competence persuaded me that literal and physical interpretations share the same sort of milk teeth. They are dismissed and suddenly replaced by finer interpretations, eventually coexisting with them for a certain time window.

Hence, a systematic study of children's error types during figurative meaning acquisition combined with robust quantitative methods can provide the testing ground of a *Milk Pragmatics*, thus paving the way toward a novel developmental account of metaphor competence.

### **3. Future directions**

Gardner and Winner, in 1978, sealed their heartfelt doubts in a paper on the implications of metaphor research on humanistic disciplines: [...] *the adequacy of paraphrase as a test for metaphoric competence; [...] the ontological status of "early metaphor" and the kinds of evidence on which one would want to*

*make inferences about the "awareness," "intention," and "consciousness" of young subjects, the possible nature and meaning of a "literal stage" [...].*

Almost fifty years later, I trust that this work contributes, at least in some small way, to the good hope of answering these questions. However, novel queries emerged after this work: How many stages characterize the development of metaphorical competence? Are they considered as proper stages? And why are they ordered the way they are? Do precursors of metaphorical competence exist before early childhood? And what about adolescence, is it only a period of fine-tuning?

Other crucial questions concern the psychological status of metaphors, reported as being dynamic clinical tools in developmental psychotherapy, even if in the absence of any quantitative evidence. Future studies should pursue the investigation of metaphorical competence development to complete the exegetical timeline and describe when, and if, metaphors become a bridge to enter the child's symbolic life.

Gardner and Winner concluded their work with a yearning:

*If humanistically oriented scholars will add to their judicious consideration of texts an equally critical attitude toward the findings reported by psychologists of various stripes, future treatments of figurative language should be significantly enriched.*

Undoubtedly, my future research efforts will be guided by the ambition to restore the harmony between a formal description of the linguistic phenomenon and a genuine ontogenetic consideration of its cognitive and psychological nature.

## BIBLIOGRAPHY

- Adler, D., & Murdoch, D. (2025). *Package 'rgl'* (R package version 1.3.17). <https://github.com/dmurdoch/rgl>
- Adler, N. (2007). The MacArthur Scale of Subjective Social Status. *MacArthur Research Network on SES & Health*.
- Al-Azary, H., & Katz, A. N. (2021). Do metaphorical sharks bite? Simulation and abstraction in metaphor processing. *Memory and Cognition*, *49*(3), 557–570. <https://doi.org/10.3758/s13421-020-01109-2>
- American Psychological Association. (2022). *APA guidelines for Assessment and Intervention with Persons with Disabilities*. February, 61. <https://www.apa.org/about/policy/guidelines-assessment-intervention-disabilities.pdf>
- Amir, D., Vaggia, C., Srinivasan, M., Sugiyama, L. S., & Dunham, Y. (2019). Measuring subjective social status in children of diverse societies. *PLoS ONE*, *14*(12). <https://doi.org/10.1371/journal.pone.0226550>
- Asch, S. E., & Nerlove, H. (1960). The development of double-function terms in children. In B. Kaplan & S. Wapner (Eds.), *Perspectives in Psychological Theory*. International Universities Press.
- Ayhan, A. B., Aki, E., Mutlu, B., & Aral, N. (2015). A study of conceptual development and visual perception in six-year-old children. *Perceptual and Motor Skills*, *121*(3), 832–839. <https://doi.org/10.2466/24.10.PMS.121c22x7>
- Baldi, B. (2020). Persuasion we live by: symbols, metaphors and linguistic strategies. *Quaderni Di Linguistica e Studi Orientali*, *7220*, 337–382. <https://oaj.fupress.net/index.php/bsfm-qulso/article/view/9706>
- Bambini, V., Arcara, G., Bosinelli, F., Buonocore, M., Bechi, M., Cavallaro, R., & Bosia, M. (2020). A leopard cannot change its spots: A novel pragmatic account of concretism in schizophrenia. *Neuropsychologia*, *139*(January), 107332. <https://doi.org/10.1016/j.neuropsychologia.2020.107332>
- Bambini, V., & Domaneschi, F. (2024). Twenty years of experimental pragmatics. New advances in scalar implicature and metaphor processing. *Cognition*, *244*, 105708. <https://doi.org/10.1016/j.cognition.2023.105708>
- Bambini, V., & Frau, F. (2022). Neuropragmatics. In F. Domaneschi & V. Bambini (Eds.), *Pragmatica Sperimentale*. Il Mulino.
- Bambini, V., Frau, F., Bischetti, L., Agostoni, G., Mevio, C., Battaglini, C., Bechi, M., Buonocore, M., Sapienza, J., Spangaro, M., Guglielmino, C., Cocchi, F., Cavallaro, R., & Bosia, M. (2025). From semantic concreteness to concretism in schizophrenia: An automated linguistic analysis of speech produced in figurative language interpretation. *Clinical Linguistics & Phonetics*, 1–23. <https://doi.org/10.1080/02699206.2025.2451961>
- 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), 203–216. <https://doi.org/10.1016/j.brainresbull.2011.07.015>
- Bambini, V., Ghio, M., Moro, A., & Schumacher, P. B. (2013). Differentiating among pragmatic uses of words through timed sensicality judgments. *Frontiers in Psychology*, *4*, 938. <https://doi.org/10.3389/fpsyg.2013.00938>
- Bambini, V., Pompei, C., & Scalingi, B. (2022). Metafora. In F. Domaneschi & V. Bambini (Eds.), *Pragmatica Sperimentale* (pp. 55–66). Il Mulino.
- Bambini, V., Ranieri, G., Bischetti, L., Scalingi, B., Bertini, C., Ricci, I., Schaeken, W., & Canal, P. (2024). The costs of multimodal metaphors: comparing ERPs to figurative expressions in verbal and verbo-

- pictorial formats. *Discourse Processes*, 61(1–2), 44–68.  
<https://doi.org/10.1080/0163853X.2023.2282895>
- Bambini, V., Van Looy, L., Demiddele, K., & Schaeken, W. (2021). What is the contribution of executive functions to communicative-pragmatic skills? Insights from aging and different types of pragmatic inference. *Cognitive Processing*, 22(3), 435–452. <https://doi.org/10.1007/s10339-021-01021-w>
- Barajas, C., Linero, M.-J., & Alarcón, R. (2022). Persuasion ability in children from 6 to 12 years old: Relations to cognitive and affective theory of mind. *Frontiers in Psychology*, 13.  
<https://doi.org/10.3389/fpsyg.2022.966102>
- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3), 209–226. <https://doi.org/10.1007/s10579-009-9081-4>
- Barrett, M. D. (1986). Early Semantic Representations and Early Word-Usage. In S. A. Kuczaj & M. D. Barret (Eds.), *The Development of Word Meaning*. (pp. 39–67). Springer Series in Cognitive Development. Springer. [https://doi.org/10.1007/978-1-4612-4844-6\\_2](https://doi.org/10.1007/978-1-4612-4844-6_2)
- Bartsch, K., Wade, C. E., & Estes, D. (2011). Children’s Attention to Others’ Beliefs during Persuasion: Improvised and Selected Arguments to Puppets and People. *Social Development*, 20(2), 316–333.  
<https://doi.org/10.1111/j.1467-9507.2010.00580.x>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Haubo, R., Christensen, B., Singmann, H., Dai, B., & Scheipl, F. (2024). *Package ‘lme4’* (R package version 1.1-36).
- Beard, E., Henninger, N. M., & Venkatraman, V. (2024). Making Ads Stick: Role of Metaphors in Improving Advertising Memory. *Journal of Advertising*, 53(1), 86–103.  
<https://doi.org/10.1080/00913367.2022.2089302>
- Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2), 757–780.  
<https://doi.org/10.3758/s13428-020-01453-w>
- Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, 39(12), 2115–2132.  
<https://doi.org/10.1016/j.pragma.2007.05.009>
- Billow, R. M. (1981). Observing spontaneous metaphor in children. *Journal of Experimental Child Psychology*, 31(3), 430–445. [https://doi.org/10.1016/0022-0965\(81\)90028-X](https://doi.org/10.1016/0022-0965(81)90028-X)
- Bird, L. H., Thomas, E. F., Wenzel, M., & Lizzio-Wilson, M. (2024). Thinking about the future: Examining the exacerbating and attenuating factors of despair-induced climate burnout. *Journal of Environmental Psychology*, 98, 102382. <https://doi.org/10.1016/j.jenvp.2024.102382>
- Bischetti, L., Frau, F., & Bambini, V. (2024). Neuropragmatics. In *The Handbook of Clinical Linguistics, Second Edition* (pp. 41–54). Wiley. <https://doi.org/10.1002/9781119875949.ch4>
- Boeynaems, A., Burgers, C., Konijn, E. A., & Steen, G. J. (2017). The Effects of Metaphorical Framing on Political Persuasion: A Systematic Literature Review. *Metaphor and Symbol*, 32(2), 118–134.  
<https://doi.org/10.1080/10926488.2017.1297623>
- Bohn, M., & Frank, M. C. (2019). The Pervasive Role of Pragmatics in Early Language. *Annual Review of Developmental Psychology*, 1(1), 223–249. <https://doi.org/10.1146/annurev-devpsych-121318-085037>
- Bowerman, M. (1982). Starting to talk worse: Clues to language acquisition from children’s late speech errors. In M. Bowerman, S. Strauss, & R. Stavy (Eds.), *U shaped behavioral growth* (pp. 101–145). Academic Press.
- Bowes, A., & Katz, A. (2015). Metaphor creates intimacy and temporarily enhances theory of mind. *Memory and Cognition*, 43(6), 953–963. <https://doi.org/10.3758/s13421-015-0508-4>

