LANGUAGE DEVELOPMENT RESEARCH

An Open Science Journal

Volume 5 | Issue 2 | November 2025
ISSN 2771-7976

About the Journal

Language Development Research: An Open-Science Journal was established in 2020 to meet the field's need for a peer-reviewed journal that is committed to fully open science: LDR charges no fees for readers or authors, and mandates full sharing of materials, data and analysis code. The intended audience is all researchers and professionals with an interest in language development and related fields: first language acquisition; typical and atypical language development; the development of spoken, signed or written languages; second language learning; bi- and multilingualism; artificial language learning; adult psycholinguistics; computational modeling; communication in nonhuman animals etc. The journal is managed by its editorial board and is not owned or published by any public or private company, registered charity or nonprofit organization.

TalkBank System

Language Development Research is the official journal of the **TalkBank system**, comprising the CHILDES, PhonBank, HomeBank, FluencyBank, Multilingualism and Clinical Banks, the CLAN software (used by hundreds of researchers worldwide to analyze children's spontaneous speech data), and the Info-CHILDES mailing list, the de-facto mailing list for the field of child language development with over 1,600 subscribers.

Diamond Open Access

Language Development Research is published using the Diamond Open Access model (also known as "Platinum" or "Universal" OA). The journal does not charge users for access (e.g., subscription or download fees) or authors for publication (e.g., article processing charges).

Hosting

The **Carnegie Mellon University Library Publishing Service** (LPS) hosts the journal on a Janeway Publishing Platform with its manuscript management system (MMS) used for author submissions.

License

Language Development Research is published by TalkBank and the Carnegie Mellon University Library Publishing Service. Authors retain the copyright to their published content. This work is distributed under the terms of the **Creative Commons Attribution-Noncommercial 4.0 International license** (https://creativecommons.org/licenses/by-nc/4.0/), which permits any use, reproduction and distribution of the work for noncommercial purposes with no further permissions required provided the original work is attributed as specified under the terms of this Creative Commons license.

Peer Review and Submissions

All submissions are reviewed by a minimum of two peer reviewers, and one of our Action Editors, all well- established senior researchers, chosen to represent a wide range of theoretical and methodological expertise. Action Editors select peer reviewers based on their expertise and experience in publishing papers in the relevant topic area.

Submissions and Publication Cycle

We invite submissions that meet our criteria for rigour, without regard to the perceived novelty or importance of the findings. We publish general and special-topic articles ("Special Collections") on a rolling basis to ensure rapid, cost-free publication for authors.

Language Development Research is published once a year, in December, with each issue containing the articles produced over the previous 12 months. Individual articles are published online as soon as they are produced. For citation purposes, articles are identified by the year of first publication and digital object identifier (DOI).

Editors in Chief

Patricia J. Brooks, City University of New York, USA **Vera Kempe,** Abertay University, UK	Email: ldr-journal@andrew.cmu.edu

Action Editors

Erin Conwell North Dakota State University, USA	**Natalia Gagarina** Leibniz-Zentrum Allgemeine Sprachwissenschaft, Germany
Josh Hartshorne MGH Institute of Health Professions, USA	**Victoria Knowland** Newcastle University, UK
Ingrid Lossius Falkum University of Oslo, Norway	**Katherine Messenger** Lancaster University, UK
Monika Molnar University of Toronto, Canada	**Alan Rumsey** Australian National University, Australia
Erik Thiessen Carnegie Mellon University, USA	**Daniel Walter** Emory University, USA

Founders

Ben Ambridge University of Manchester, UK	**Brian MacWhinney** Carnegie Mellon University, USA