- Boyes, E., Myers, G., Skamp, K., Stanisstreet, M., & Yeung, S. (2007). Air quality: a comparison of students' conceptions and attitudes across the continents. *Compare: A Journal of Comparative and International Education*, 37(4), 425–445. <https://doi.org/10.1080/03057920701366176>
- Burgers, C., Konijn, E. A., & Steen, G. J. (2016). Figurative Framing: Shaping Public Discourse Through Metaphor, Hyperbole, and Irony. *Communication Theory*, 26(4), 410–430. <https://doi.org/10.1111/comt.12096>
- Burns, G. W. (2024). Using metaphors to build hope and hopefulness from depression. *American Journal of Clinical Hypnosis*, 66(1), 48–60. <https://doi.org/10.1080/00029157.2023.2210178>
- Cacciari, C. (2008). Crossing the Senses in Metaphorical Language. In *The Cambridge Handbook of Metaphor and Thought* (pp. 425–444). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816802.026>
- Cacciari, C., & Padovani, R. (2012). The development of figurative language. In M. J. Spivey, K. McRae, & M. F. Joannisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 505–522). Cambridge University Press. <https://doi.org/10.1017/CBO9781139029377.034>
- Cameron-Faulkner, T., Macdonald, R., Serratrice, L., Melville, J., & Gattis, M. (2017). Plant Yourself Where Language Blooms: Direct Experience of Nature Changes How Parents and Children Talk about Nature. *Children, Youth and Environments*, 27(2), 110. <https://doi.org/10.7721/chilyoutenvi.27.2.0110>
- Cameron, L. (2003). *Metaphor in educational discourse*. Continuum.
- Canal, P., Bischetti, L., Bertini, C., Ricci, I., Lecce, S., & Bambini, V. (2022). N400 differences between physical and mental metaphors: The role of Theories of Mind. *Brain and Cognition*, 161, 105879. <https://doi.org/10.1016/j.bandc.2022.105879>
- Carapezza, M., & Garello, S. (2025). The Role of Literal Meaning in the Construction of Metaphorical Meaning. In A. Capone, R. Graci, & P. Perconti (Eds.), *New Frontiers in Pragmalinguistic Studies. Perspectives in Pragmatics, Philosophy & Psychology*, vol. 37 (pp. 35–48). Springer, Cham. [https://doi.org/10.1007/978-3-031-65502-9\\_3](https://doi.org/10.1007/978-3-031-65502-9_3)
- 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. <https://doi.org/10.1371/journal.pone.0150289>
- Carston, R. (2010a). Lexical pragmatics, ad hoc concepts and metaphor: A Relevance Theory perspective. *Italian Journal of Linguistics*, 22(1), 153–180.
- Carston, R. (2010b). Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images. *Proceedings of the Aristotelian Society (Hardback)*, 110(3), 295–321. <https://doi.org/10.1111/j.1467-9264.2010.00288.x>
- Carston, R. (2018). Figurative Language, Mental Imagery, and Pragmatics. *Metaphor and Symbol*, 33(3), 198–217. <https://doi.org/10.1080/10926488.2018.1481257>
- Ceccato, I., Lecce, S., Bischetti, L., Mangiaterra, V., di San Pietro, C. B., Cavallini, E., & Bambini, V. (2025). Aging and the Division of Labor of Theory of Mind Skills in Metaphor Comprehension. *Topics in Cognitive Science*. <https://doi.org/10.1111/tops.12785>
- Chamila Roshani Perera, L., & Rathnasiri Hewege, C. (2013). Climate change risk perceptions and environmentally conscious behaviour among young environmentalists in Australia. *Young Consumers*, 14(2), 139–154. <https://doi.org/10.1108/17473611311325546>
- Chen, J. C., & Martin, A. R. (2015). Role-Play Simulations as a Transformative Methodology in Environmental Education. *Journal of Transformative Education*, 13(1), 85–102. <https://doi.org/10.1177/1541344614560196>
- Chen, J., Lv, J., & Chen, B. (2025). Crossing the cultural bridge: The role of inhibitory control during second language metaphor comprehension. *Bilingualism: Language and Cognition*, 1–17.

<https://doi.org/10.1017/S1366728924001081>

- Cheng, J. C. H., & Monroe, M. C. (2012). Connection to nature: Children's affective attitude toward nature. *Environment and Behavior*, *44*(1), 31–49. <https://doi.org/10.1177/0013916510385082>
- Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language*, *56*(2), 172–188. <https://doi.org/10.1016/j.jml.2006.11.006>
- Christensen, R. H. B. (2015). *Package "ordinal"* (R package version 2013.9-30).
- Christidou, V., Dimopoulos, K., & Koulaidis, V. (2004). Constructing social representations of science and technology: the role of metaphors in the press and the popular scientific magazines. *Public Understanding of Science*, *13*(4), 347–362. <https://doi.org/10.1177/0963662504044108>
- Cicone, M., Gardner, H., & Winner, E. (1981). Understanding the psychology in psychological metaphors. *Journal of Child Language*, *8*(1), 213–216. <https://doi.org/10.1017/S0305000900003123>
- Cisneros, J. D. (2008). Contaminated Communities: The Metaphor of “Immigrant as Pollutant” in Media Representations of Immigration. *Rhetoric & Public Affairs*, *11*(4), 569–601. <https://doi.org/10.1353/rap.0.0068>
- Clark, E. V. (2020). Perspective-taking and pretend-play: Precursors to figurative language use in young children. *Journal of Pragmatics*, *156*, 100–109. <https://doi.org/10.1016/j.pragma.2018.12.012>
- Clevenger, T., & Edwards, R. (1988). Semantic distance as a predictor of metaphor selection. *Journal of Psycholinguistic Research*, *17*(3), 211–226. <https://doi.org/10.1007/BF01686356>
- Colston, H. L., & Rasse, C. (2022). Figurativity. Cognitive, because it's social. In H. L. Colston, T. Matlock, & G. J. Steen (Eds.), *Dynamism in metaphor and beyond* (pp. 243–264). John Benjamins Publishing Company. <https://doi.org/10.1075/milcc.9.13col>
- Condit, C. M., & Condit, D. M. (2001). Blueprints and recipes: Gendered metaphors for genetic medicine. *Journal of Medical Humanities*, *22*(1), 29–39. <https://doi.org/10.1023/A:1026634010579>
- Conway, A. R. A., & Kovacs, K. (2013). *Individual Differences in Intelligence and Working Memory* (Vol. 58, pp. 233–270). <https://doi.org/10.1016/B978-0-12-407237-4.00007-4>
- Cornoldi, C., & Colpo, G. (1998). *Nuove Prove di Lettura MT*. Giunti O.S.
- Cortés, A. I. R., Cobos, F. J. M., & Tarbox, J. (2018). Teaching children to create metaphorical expressions. *International Journal of Psychology and Psychological Therapy*, *18*(1), 27–38.
- Crowhurst, M., & Piche, G. L. (1979). Audience and Mode of Discourse Effects on Syntactic Complexity in Writing at Two Grade Levels. *Research in the Teaching of English*, *13*(2), 101–109. <https://doi.org/10.58680/rte197917847>
- Cuevas, K., Calkins, S. D., & Bell, M. A. (2016). To Stroop or not to Stroop: Sex-related differences in brain-behavior associations during early childhood. *Psychophysiology*, *53*(1), 30–40. <https://doi.org/10.1111/psyp.12464>
- Cundiff, J. M., Smith, T. W., Uchino, B. N., & Berg, C. A. (2013). Subjective Social Status: Construct Validity and Associations with Psychosocial Vulnerability and Self-Rated Health. *International Journal of Behavioral Medicine*, *20*(1), 148–158. <https://doi.org/10.1007/s12529-011-9206-1>
- Daysh, S., Thomas, E. F., Lizzio-Wilson, M., Bird, L., & Wenzel, M. (2024). “The future will be green, or not at all”: How positive (utopian) and negative (dystopian) thoughts about the future shape collective climate action. *Global Environmental Psychology*, *2*. <https://doi.org/10.5964/gep.11153>
- DeCaro, M. S., & Van Stockum, C. A. (2018). Ego depletion improves insight. *Thinking & Reasoning*, *24*(3), 315–343. <https://doi.org/10.1080/13546783.2017.1396253>
- DeCaro, M. S., Van Stockum, C. A., & Wieth, M. B. (2016). When higher working memory capacity

- hinders insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(1), 39–49.  
<https://doi.org/10.1037/xlm0000152>
- Deignan, A., & Semino, E. (2019). Translating science for young people through metaphor. *Translator*, 25(4), 369–384. <https://doi.org/10.1080/13556509.2020.1735759>
- Deignan, A., & Semino, E. (2022). Metaphors and meaning-making in young people’s talk about climate change. In H. L. Colston, T. Matlock, & G. J. Steen (Eds.), *Dynamism in metaphor and beyond. Metaphor in language, cognition and communication* (Vol. 9, pp. 191–204). John Benjamins Publishing Company.  
<https://doi.org/10.1075/milcc.9>
- Deignan, A., Semino, E., & Paul, S.-A. (2019). Metaphors of Climate Science in Three Genres: Research Articles, Educational Texts, and Secondary School Student Talk. *Applied Linguistics*, 40(2), 379–403.  
<https://doi.org/10.1093/applin/amx035>
- Del Sette, P., Bambini, V., Bischetti, L., & Lecce, S. (2020). Longitudinal associations between theory of mind and metaphor understanding during middle childhood. *Cognitive Development*, 56(100958).  
<https://doi.org/10.1016/j.cogdev.2020.100958>
- Del Sette, P., Bambini, V., Tonini, E., & Lecce, S. (2024). Are theory of mind and metaphor comprehension causally related? A training study in middle childhood. *Language Acquisition*, 1–21.  
<https://doi.org/10.1080/10489223.2024.2345324>
- Del Sette, P., Ronchi, L., Bambini, V., & Lecce, S. (2021). Longitudinal associations between metaphor understanding and peer relationships in middle childhood. *Infant and Child Development*, 30(4), 1–22.  
<https://doi.org/10.1002/icd.2232>
- Despot, K. Š., Sekulić Sović, M., Vilibić, M., & Mimica, N. (2021). Metaphor Production by Patients with Schizophrenia – A Case Analysis. *Metaphor and Symbol*, 36(3), 119–140.  
<https://doi.org/10.1080/10926488.2021.1893606>
- DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, 23(3), 1616–1641. <https://doi.org/10.1016/j.chb.2005.08.012>
- Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*, 149, 23–40. <https://doi.org/10.1016/j.jecp.2015.07.011>
- Di Paola, S., Domaneschi, F., & Pouscoulous, N. (2020). Metaphorical developing minds: The role of multiple factors in the development of metaphor comprehension. *Journal of Pragmatics*, 156, 235–251.  
<https://doi.org/10.1016/j.pragma.2019.08.008>
- Diesendruck, G., & Bloom, P. (2003). How Specific is the Shape Bias? *Child Development*, 74(1), 168–178.  
<https://doi.org/10.1111/1467-8624.00528>
- Dillard, J. P. (2010). Persuasion. In *The Handbook of Communication Science* (pp. 203–218). SAGE Publications, Inc. <https://doi.org/10.4135/9781412982818.n12>
- Dillard, J. P., & Seo, K. (2012). Affect and persuasion. In J. P. Dillard & L. Shen (Eds.), *The SAGE Handbook of Persuasion: Developments in Theory and Practice* (Issue January, pp. 150–166). SAGE Publications. <https://doi.org/10.4135/9781452218410.n10>
- Doherty, M., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development*, 13(3), 279–305. [https://doi.org/10.1016/S0885-2014\(98\)90012-0](https://doi.org/10.1016/S0885-2014(98)90012-0)
- Domaneschi, F., & Bambini, V. (2020). Pragmatic Competence. In E. Fridland & C. Pavese (Eds.), *The Routledge Handbook of Philosophy of Skill And Expertise* (pp. 419–430). Routledge.  
<https://doi.org/10.4324/9781315180809-40>
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test*. American Guidance Service.
- Elvevåg, B., Helsen, K., De Hert, M., Sweers, K., & Storms, G. (2011). Metaphor interpretation and use:

- A window into semantics in schizophrenia. *Schizophrenia Research*, 133(1–3), 205–211.  
<https://doi.org/10.1016/j.schres.2011.07.009>
- Estlein, R. (2021). Parenting as a Communication Process: Integrating Interpersonal Communication Theory and Parenting Styles Conceptualization. *Journal of Family Theory & Review*, 13(1), 21–33.  
<https://doi.org/10.1111/jftr.12407>
- Fainsilber, L., & Ortony, A. (1987). Metaphorical Uses of Language in the Expression of Emotions. *Metaphor and Symbolic Activity*, 2(4), 239–250. [https://doi.org/10.1207/s15327868ms0204\\_2](https://doi.org/10.1207/s15327868ms0204_2)
- Falkum, I. L. (2018). Pragmatic development: Learning to use language to communicate. In J. S. Horst & J. Von Koss Torkildsen (Eds.), *International Handbook of Language Development* (Issue October, pp. 1–30). Routledge. <https://doi.org/10.13140/RG.2.2.27698.86721>
- Falkum, I. L. (2022). The development of non-literal uses of language: Sense conventions and pragmatic competence. *Journal of Pragmatics*, 188, 97–107. <https://doi.org/10.1016/j.pragma.2021.12.002>
- Fasiolo, M., Nedellec, R., Goude, Y., Capezza, C., & Wood, S. N. (2023). *Package 'mgcVis'* (R package version 0.1.11).
- Figuerola, M., & Gerken, L. (2019). Experience with morphosyntactic paradigms allows toddlers to tacitly anticipate overregularized verb forms months before they produce them. *Cognition*, 191, 103977.  
<https://doi.org/10.1016/j.cognition.2019.05.014>
- Flavell, J. H., Green, F. L., Flavell, E. R., Watson, M. W., & Campione, J. C. (1986). Development of Knowledge about the Appearance-Reality Distinction. *Monographs of the Society for Research in Child Development*, 51(1), i. <https://doi.org/10.2307/1165866>
- Fleer, M. (2013). Affective Imagination in Science Education: Determining the Emotional Nature of Scientific and Technological Learning of Young Children. *Research in Science Education*, 43(5), 2085–2106. <https://doi.org/10.1007/s11165-012-9344-8>
- Florit, E., Cain, K., & Mason, L. (2020). Going beyond children's single-text comprehension: The role of fundamental and higher-level skills in 4 th graders' multiple-document comprehension. *British Journal of Educational Psychology*, 90(2), 449–472. <https://doi.org/10.1111/bjep.12288>
- Flusberg, S. J., Matlock, T., & Thibodeau, P. H. (2017). Metaphors for the War (or Race) against Climate Change. *Environmental Communication*, 11(6), 769–783.  
<https://doi.org/10.1080/17524032.2017.1289111>
- Flusberg, S. J., & Thibodeau, P. H. (2023). Why Is Mother Earth on Life Support? Metaphors in Environmental Discourse. *Topics in Cognitive Science*, 15(3), 522–545.  
<https://doi.org/10.1111/tops.12651>
- Forgács, B., & Pléh, C. (2022). The Fluffy Metaphors of Climate Science. In S. Wuppuluri & A. C. Grayling (Eds.), *Metaphors and Analogies in Sciences and Humanities. Synthese Library, vol 453* (pp. 447–477). Springer, Cham. [https://doi.org/10.1007/978-3-030-90688-7\\_22](https://doi.org/10.1007/978-3-030-90688-7_22)
- Frau, F., Cerami, C., Dodich, A., Bosia, M., & Bambini, V. (2024). Weighing the role of social cognition and executive functioning in pragmatics in the schizophrenia spectrum: A systematic review and meta-analysis. *Brain and Language*, 252, 105403. <https://doi.org/10.1016/j.bandl.2024.105403>
- Freeley, A. J., & Steinberg, D. L. (2009). *Argumentation and debate : critical thinking for reasoned decision making*. Wadsworth/Cengage Learning.
- Gardner, H., Kircher, M., Winner, E., & Perkins, D. (1975). Children's metaphoric productions and preferences. *Journal of Child Language*, 2(1), 125–141. <https://doi.org/10.1017/S0305000900000921>
- Gardner, H., & Winner, E. (1978). The Development of Metaphoric Competence: Implications for Humanistic Disciplines. *Critical Inquiry*, 5(1), 123–141.
- Gaskins, D., & Rundblad, G. (2023). Metaphor production in the bilingual acquisition of English and