Head of Editorial Board

Danielle Matthews, University of Sheffield, UK

Editorial Board

Javier Aguado-Orea Sheffield Hallam University, UK	**Ben Ambridge** Manchester University, UK
David Barner University of California, San Diego, USA	**Dorothy Bishop** University of Oxford, UK
Arielle Borovsky Purdue University, USA	**Ana Castro** Universidade NOVA de Lisboa, Portugal
Jean-Pierre Chevrot Université Grenoble Alpes, France	**Beatriz de Diego-Lázaro** Universidad de Valladolid, Spain
Seamus Donnelly Australian National University, Australia	**Ruthe Foushee** New School for Social Research, USA
Michael C. Frank Stanford University, USA	**Steven Gillis** Universiteit Antwerpen, Belgium
Lisa Hsin University of Alabama, USA	**Sam Jones** University of Lancaster, UK
Jeff Lidz University of Maryland, USA	**Weiyi Ma** University of Arkansas, USA
Brian MacWhinney Carnegie Mellon University, USA	**Monique Mills** University of Houston, USA
Toby Mintz University of Southern California, USA	**Anat Prior** University of Haifa, Israel
Kirsten Read Santa Clara University, USA	**Tom Roeper** University of Massachusetts, Amherst, USA
Caroline Rowland Max Planck Institute for Psycholinguistics, Netherlands	**Melanie Soderstrom** University of Manitoba, Canada
Virve-Anneli Vihman Tartu Ulikooli, Estonia	**Erik Thiessen** Carnegie Mellon University, USA
Frank Wijnen Utrecht University, Netherlands	**Tania Zamuner** University of Ottawa, Canada

Table of Contents

Volume 5, Issue 2, November 2025

Special Issue: Development of Metaphor Comprehension

Special Issue Editors: Ingrid Lossius Falkum and Mary Beth Neff

University of Oslo, Norway

Metaphor comprehension in preschool children

Ariel Starr
Taylor Petersen
Sahana Sridhar
University of Washington, USA

Abstract: Recent work suggests that though young children can comprehend metaphors based on shared perceptual or functional features of objects, comprehending metaphors based on abstract relations across domains presents a greater challenge. We conducted two pre-registered studies (n = 272; mean age = 3.77 years; 143 female) to investigate children's ability to understand metaphors based on object and abstract similarities. We also assessed how children's language learning environments (monolingual or bilingual) relate to their metaphor comprehension. Children were successful in understanding both types of metaphors. In addition, monolingual and bilingual children were equally proficient in metaphor comprehension. These findings highlight the sophisticated ways that preschool-aged children can use their rapidly developing lexicons.

Keywords: metaphor comprehension, bilingualism, language development

Corresponding author: Ariel Starr, Department of Psychology, University of Washington, Seattle, WA 98195, USA. Email: abstarr@uw.edu.

ORCID ID(s): https://orcid.org/0000-0003-0433-5003 (Starr, A.); https://orcid.org/0009-0000-6272-0191 (Sridhar, S.)

Citation: Starr, A., Petersen, T., & Sridhar, S. (2025). Metaphor comprehension in preschool children. *Language Development Research*, 5(2), 1–27. http://doi.org/10.34842/starr

Introduction

Metaphors are a powerful tool for expression. By highlighting similarities between unrelated domains, metaphors drive creativity in artistic and scientific fields and shape how we think about the world around us (Bowdle & Gentner, 2005; Kuhn, 1993; Lakoff & Johnson, 1980). However, understanding metaphors requires understanding figurative language, which can be difficult for young children. Adults can rest reasonably assured that when someone is described as a "busy bee," there are no actual bees present. To young children, however, this statement might cause confusion (Vosniadou, 1987). How do children come to be able to understand, and even appreciate, metaphors?

A classic view has been that preschool-aged children do not understand metaphors and instead interpret all language literally (e.g., Silberstein et al., 1982; Winner et al., 1976, 1980). However, as many have noted (e.g., Gentner, 1988; Pouscoulous, 2011; Vosniadou et al., 1984; Winner et al., 1976), the literature is rife with examples of children seemingly producing metaphorical language (e.g., an 18-month-old referring to a toy car as a "snake" while twisting it up and down his mother's arm (Winner et al., 1979)). While a more charitable interpretation of these observations might be that young children produce metaphors before they are able to comprehend metaphorical languages produced by others, these instances were often interpreted as overextension or pretend play rather than use of metaphorical language. However, early studies used metaphorical language and concepts that are unlikely to be familiar to young children (Winner et al., 1976; see Winner et al., 1980 for a detailed discussion of metaphor task demands). As a result, these studies may have mischaracterized children's understanding of metaphorical language (Vosniadou, 1987). When young children are presented with metaphors that use familiar language and concepts, laboratory studies do reveal early metaphoric competence. For example, by age 5, children can use toys to enact metaphorical phrases like "Billy was a squirrel burying the nuts" (correctly interpreted as meaning that Billy was hiding cookies rather that literally burying nuts; Vosniadou et al., 1984) and provide explanations of metaphorical phrases like "plant stems are drinking straws" (Gentner, 1988).

In addition to ensuring that concepts are age-appropriate, investigating children's metaphorical understanding also requires recognizing that metaphors can take many different forms. Gentner (1988) argued that metaphors can be divided into the partially overlapping categories of attributional metaphors (those based on perceptual similarities), relational metaphors (those based on a common relational structure, such as function), double metaphors (those that involve both perceptual and relational similarities), and complex metaphors (those that are idiosyncratic). Gentner found that children as young as 5 were able to understand attributional metaphors, and a preference for relational metaphors increased with age (1988). These results let to the hypothesis that the ability to understand metaphors based on shared perceptual properties emerges earlier than the ability to understand metaphors based on shared relations.

Recent studies using simplified paradigms have demonstrated that even preschool-aged children can understand metaphorical language when it involves familiar object concepts (Pouscoulous & Tomasello, 2020; Zhu et al., 2024; Zhu & Gopnik, 2023).In Pouscoulous and Tomasello (2020), for example, 3-year-old participants were presented with two toys, and an experimenter asked the child for one of the toys using a novel attributional metaphorical phrase structured in the form of “the X with the Y,” where X refers to the object and Y refers to a figurative perceptual attribute. The metaphorical phrase “the dog with the brown shoes,” for instance, referred to a toy dog with brown feet. In this simplified task, 3-year-olds succeeded in choosing the toys that corresponded to the metaphorical phrases. The metaphorical phrases used by Pouscoulous and Tomasello (2020) involved metaphors that evoked shared perceptual features and common knowledge about objects. In the example metaphor “the dog with the brown shoes,” children can use their knowledge of shoes (i.e., they go on feet) to solve the metaphor and match the phrase to a picture of a dog with brown feet. These results therefore extend Genter’s 1998 findings down to younger children and demonstrate that the ability to understand attributional metaphors is present by age three.

In addition to succeeding on attributional metaphors based on perceptual similarities, recent studies have shown that preschoolers can also comprehend relational metaphors that are based on shared relations between objects, namely object functions (Zhu et al., 2024; Zhu & Gopnik, 2023). Zhu and colleagues (2024) found that children as young as 3 could understand relational metaphors that make comparisons between attributes of two objects that serve the same purpose (e.g., “pools are bowls,” which makes use of the fact that both pools and bowls hold liquids). Three-year-olds could also use their object knowledge to make novel metaphorical inferences (Zhu & Gopnik, 2023). For example, after hearing the phrase “Blickets are eyes,” 3-year-olds decided that blickets would be more likely to help you see things than help you chop things. These recent studies demonstrate that preschoolers are adept at comprehending metaphors based on attributional and functional features of familiar objects.

In contrast, young children have more difficulty with metaphors that involve abstract properties without direct physical or perceptual correlates, such as the metaphorical motion of time (e.g., “time *flew* by”) (Özçalışkan, 2005; Özçalışkan, 2007; Stites & Özçalışkan, 2012). Özçalışkan found that while 3-year-old children performed at chance in understanding these abstract metaphors, 4- and 5-year-old children understood their meanings. By age 5, children could also provide verbal explanations of the concepts represented in metaphorical motion. Taken together, the full picture of prior work on metaphor comprehension in preschool-aged children suggests that comprehension may development gradually, with understanding of metaphors based on perceptual and functional similarities developing earlier than understanding of abstract metaphors (Gentner, 1988; Vosniadou, 1987).

The main goal of the present study was to further investigate preschoolers' ability to comprehend abstract metaphors. We use the term abstract metaphor to refer to metaphors that involve a conceptual mapping between two unrelated domains (Casasanto, 2009; Lakoff & Johnson, 1980) instead of highlighting a shared perceptual or functional feature of a physical object. An example of an abstract metaphor is "the girl is feeling *down*," which involves mapping emotional states to spatial positions. When making sense of metaphors based on perceptual or functional similarities, children are making mappings between physical objects (e.g., shoes and feet). By contrast, abstract metaphors require making associations between two unrelated concepts (e.g., spatial locations and emotional valence) or mapping cross-modally across different sensory perceptions (e.g., mapping visual brightness to auditory volume, as in describing a bright color as *loud*). Although prior work by Özçalışkan and colleagues (2005, 2007, 2012) has indicated that abstract metaphors can be difficult for young children, we attempted to facilitate preschoolers' understanding by providing visual illustrations to go with the verbal metaphors. With this simplified presentation, we predicted that preschool-aged children would be able to understand abstract metaphors that involved familiar words and concepts, though they might find these metaphors more difficult than metaphors based on object knowledge.