- Polish. *Frontiers in Psychology*, 14(August), 1–18. <https://doi.org/10.3389/fpsyg.2023.1162486>
- Gass, R. H., & Seiter, J. S. (2022). *Persuasion*. Routledge. <https://doi.org/10.4324/9781003081388>
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 3-7 years old on a stroop- like day-night test. *Cognition*, 53(2), 129–153. [https://doi.org/10.1016/0010-0277\(94\)90068-X](https://doi.org/10.1016/0010-0277(94)90068-X)
- Giatti, L., Camelo, L. D. V., Rodrigues, J. F. D. C., & Barreto, S. M. (2012). Reliability of the MacArthur scale of subjective social status - Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *BMC Public Health*, 12(1096). <https://doi.org/10.1186/1471-2458-12-1096>
- Gibbs, R. W. (2006). Metaphor Interpretation as Embodied Simulation. *Mind and Language*, 21(3), 434–458. <https://doi.org/10.1111/j.1468-0017.2006.00285.x>
- Gopinath, S., & Kumar, A. (2024). The effect of an environmental education program based on empathy and reflective thinking on preadolescents' environmental values and knowledge. *International Research in Geographical and Environmental Education*, 1–20. <https://doi.org/10.1080/10382046.2024.2349430>
- Gopnik, A., & Astington, J. W. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child Development*, 59(1), 26. <https://doi.org/10.2307/1130386>
- Gottfried, G. M. (1997). Using metaphors as modifiers: Children's production of metaphoric compounds. *Journal of Child Language*, 24(3), 567–601. <https://doi.org/10.1017/S0305000997003176>
- Grazzani, I., & Ornaghi, V. (2012). How do use and comprehension of mental-state language relate to theory of mind in middle childhood? *Cognitive Development*, 27(2), 99–111. <https://doi.org/10.1016/j.cogdev.2012.03.002>
- Grigoroglou, M., & Papafragou, A. (2019). The development of pragmatic abilities. In K. Scott, B. Clark, & R. Carston (Eds.), *Relevance, Pragmatics and Interpretation* (pp. 102–112). Cambridge University Press. <https://doi.org/10.1017/9781108290593.010>
- Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, 47(4), 930–944. <https://doi.org/10.3758/s13428-014-0529-0>
- Haladyna, T. M., Rodriguez, M. C., & Stevens, C. (2019). Are Multiple-choice Items Too Fat? *Applied Measurement in Education*, 32(4), 350–364. <https://doi.org/10.1080/08957347.2019.1660348>
- Hanne, M., & Hawken, S. J. (2007). Metaphors for illness in contemporary media. *Medical Humanities*, 33(2), 93–99. <https://doi.org/10.1136/jmh.2006.000253>
- 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), 129–154. <https://doi.org/10.1007/BF02172093>
- Happé, F., & Loth, E. (2002). 'Theory of Mind' and Tracking Speakers' Intentions. *Mind & Language*, 17(1–2), 24–36. <https://doi.org/10.1111/1468-0017.00187>
- Harris, Z. S. (1954). Distributional Structure. *Word*, 10(2–3), 146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- Hartig, F. (2024). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. In *CRAN: Contributed Packages* (R package version 0.4.7). <https://doi.org/10.32614/CRAN.package.DHARMa>
- Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–310. <https://doi.org/10.1214/ss/1177013604>
- Heilmann, J., Malone, T. O., & F. Westerveld, M. (2020). Properties of Spoken Persuasive Language

- Samples From Typically Developing Adolescents. *Language, Speech, and Hearing Services in Schools*, 51(2), 441–456. [https://doi.org/10.1044/2019\\_LSHSS-19-00078](https://doi.org/10.1044/2019_LSHSS-19-00078)
- Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, 120(8), 1887–1897. <https://doi.org/10.1016/j.lingua.2010.02.002>
- Hiller, R. M., Weber, N., & Young, R. L. (2014). The validity and scalability of the Theory of Mind Scale with toddlers and preschoolers. *Psychological Assessment*, 26(4), 1388–1393. <https://doi.org/10.1037/a0038320>
- Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin*, 144(6), 641–671. <https://doi.org/10.1037/bul0000145>
- James, C. (2013). *Errors in Language Learning and Use*. Routledge. <https://doi.org/10.4324/9781315842912>
- Johnstone, C., & Stickles, E. (2024). Navigating the climate change minefield: the influence of metaphor on climate doomism. *Frontiers in Communication*, 9(May), 1–15. <https://doi.org/10.3389/fcomm.2024.1380092>
- Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible Words are Recognized Faster: The Grounding of Meaning in Sensory and Perceptual Systems. *Quarterly Journal of Experimental Psychology*, 64(9), 1683–1691. <https://doi.org/10.1080/17470218.2011.605150>
- Kalandadze, T., Bambini, V., & Næss, K. A. B. (2019). A systematic review and meta-analysis of studies on metaphor comprehension in individuals with autism spectrum disorder: Do task properties matter? *Applied Psycholinguistics*, 40(6), 1421–1454. <https://doi.org/10.1017/S0142716419000328>
- Kasirer, A., & Mashal, N. (2016). Comprehension and generation of metaphors by children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 32, 53–63. <https://doi.org/10.1016/j.rasd.2016.08.003>
- Kasirer, A., & Mashal, N. (2018). Fluency or Similarities? Cognitive Abilities that Contribute to Creative Metaphor Generation. *Creativity Research Journal*, 30(2), 205–211. <https://doi.org/10.1080/10400419.2018.1446747>
- Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1), 67–81. <https://doi.org/10.1016/j.cognition.2011.02.015>
- Katz, A. N. (1989). On choosing the vehicles of metaphors: Referential concreteness, semantic distances, and individual differences. *Journal of Memory and Language*, 28(4), 486–499. [https://doi.org/10.1016/0749-596X\(89\)90023-5](https://doi.org/10.1016/0749-596X(89)90023-5)
- Katz, A. N. (1992). Psychological Studies in Metaphor Processing: Extensions to the Placement of Terms in Semantic Space. *Poetics Today*, 13(4), 607. <https://doi.org/10.2307/1773291>
- Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, 68(8), 1665–1692. <https://doi.org/10.1080/17470218.2015.1022560>
- Khafaga, A., Bosli, R., & Al-Anazi, M. (2023). The Pragmatics of Persuasion in Fictional Communication. *World Journal of English Language*, 13(8), 474. <https://doi.org/10.5430/wjel.v13n8p474>
- Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7(2), 257–266. <https://doi.org/10.3758/BF03212981>
- Kintsch, W., & Bowles, A. R. (2002). Metaphor Comprehension: What Makes a Metaphor Difficult to Understand? *Metaphor and Symbol*, 17(4), 249–262. [https://doi.org/10.1207/s15327868ms1704\\_1](https://doi.org/10.1207/s15327868ms1704_1)
- Knudsen, S. (2003). Scientific metaphors going public. *Journal of Pragmatics*, 35(8), 1247–1263. [https://doi.org/10.1016/S0378-2166\(02\)00187-X](https://doi.org/10.1016/S0378-2166(02)00187-X)