A second goal of the present study was to investigate how children's language learning environments influence the development of metaphor comprehension skills. In addition to pragmatic and conceptual constraints that contribute to children's difficulties with metaphorical language, metaphorical language may also be difficult for young children to understand because they are resistant to assigning multiple labels to a single concept (Rubio-Fernández & Grassmann, 2016). In their study, Rubio-Fernández and Grassmann found that 3- and 4-year-old children were more successful at choosing the object pair that matched a metaphorical phrase when the target of the metaphor was a toy block compared to when it was a familiar object (e.g., "the train with a hat" with the hat represented by either a block or a toy dog). Children's difficulty with metaphorically extending a word to an object with a known label (e.g., referring to a dog as a hat) may stem from the mutual exclusivity heuristic. When young children are learning new words, they frequently operate under the heuristic that each concept has a single label and therefore, if a concept already has a known label, a newly introduced word likely labels a new concept (Markman & Wachtel, 1998). The mutual exclusivity heuristic has been documented in children as young as 17 months, who will preferentially look at a novel object compared to a familiar object when they are presented with a novel object label (Halberda, 2003).

However, children's reliance on mutual exclusivity and their expectation that there is a one-to-one mapping between labels and concepts is shaped by their language learning environment (Houston-Price et al., 2010). Children who are learning multiple languages are frequently learning multiple labels for the same concepts. As might be expected based on their learning history, bilingual toddlers are less likely than monolingual toddlers to assume that a novel label refers to a novel object rather than a familiar object (Houston-Price et al., 2010; Repnik et al., 2021). Similarly, bilingual preschoolers are less likely to use mutual exclusivity to learn new adjectives in comparison to monolingual preschoolers (Groba et al., 2019), which demonstrates that the phenomenon is not specific to learning nouns. Some studies have also suggested that bilingual children have stronger metalinguistic skills in comparison to monolingual skills (Ben-Zeev, 1977; Diaz & Farrar, 2018). Metalinguistic awareness includes a range of abilities including phonological awareness and understanding of homonyms and synonyms. Understanding homonyms and synonyms requires children to flexibly switch between different meanings of a word depending on the context – an ability that is also relevant for understanding metaphors. Because bilingual children are more open to the possibility that known concepts can have more than one label and labels can refer to multiple concepts, they might have an advantage in understanding metaphorical language. Although rates of bilingualism are increasing in the US and around the world (Luk, 2017), studies on metaphor comprehension in children have focused on monolingual children. To better understand different factors that influence the development of metaphor comprehension, the present study included monolingual and bilingual children.

In the present study, we compared monolingual and bilingual preschoolers' comprehension of two types of metaphors. Prior studies have shown that preschool-aged children can understand metaphors that involve shared perceptual and functional features of objects. We compared children's performance on these types of metaphors, which we term object metaphors, with performance on metaphors we term abstract metaphors, metaphors that involve children flexibly extending words across modalities and perceptual domains. We predicted that children would be able to understand both types of metaphors, but that they would be more successful in understanding perceptual metaphors compared to abstract metaphors. The second goal of the study was to explore how children's early language learning environments affect metaphor comprehension. We predicted that because of their regular exposure to second labels, bilingual children might be more successful in understanding both types of metaphorical language in comparison to monolingual children.

We tested these hypotheses in two pre-registered studies that used a metaphor comprehension paradigm adapted from Pouscoulous and Tomasello (2020). This paradigm involves matching a metaphorical phrase to one of two pictures, which may be an easier task for preschool-aged children than comprehending abstract metaphors presented in speech alone, as in Özçalışkan (2005; 2007). Children were also asked to select pictures matching the literal meaning of the target words. The vocabularies of young children typically contain more words for concrete nouns relative to words for more abstract concepts and more nouns in general relative to adjectives and verbs (L.

Bloom et al., 1975; Braginsky et al., 2019; Casasanto et al., 2010; Gentner, 1978). Furthermore, children's understanding of concrete word meanings precedes their understanding of more abstract meanings (Andrews & Madeira, 1977; P. Bloom, 2000; Braginsky et al., 2019; Casasanto et al., 2010; Clark, 1973; Gentner & Boroditsky, 2001). We predicted that children would need to understand the literal meaning of words in order to extend them metaphorically. Prior work has implicitly built off this assumption by testing children's knowledge of the literal meanings of the target word stimuli and ensuring that children understood the majority of the words (Pouscoulous & Tomasello, 2020), assuming that children would know the meaning of the target words without explicitly testing (Zhu et al., 2024), or circumnavigating this issue by using novel words as targets (Zhu & Gopnik, 2023). In Experiment 1, we opted to test children's understanding of the literal meaning of words and set a threshold performance level to include children in our analyses, as in Pouscoulous and Tomasello (2020). However, this method involved analyzing some metaphor trials for which children had not succeeded in identifying the literal meaning of the target word. In Experiment 2, we opted to only analyze metaphor comprehension trials for which children answered the corresponding vocabulary question correctly. Experiment 1 was conducted remotely with an experimenter present. Experiment 2 was a conceptual replication to test the robustness of the Experiment 1 results and was conducted remotely in an unmoderated format.