- Köder, F., & Falkum, I. L. (2020). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics*, *156*, 191–205. <https://doi.org/10.1016/j.pragma.2019.07.007>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental Psychology*, *52*(3), 423–429. <https://doi.org/10.1037/dev0000089>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. <https://doi.org/10.18637/JSS.V082.I13>
- Lakoff, G. (2012). Metaphor and War: The Metaphor System Used to Justify War in the Gulf. *Cognitive Semiotics*, *4*(2), 5–19. <https://doi.org/10.1515/cogsem.2012.4.2.5>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago press.
- Lampri, S., Peristeri, E., Marinis, T., & Andreou, M. (2024). Metaphor comprehension and production in verbally able children with Autism Spectrum Disorder. *Autism Research*, *17*(11), 2292–2304. <https://doi.org/10.1002/aur.3210>
- Lanfranchi, S., Cornoldi, C., & Vianello, R. (2004). Verbal and Visuospatial Working Memory Deficits in Children With Down Syndrome. *American Journal on Mental Retardation*, *109*(6), 456–466. [https://doi.org/10.1352/0895-8017\(2004\)109<456:VAVWMD>2.0.CO;2](https://doi.org/10.1352/0895-8017(2004)109<456:VAVWMD>2.0.CO;2)
- Lavoie, R., Main, K., King, C., & King, D. (2021). Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay. *Virtual Reality*, *25*(1), 69–81. <https://doi.org/10.1007/s10055-020-00440-y>
- Lavrič, M., & Naterer, A. (2020). The power of authoritative parenting: A cross-national study of effects of exposure to different parenting styles on life satisfaction. *Children and Youth Services Review*, *116*, 105274. <https://doi.org/10.1016/j.childyouth.2020.105274>
- Lecce, S., Bianco, F., & Hughes, C. (2021). Reading minds and reading texts: Evidence for independent and specific associations. *Cognitive Development*, *57*, 101010. <https://doi.org/10.1016/j.cogdev.2021.101010>
- Lecce, S., Pompei, C., Del Sette, P., Canal, P., & Bambini, V. (n.d.). *Developmental trajectories of physical and mental metaphor understanding in preschoolers*.
- Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language*, *46*(2), 393–407. <https://doi.org/10.1017/S030500091800048X>
- Lenci, A. (2008). Distributional semantics in linguistics and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31.
- Lenth, R. V. (2024). emmeans: Estimated Marginal Means, aka Least-Squares Means. In *CRAN: Contributed Packages* (R package version 1.10.6-090001). <https://doi.org/10.32614/CRAN.package.emmeans>
- Lillard, A. S. (1993). Pretend Play Skills and the Child's Theory of Mind. *Child Development*, *64*(2), 348–371. <https://doi.org/10.1111/j.1467-8624.1993.tb02914.x>
- Lindner, M. A., Lüdtkke, O., Grund, S., & Köller, O. (2017). The merits of representational pictures in educational assessment: Evidence for cognitive and motivational effects in a time-on-task analysis. *Contemporary Educational Psychology*, *51*, 482–492. <https://doi.org/10.1016/j.cedpsych.2017.09.009>
- Littlemore, J. (2016). Metaphor use in educational contexts: functions and variations. In E. Semino & Z.

- Demjén (Eds.), *The Routledge Handbook of Metaphor and Language*. Routledge.  
<https://doi.org/10.4324/9781315672953>
- Littlemore, J. (2019). *Metaphors in the Mind*. Cambridge University Press.
- Lombardi Vallauri, E. (2022). Implicit strategies aimed at persuading the audience in public debates. *Intercultural Pragmatics*, 19(3), 299–319. <https://doi.org/10.1515/ip-2022-3002>
- Lombrozo, T. (2020). “Learning by Thinking” in Science and in Everyday Life. In A. Levy & P. Godfrey-Smith (Eds.), *The Scientific Imagination* (pp. 230–249). Oxford University Press New York.  
<https://doi.org/10.1093/oso/9780190212308.003.0010>
- Long, M., Shukla, V., & Rubio-Fernandez, P. (2021). The Development of Simile Comprehension: From Similarity to Scalar Implicature. *Child Development*, 92(4), 1439–1457.  
<https://doi.org/10.1111/cdev.13507>
- Lonigro, A., Baiocco, R., Baumgartner, E., Sette, S., & Laghi, F. (2017). Persuasion in school-aged children: How does it change if the persuadee is the mother or the peer? *Cognitive Processing*, 18(1), 67–74. <https://doi.org/10.1007/s10339-016-0776-3>
- Madan, C. R., Bayer, J., Gamer, M., Lonsdorf, T. B., & Sommer, T. (2018). Visual Complexity and Affect: Ratings Reflect More Than Meets the Eye. *Frontiers in Psychology*, 8.  
<https://doi.org/10.3389/fpsyg.2017.02368>
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236.  
<https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development*, 57(4). <https://doi.org/10.2307/1166115>
- Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., & Bailenson, J. N. (2018). Immersive Virtual Reality Field Trips Facilitate Learning About Climate Change. *Frontiers in Psychology*, 9.  
<https://doi.org/10.3389/fpsyg.2018.02364>
- Martín-González, I., R. Ronderos, C., Castroviejo, E., Schroeder, K., Lossius-Falkum, I., & Vicente, A. (2024). That kid is a grasshopper! Metaphor development from 3 to 9 years of age. *Journal of Child Language*, 1–26. <https://doi.org/10.1017/S0305000924000187>
- Mashal, N. (2013). The role of working memory in the comprehension of unfamiliar and familiar metaphors. *Language and Cognition*, 5(4), 409–436. <https://doi.org/10.1515/langcog-2013-0024>
- Mashal, N., & Kasirer, A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities*, 32(6), 2045–2054.  
<https://doi.org/10.1016/j.ridd.2011.08.012>
- Mashal, N., Zeev-Wolf, M., Citron, F. M. M., Papagno, C., Rossetti, I., & Brambilla, P. (2018). *Metaphor Comprehension in Schizophrenic Patients*. <https://doi.org/10.3389/fpsyg.2018.00670>
- Matthews, D. (2014). *Pragmatic Development in First Language Acquisition* (D. Matthews (ed.); Vol. 10). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.10>
- 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).  
<https://doi.org/10.1080/15475441.2018.1455584>
- Mazzarella, D., & Vaccargiu, E. (2024). Communication: Inferring speaker intentions or perceiving the world? Insights from developmental research. *Journal of Pragmatics*, 221, 123–136.  
<https://doi.org/10.1016/j.pragma.2023.12.014>

- Mazzola, V., & Camaioni, L. (2002). *Strane storie*. Dipartimento di Psicologia dinamica e clinica, Università La Sapienza.
- McQuarrie, E. F., & Phillips, B. J. (2005). INDIRECT PERSUASION IN ADVERTISING: How Consumers Process Metaphors Presented in Pictures and Words. *Journal of Advertising*, 34(2), 7–20. <https://doi.org/10.1080/00913367.2005.10639188>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12. <http://arxiv.org/abs/1301.3781>
- 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), 622–646. <https://doi.org/10.1111/j.1467-8624.2007.01018.x>
- Mio, J., & Katz, A. N. (1966). *Metaphor: implications and applications*. Lawrence Erlbaum Associates Inc.
- Monroe, M. C., Plate, R. R., Oxarart, A., Bowers, A., & Chaves, W. A. (2019). Identifying effective climate change education strategies: a systematic review of the research. *Environmental Education Research*, 25(6), 791–812. <https://doi.org/10.1080/13504622.2017.1360842>
- Montanari, F. (1996). Introduzione. In M. Dorati (Ed.), *Aristotele, Retorica*. Mondadori.
- Moore, C., & Davidge, J. (1989). The development of mental terms: pragmatics or semantics? *Journal of Child Language*, 16(3), 633–641. <https://doi.org/10.1017/S030500090001076X>
- Moro, A. (2011). “Kataptation” or the Qwerty-Effect in Language Evolution. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00050>
- Morote, Á. F., & Hernández, M. (2022). What Do School Children Know about Climate Change? A Social Sciences Approach. *Social Sciences*, 11(4). <https://doi.org/10.3390/socsci11040179>
- Muran, J. (1990). Towards a cognitive formulation of metaphor use in psychotherapy. *Clinical Psychology Review*, 10(1), 69–85. [https://doi.org/10.1016/0272-7358\(90\)90107-L](https://doi.org/10.1016/0272-7358(90)90107-L)
- Naglieri, J. A., & Maxwell, S. (1981). Inter-Rater Reliability and Concurrent Validity of the Goodenough-Harris and McCarthy Draw-a-Child Scoring Systems. *Perceptual and Motor Skills*, 53(2), 343–348. <https://doi.org/10.2466/pms.1981.53.2.343>
- Neff, M., & Falkum, I. L. (2022). *Exploring the literal bias in children’s metaphor development*. <https://doi.org/https://doi.org/10.17605/OSF.IO/VAUW2>
- Neff, M., Ganea, P. A., & Falkum, I. L. (2023). *Do children interpret metaphors literally because they are fantastical thinkers?* <https://doi.org/https://doi.org/10.17605/OSF.IO/B5DU8>
- Neill, D. K. O., & Astington, J. W. (1992). *Young Children’s Understanding of the Role That Sensory Experiences Play in Knowledge Acquisition Author (s): Daniela K. O’Neill, Janet Wilde Astington and John H. Flavell Published by: Wiley on behalf of the Society for Research in Child Develo.* 63(2), 474–490.
- Nelkin, D. (2001). Molecular metaphors: the gene in popular discourse. *Nature Reviews Genetics*, 2(7), 555–559. <https://doi.org/10.1038/35080583>
- Niebert, K., & Gropengiesser, H. (2013). Understanding and communicating climate change in metaphors. *Environmental Education Research*, 19(3), 282–302. <https://doi.org/10.1080/13504622.2012.690855>
- Nippold, M. A., Leonard, L. B., & Kail, R. (1984). Syntactic and conceptual factors in children’s understanding of metaphors. *Journal of Speech and Hearing Research*, 27(2), 197–205. <https://doi.org/10.1044/jshr.2702.197>
- Nippold, M. A., & Sullivan, M. P. (1987). Verbal and Perceptual Analogical Reasoning and Proportional Metaphor Comprehension in Young Children. *Journal of Speech, Language, and Hearing Research*, 30(3),

- 367–376. <https://doi.org/10.1044/jshr.3003.367>
- Nippold, M. A., Ward-Loneragan, J. M., & Fanning, J. L. (2005). Persuasive Writing in Children, Adolescents, and Adults. *Language, Speech, and Hearing Services in Schools*, *36*(2), 125–138. [https://doi.org/10.1044/0161-1461\(2005/012\)](https://doi.org/10.1044/0161-1461(2005/012))
- Norbury, C. F. (2005a). 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), 383–399. <https://doi.org/10.1348/026151005X26732>
- Norbury, C. F. (2005b). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *The British Psychological Society*, *23*, 383–399. <https://doi.org/10.1348/026151005X26732>
- Nord, C. (2008). Persuading by addressing: a functional approach to speech-act comparison. *Southern African Linguistics and Applied Language Studies*, *26*(2), 283–293. <https://doi.org/10.2989/SALALS.2008.26.2.8.572>
- Noveck, I. A. (2018). *Experimental Pragmatics: the making of a cognitive science*. Cambridge University Press. <https://doi.org/https://hal.science/hal-01879965v1>
- O'Reilly, D., & Marsden, E. (2023). Elicited metaphoric competence in a second language: a construct associated with vocabulary knowledge and general proficiency? *IRAL - International Review of Applied Linguistics in Language Teaching*, *61*(2), 287–327. <https://doi.org/10.1515/iral-2020-0054>
- Okanda, M., Asada, K., Moriguchi, Y., & Itakura, S. (2015). Understanding violations of Gricean maxims in preschoolers and adults. *Frontiers in Psychology*, *6*(July), 1–7. <https://doi.org/10.3389/fpsyg.2015.00901>
- Ortony, A. (1975). Why Metaphors Are Necessary and Not Just Nice 1. *Educational Theory*, *25*(1), 45–53. <https://doi.org/10.1111/j.1741-5446.1975.tb00666.x>
- Ottati, V. C., & Renstrom, R. A. (2010). Metaphor and Persuasive Communication: A Multifunctional Approach. *Social and Personality Psychology Compass*, *4*(9), 783–794. <https://doi.org/10.1111/j.1751-9004.2010.00292.x>
- Paivio, A., & Walsh, M. (2012). Psychological processes in metaphor comprehension and memory. *Metaphor and Thought*, *1975*, 307–328. <https://doi.org/10.1017/cbo9781139173865.016>
- Pascual, B., Aguado, G., Sotillo, M., & Masdeu, J. C. (2008). Acquisition of mental state language in Spanish children: a longitudinal study of the relationship between the production of mental verbs and linguistic development. *Developmental Science*, *11*(4), 454–466. <https://doi.org/10.1111/j.1467-7687.2008.00691.x>
- Pérez-Hernández, L., & Duvignau, K. (2016). Metaphor, metonymy, and their interaction in the production of semantic approximations by monolingual children: A corpus analysis. *First Language*, *36*(4), 383–406. <https://doi.org/10.1177/0142723716648845>
- Pérez-Sobrinho, P. (2016). Multimodal Metaphor and Metonymy in Advertising: A Corpus-Based Account. *Metaphor and Symbol*, *31*(2), 73–90. <https://doi.org/10.1080/10926488.2016.1150759>
- Perloff, R. M. (1993). *The Dynamics of Persuasion*. Routledge. <https://doi.org/10.4324/9781410606884>
- Peterson, C. C., Slaughter, V., & Wellman, H. M. (2018). Nimble negotiators: How theory of mind (ToM) interconnects with persuasion skills in children with and without ToM delay. *Developmental Psychology*, *54*(3), 494–509. <https://doi.org/10.1037/dev0000451>
- Petit, N., Bambini, V., Bischetti, L., Prado, J., & Noveck, I. (2024). How do theory of mind and formal language skills impact metaphoric reference comprehension during children's school-age years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0001381>
- Pierce, R. S., & Chiappe, D. L. (2009). The roles of aptness, conventionality, and working memory in the

- production of metaphors and similes. *Metaphor and Symbol*, 24(1), 1–19.  
<https://doi.org/10.1080/10926480802568422>
- Pollio, H. R., Barloe, J. M., Fine, H. J., & R, P. M. (1997). *Psychology and the Poetics of Growth: Figurative Language in Psychology, Psychotherapy, and Education*. Lawrence Erlbaum Associates Inc.
- Pompei, C., Lecce, S., Canal, P., Del Sette, P., & Bambini, V. (2023). How a white dog becomes a cloud: behavioral and computational approaches to metaphor production in preschoolers. *Crossing The Borders: The Interplay of Language, Cognition, and the Brain in Early Human Development*.  
<https://doi.org/https://crossing-project.de/ctb-conference-2023/>
- Pompei, C., Mangiaterra, V., Frau, F., Bischetti, L., Lecce, S., & Bambini, V. (2024). Metaphors as the earth's parachute: The effectiveness of the MetaCom green program in fostering climate hope in children. *Communicating Climate Hope: Countering Eco-Anxiety and Climate Doomism in Research and Practice*.  
<https://doi.org/https://climatehope.sites.olt.ubc.ca/attendees-only-area/book-of-abstracts/>
- Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics*, 25(2011), 51–79.  
<https://doi.org/10.1075/bjl.25.04pou>
- Pouscoulous, N. (2014). The elevator's buttocks. In *Metaphorical abilities in children*. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 239–259). John Benjamins Publishing Company.
- Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, 156, 160–167. <https://doi.org/10.1016/j.jpragma.2019.05.021>
- R Core Team. (2023). *RStudio: Integrated Development Environment for R* (4.3.2).
- Rakison, D. H. (2005). The Perceptual to Conceptual Shift in Infancy and Early Childhood: A Surface or Deep Distinction? In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building Object Categories in Developmental Time*. Psychology Press. <https://doi.org/10.4324/9781410612908>
- Reid, J. N., & Katz, A. N. (2018). Vector Space Applications in Metaphor Comprehension. *Metaphor and Symbol*, 33(4), 280–294. <https://doi.org/10.1080/10926488.2018.1549840>
- Rochat, P. (2007). Intentional action arises from early reciprocal exchanges. *Acta Psychologica*, 124(1), 8–25.  
<https://doi.org/10.1016/j.actpsy.2006.09.004>
- Rodriguez, M. C. (2005). Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. *Educational Measurement: Issues and Practice*, 24(2), 3–13.  
<https://doi.org/10.1111/j.1745-3992.2005.00006.x>
- Rooney-Varga, J. N., Kapmeier, F., Sterman, J. D., Jones, A. P., Putko, M., & Rath, K. (2020). The Climate Action Simulation. *Simulation & Gaming*, 51(2), 114–140.  
<https://doi.org/10.1177/1046878119890643>
- Rossi, M. G., & Macagno, F. (2021). The Communicative Functions of Metaphors Between Explanation and Persuasion. In F. Macagno & A. Capone (Eds.), *Inquiries in Philosophical Pragmatics. Perspectives in Pragmatics, Philosophy & Psychology* (pp. 171–191). Springer. [https://doi.org/10.1007/978-3-030-56437-7\\_12](https://doi.org/10.1007/978-3-030-56437-7_12)
- Rousell, D., & Cutter-Mackenzie-Knowles, A. (2020). A systematic review of climate change education: giving children and young people a 'voice' and a 'hand' in redressing climate change. *Children's Geographies*, 18(2), 191–208. <https://doi.org/10.1080/14733285.2019.1614532>
- Rubio-Fernández, P., & Grassmann, S. (2016). Metaphors as Second Labels: Difficult for Preschool Children? *Journal of Psycholinguistic Research*, 45(4), 931–944. <https://doi.org/10.1007/s10936-015-9386-y>
- Rubio Fernandez, P. (2007). Suppression in Metaphor Interpretation: Differences between Meaning Selection and Meaning Construction. *Journal of Semantics*, 24(4), 345–371.  
<https://doi.org/10.1093/jos/ffm006>