Study 1

Method

Participants

Data from 123 participants aged 2.5- to 4-years-old (mean = 3.6 years, range = 2.5-4.8, 64 female) were included in the final analyses. Participants were classified as monolingual or bilingual based on their language learning environments as described by their caregivers. Parents were asked if their child was regularly exposed to another language besides English, and if the answer was yes, to estimate the proportion of time the child was exposed to each language. Participants who were exposed to English 90% of the time or more were considered monolingual (n = 74) (Byers-Heinlein, 2017; Jardak & Byers-Heinlein, 2019). Participants who were exposed to English and a second language at least 20% of the time were considered bilingual (n = 49; see Supplement for list of second languages spoken by participants and for the language exposure survey). We preregistered a sample size of 50 monolingual children and 50 bilingual children, but slightly more monolingual children were included in the analyses as we continued recruiting participants until we had reached 50 bilingual children. However, after data collection was stopped, we realized that one child was initially incorrectly classified as bilingual. Data from an additional 70 participants were excluded for performing below 70% on the vocabulary test (n = 34), not fitting into one of our language categories (e.g., regularly exposed to more than two languages;

n = 14), caregiver interference (n = 5), consistently choosing one character over the other (n = 4), parent-reported developmental disabilities (n = 3), distractedness (n = 3), performing below 66% on the attention checks (n = 2), unclear parental language exposure reports (n = 2), failing to complete the experiment (n = 2) or for poor internet connection during the session (n = 1). The University of Washington Institutional Review Board approved this study. Parents of the participants consented for their child's participation in the study and consented for video and audio recording of the study. Parents of participants were compensated with a $5 electronic Tango gift card.

Out of all families who participated in Experiment 1, 160 filled out our optional demographic questionnaire. Because demographics information was collected anonymously, we are unable to report the demographics separately for our monolingual and bilingual groups. Parents identified participants as 47% White, 26% multiracial, 16% Asian, 4% Hispanic or Latine, 0.6% Black or African American, 0.6% American Indian or Alaska Native, 3% another option not listed, and 2.5% chose not to respond. With respect to annual household income, 43% of caregivers reported less than $100,000, 40% reported $100,000-$200,000, 2% reported more than $200,000, and 15% chose not to respond. Data were collected between December 2020 and October 2021.

Procedure

Children were tested by an experimenter over Zoom while their parent sat nearby. Parents were asked to use a laptop or desktop computer to ensure that the stimuli were displayed similarly across participants, and each parent was sent a Zoom user guide prior to the study date. During the experimental session, children played a game in which they chose which of two pictures best matched a metaphorical expression provided by the experimenter. After children completed the experiment, parents completed a questionnaire about their child's language environment. Altogether, the session took approximately 15 minutes.

Materials and Design

Because it is difficult to discern the referent of children's pointing responses over Zoom, each target picture had a small picture of either Cookie Monster or Elmo in a top corner, and children were asked to verbally indicate whether Cookie Monster or Elmo had the picture that best matched the metaphorical phrase. The game began with two warm-up questions to familiarize children with the procedure. The experimenter asked the participant, "Whose picture has a red dot, Elmo or Cookie Monster?" and one character had a red dot below them whereas the other character did not. For the practice questions, the experimenter corrected the child if they were incorrect, but only neutral feedback was provided for the test questions. Children also completed three attention check questions interspersed with the test questions that did not involve metaphorical language (e.g., "Whose picture has the chair?"). Children needed to answer at least two out of three attention check questions correctly for their data to be included in the final analyses.