- Rueda, M. R., Cómbita, L. M., & Pozuelos, J. P. (2021). Cognitive Training in Childhood and Adolescence. In *Cognitive Training* (pp. 127–139). Springer International Publishing. [https://doi.org/10.1007/978-3-030-39292-5\\_9](https://doi.org/10.1007/978-3-030-39292-5_9)
- Rundblad, G., & Annaz, D. (2010). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*(3), 547–563. <https://doi.org/10.1348/026151009X454373>
- Sana, F., Park, J., Gagné, C. L., & Spalding, T. L. (2021). The interplay between inhibitory control and metaphor conventionality. *Memory and Cognition*, *49*(6), 1267–1284. <https://doi.org/10.3758/s13421-021-01152-7>
- Sapp, F., Lee, K., & Muir, D. (2000). Three-year-olds' difficulty with the appearance–reality distinction: Is it real or is it apparent? *Developmental Psychology*, *36*(5), 547–560. <https://doi.org/10.1037/0012-1649.36.5.547>
- Satchwell, C., Walley, B., Dodding, J., & Lagi, M. D. A. (2025). Conversations across international divides: Children learning through empathy about climate change. *Geographical Research*, *63*(1), 103–117. <https://doi.org/10.1111/1745-5871.12638>
- Sauerbrei, W., & Patrick, R. (2010). Continuous variables: to categorize or to model. *8th International Conference on Teaching Statistics—Data and Context in Statistics Education: Towards an Evidence Based Society. International Statistical Institute*.
- Sbisà, M. (2013). Locution, illocution, perlocution. In M. Sbisà & K. Turner (Eds.), *Pragmatics of speech actions* (pp. 25–75). Mouton de Gruyter.
- Schaeken, W., Haeren, M. Van, & Bambini, V. (2018). The understanding of scalar implicatures in children with autism spectrum disorder: Dichotomized responses to violations of informativeness. *Frontiers in Psychology*, *9*(JUL), 1–16. <https://doi.org/10.3389/fpsyg.2018.01266>
- Scherer, A. M., Scherer, L. D., & Fagerlin, A. (2015). Getting ahead of illness: Using metaphors to influence medical decision making. *Medical Decision Making*, *35*(1), 37–45. <https://doi.org/10.1177/0272989X14522547>
- Schleihauf, H., Herrmann, E., Fischer, J., & Engelmann, J. M. (2022). How children revise their beliefs in light of reasons. *Child Development*, *93*(4), 1072–1089. <https://doi.org/10.1111/cdev.13758>
- Semino, E. (2021). “Not Soldiers but Fire-fighters”—Metaphors and Covid-19. *Health Communication*, *36*(1), 50–58. <https://doi.org/10.1080/10410236.2020.1844989>
- Seo, K., Dillard, J. P., & Shen, F. (2013). The Effects of Message Framing and Visual Image on Persuasion. *Communication Quarterly*, *61*(5), 564–583. <https://doi.org/10.1080/01463373.2013.822403>
- Shaw, C., & Nerlich, B. (2015). Metaphor as a mechanism of global climate change governance: A study of international policies, 1992–2012. *Ecological Economics*, *109*, 34–40. <https://doi.org/10.1016/j.ecolecon.2014.11.001>
- Skalicky, S., & Crossley, S. A. (2018). Linguistic features of sarcasm and metaphor production quality. *Proceedings of the Workshop on Figurative Language Processing, Fig-Lang 2018 at the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HTL 2018, Colston 2017*, 7–16. <https://doi.org/10.18653/v1/w18-0902>
- Skamp, K., Boyes, E., & Stanisstreet, M. (2013). Beliefs and Willingness to Act About Global Warming: Where to Focus Science Pedagogy? *Science Education*, *97*(2), 191–217. <https://doi.org/10.1002/sce.21050>
- Slaughter, V., Peterson, C. C., & Moore, C. (2013). I can talk you into it: Theory of mind and persuasion behavior in young children. *Developmental Psychology*, *49*(2), 227–231. <https://doi.org/10.1037/a0028280>
- Sopory, P., & Dillard, J. P. (2002a). Figurative language and persuasion. In J. P. Dillard & M. Pfau (Eds.),

- The persuasion handbook: Developments in theory and practice* (pp. 407–426). SAGE Publications.
- Sopory, P., & Dillard, J. P. (2002b). The Persuasive Effects of Metaphor: A Meta-Analysis. *Human Communication Research*, 28(3), 382–419. <https://doi.org/10.1111/j.1468-2958.2002.tb00813.x>
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language*, 17(1–2), 3–23. <https://doi.org/10.1111/1468-0017.00186>
- Sperber, D., & Wilson, D. (2008). A Deflationary Account of Metaphors. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84–106). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816802.007>
- Stamenković, D., Ichien, N., & Holyoak, K. J. (2019). Metaphor comprehension: An individual-differences approach. *Journal of Memory and Language*, 105, 108–118. <https://doi.org/10.1016/j.jml.2018.12.003>
- Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). *Il peabody test. test di vocabolario recettivo*. Omega.
- Stevenson, K., & Peterson, N. M. (2015). Motivating Action through Fostering Climate Change Hope and Concern and Avoiding Despair among Adolescents. *Sustainability*, 8(1), 6. <https://doi.org/10.3390/su8010006>
- Stevenson, K. T., Peterson, N. M., Bondell, H. D., Moore, S. E., & Carrier, S. J. (2014). Overcoming skepticism with education: interacting influences of worldview and climate change knowledge on perceived climate change risk among adolescents. *Climatic Change*, 126(3–4), 293–304. <https://doi.org/10.1007/s10584-014-1228-7>
- Stone, A. A., Schwartz, J. E., Broderick, J. E., & Deaton, A. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences*, 107(22), 9985–9990. <https://doi.org/10.1073/pnas.1003744107>
- Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, 30(3), 395–402. <https://doi.org/10.1037/0012-1649.30.3.395>
- Sun, Z., & Firestone, C. (2021). Curious Objects: How Visual Complexity Guides Attention and Engagement. *Cognitive Science*, 45(4). <https://doi.org/10.1111/cogs.12933>
- Taber, F., & Taylor, N. (2009). Climate of Concern--A Search for Effective Strategies for Teaching Children about Global Warming. *International Journal of Environmental and Science Education*, 4(2), 97–116.
- Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE*, 6(2). <https://doi.org/10.1371/journal.pone.0016782>
- Thibodeau, P. H., & Boroditsky, L. (2013). Natural Language Metaphors Covertly Influence Reasoning. *PLoS ONE*, 8(1). <https://doi.org/10.1371/journal.pone.0052961>
- Thibodeau, P. H., Frantz, C. M. P., & Berretta, M. (2017). The earth is our home: systemic metaphors to redefine our relationship with nature. *Climatic Change*, 142(1–2), 287–300. <https://doi.org/10.1007/s10584-017-1926-z>
- Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How Linguistic Metaphor Scaffolds Reasoning. *Trends in Cognitive Sciences*, 21(11), 852–863. <https://doi.org/10.1016/j.tics.2017.07.001>
- Tonini, E., Bischetti, L., Del Sette, P., Tosi, E., Lecce, S., & Bambini, V. (2023). The relationship between metaphor skills and Theory of Mind in middle childhood: Task and developmental effects. *Cognition*, 238, 105504. <https://doi.org/10.1016/j.cognition.2023.105504>
- Tonini, E., Lecce, S., Del Sette, P., Bianco, F., Canal, P., & Bambini, V. (2022). Efficacy and benefits of the MetaCom training to promote metaphor comprehension in typical development. *First Language*, 42(3), 466–496. <https://doi.org/10.1177/01427237221081201>

- Tourangeau, R., & Rips, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory and Language*, 30(4), 452–472. [https://doi.org/10.1016/0749-596X\(91\)90016-D](https://doi.org/10.1016/0749-596X(91)90016-D)
- Traxler, M. J., & Gernsbacher, M. A. (2011). *Handbook of Psycholinguistics*. Elsevier.
- van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: a case study of continuity, discontinuity and variability in early language development. *Infant and Child Development*, 16(1), 7–33. <https://doi.org/10.1002/icd.506>
- van Rij, J., Wieling, M., & Baayen, R. H. (2022). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. In *CRAN: Contributed Packages* (R package version 2.4.1). <https://doi.org/10.32614/CRAN.package.itsadug>
- Van Rij, J., Wieling, M., Harald Baayen, R., Van Rijn, H., & Maintainer, J. (2022). *Title Interpreting Time Series and Autocorrelated Data Using GAMMs*.
- Vosniadou, S. (1987). Children and Metaphors. *Child Development*, 58(3), 870–885.
- Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: evidence from masked priming ERP. *Frontiers in Human Neuroscience*, 8(AUG), 1–17. <https://doi.org/10.3389/fnhum.2014.00583>
- Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided Play. *Current Directions in Psychological Science*, 25(3), 177–182. <https://doi.org/10.1177/0963721416645512>
- Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential Progressions in a Theory-of-Mind Scale: Longitudinal Perspectives. *Child Development*, 82(3), 780–792. <https://doi.org/10.1111/j.1467-8624.2011.01583.x>
- Wellman, H. M., & Liu, D. (2004a). Scaling of Theory-of-Mind Tasks. *Child Development*, 75(2), 523–541.
- Wellman, H. M., & Liu, D. (2004b). Scaling of Theory-of-Mind Tasks. *Child Development*, 75(2), 523–541. <https://doi.org/10.1111/j.1467-8624.2004.00691.x>
- Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, 39(4), 703–728. <https://doi.org/10.1017/S0142716418000152>
- White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing Mentalizing Impairments in Autism. *Child Development*, 80(4), 1097–1117. <https://doi.org/10.1111/j.1467-8624.2009.01319.x>
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116. <https://doi.org/10.1016/j.wocn.2018.03.002>
- Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology*, 39(7), 1257–1267. <https://doi.org/10.1002/ejsp.665>
- Williams, M. E., Sando, L., & Soles, T. G. (2014). Cognitive Tests in Early Childhood. *Journal of Psychoeducational Assessment*, 32(5), 455–476. <https://doi.org/10.1177/0734282913517526>
- Williams, P. E., Weiss, L. G., & Rolfhus, E. L. (2003). WISC – IV Technical Report # 2 Psychometric Properties. *WISC-IV Technical Manual #2, March 2000*, 1–6.
- Wilson, D., & Carston, R. (2006). Metaphor, relevance and the “emergent property” issue. *Mind and Language*, 21(3), 404–433. <https://doi.org/10.1111/j.1468-0017.2006.00284.x>
- Wilson, E., & Katsos, N. (2020). Acquiring implicatures. In K. P. Schneider & E. Ifantidou (Eds.), *Developmental and clinical pragmatics* (pp. 119–148). De Gruyter Mouton. <https://doi.org/https://doi.org/10.1515/9783110431056-005>
- Wilson, S. R., Guntzville, L. M., & Munz, E. A. (2012). Persuasion and Families. In *The Routledge*

- Handbook of Family Communication*. Routledge. <https://doi.org/10.4324/9780203848166.ch22>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Winner, E. (1979). New names for old things: the emergence of metaphoric language. *Journal of Child Language*, 6(3), 469–491. <https://doi.org/10.1017/S030500090000249X>
- Winner, E. (1997). *The point of words: Children's understanding of metaphor and irony*. Harvard University press.
- Winner, E., McCarthy, M., & Gardner, H. (1980). The ontogenesis of metaphor. In P. R. Honeck & R. R. Hoffman (Eds.), *Cognition and figurative language* (pp. 341–361). Hillsdale.
- Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12(4), 289–297. <https://doi.org/10.1037/0012-1649.12.4.289>
- Winnicott, D. W., Winnicott, C. E., Shepard, R. E., & Davis, M. E. (1989). *Psycho-analytic explorations*. Harvard University Press.
- Winter, B., & Strik-Lievers, F. (2023). Semantic distance predicts metaphoricity and creativity judgments in synesthetic metaphors. *Metaphor and the Social World*, 13(1), 59–80. <https://doi.org/10.1075/msw.00029.win>
- Witztum, E., van der Hart, O., & Friedman, B. (1988). The use of metaphors in psychotherapy. *Journal of Contemporary Psychotherapy*, 18(4), 270–290. <https://doi.org/10.1007/BF00946010>
- Wojcik, E. H. (2018). The Development of Lexical–Semantic Networks in Infants and Toddlers. *Child Development Perspectives*, 12(1), 34–38. <https://doi.org/10.1111/cdep.12252>
- Wojcik, E. H., & Kandhadai, P. (2020). Paradigmatic associations and individual variability in early lexical-semantic networks: Evidence from a free association task. *Developmental Psychology*, 56(1), 53–69. <https://doi.org/10.1037/dev0000844>
- Wood, S. N. (2017). *Generalized Additive Models*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>
- Wood, S. N. (2023). *Package 'mgcv'* (R package version 1.9-1).
- Yaffe, Y. (2023). Is Authoritative Parenting the Dominant Style in the Contemporary Western Family? A Report on a Cross-Cultural Israeli Sample. *Psychological Studies*, 68(1), 24–32. <https://doi.org/10.1007/s12646-022-00699-y>
- Yoder, P. J., & Warren, S. F. (1998). Maternal Responsivity Predicts the Prelinguistic Communication Intervention That Facilitates Generalized Intentional Communication. *Journal of Speech, Language, and Hearing Research*, 41(5), 1207–1219. <https://doi.org/10.1044/jslhr.4105.1207>
- Zhu, R., & Gopnik, A. (2023). Preschoolers and Adults Learn From Novel Metaphors. *Psychological Science*, 34(6), 696–704. <https://doi.org/10.1177/09567976231165267>
- Zurru, E. (2024). Communicating the urgency of the climate emergency through verbal and non-verbal metaphors. Ecological Communication and Ecoliteracy: Discourses of Awareness and Action for the Lifescape. In M. Bartoluzzi & E. Zurru (Eds.), *Ecological communication and ecoliteracy: discourses of awareness and action for the lifescape* (pp. 88–111). Bloomsbury Academic.

## APPENDIX

Below are reported some spontaneous metaphor productions that occurred during testing and training sessions.

<b>Metaphor</b>	<b>Age</b>	<b>Note</b>	<b>Context</b>
“So he is a rabbit”	5		Wellman and Liu scale (Instructions)
“I should be a broken machine”	5		Digit Span Forward (Instructions)
“Today I am a carrot”	7		Debriefing
“Happiness is running free as a dog”	7	Diagnosed with neurodevelopmental disorder	Spontaneous sharing
“This is not a court”	8		Guided-play session
“She has a computer in her mind”	9		Digit Span Backward (Peer performance)
“Pandas are parachutes, saving them means saving the earth”	9	Bilingual	Training session
“Work together to do...a chain”	9	Diagnosed with dyslexia	Training session
“Climate change is like a pencil, we have made mistakes, but we can still erase them!”	9		Training session
“Thank you! You are getting me out of hell”	10	Diagnosed with dyscalculia	On the way from the classroom to the experimental room