We created ten novel metaphors that involved concepts and language familiar to preschool-aged children. Five of the metaphors were classified as object metaphors, which were based on shared, visually perceptible features of objects (e.g., a java jacket for "the cup with the sweater") and were similar to those used in Pouscoulous & Tomasello (2020). The other five metaphors were classified as abstract metaphors, which were based on cross-modal mappings rather than physical features (e.g., a girl frowning for "the girl is feeling down"). Participants answered all ten metaphor questions in a randomized order. There were two versions of the experiment, A and B, which were identical except for the abstract metaphor questions involved opposite pairs of words. For example, version A asked about,"a girl who is feeling down," and version B asked about,"a girl who is feeling up." Participants were randomly assigned to a version. See Figure 1 for examples of the stimuli and Table 1 for a list of all metaphors. The mean ages of acquisition (Kuperman et al., 2012) for the target object and abstract metaphor words were not statistically different ($t(10.7) = 0.42$, $p = .68$). Adult English-speaking participants (N = 23) completed a pilot experiment to confirm that the novel metaphors were reliably associated with the target images. This pilot experiment used the same format and questions as the child experiment, except that participants read the questions on their own without an experimenter present. Adult participants chose the target image 99% of the time.

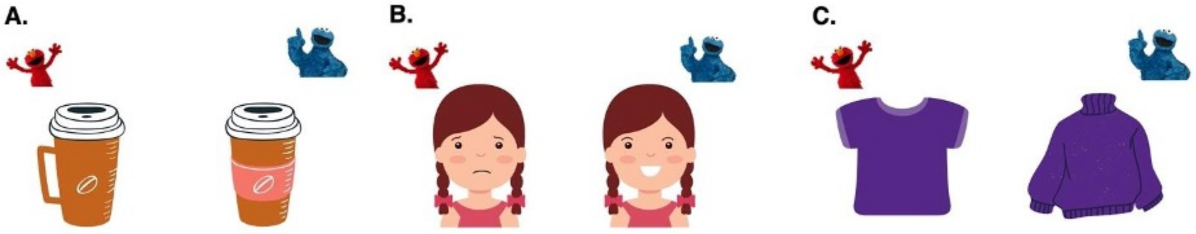


Figure 1. ***Example stimuli for the different types of experimental questions. (A) Object metaphor stimuli for the question "whose picture has the cup with a sweater, Elmo or Cookie Monster?" (B) Abstract metaphor stimuli for the question "whose picture has a girl who is feeling down, Elmo or Cookie Monster?" (C) Vocabulary stimuli for the question "whose picture has the sweater, Elmo or Cookie Monster?***

After the metaphor questions, children completed ten vocabulary questions that assessed their knowledge of the literal meaning of the words involved in the metaphor questions (e.g., "sweater" and "down" from the prior examples). These questions had the same format as the metaphorical questions (e.g., "Whose picture has the sweater?" with response options of a sweater versus a t-shirt). Because we assumed that children needed to understand the literal meaning of a word to be able to metaphorically extend it, and because children may also guess randomly if they are not engaged with the game, only children who selected the correct picture for at least eight out of the ten vocabulary questions were included in the final analyses. For all metaphor and vocabulary questions, only neutral feedback (e.g., "thank you!") was provided.

Table 1. ***Mean metaphor and vocabulary and comprehension accuracy by item.***

Question "Which picture has the..."	**Version**	**Question Type**	**Mean Metaphor Accuracy [95% CI]**	**Mean Vocabulary Accuracy [95% CI]**
Bottle with a big belly		Object	0.75 [0.64-0.86]	0.90 [0.77-1]
Cup with a sweater		Object	0.83 [0.71-0.95]	0.88 [0.75-1]
Dog with socks		Object	0.80 [0.68-0.92]	0.98 [0.83-1]
Horse with a backpack		Object	0.76 [0.65-0.87]	0.98 [0.83-1]
House with eyes*		Object	0.62 [0.53-0.71]	0.97 [0.83-1]
Bird that is happier*	A	Abstract	0.55 [0.44-0.66]	0.98 [0.78-1]
Bird that is sadder*	B	Abstract	0.56 [0.44-0.68]	0.95 [0.75-1]
Boy who is having a bumpy day	A	Abstract	0.77 [0.61-0.93]	0.91 [0.72-1]
Boy who is having a smooth day	B	Abstract	0.66 [0.52-0.80]	0.76 [0.60-0.92]
Dress that is loud*	A	Abstract	0.62 [0.49-0.75]	0.83 [0.66-1]
Dress that is quiet*	B	Abstract	0.54 [0.42-0.66]	0.86 [0.68-1]
Girl who is feeling down*	A	Abstract	0.72 [0.57-0.87]	0.95 [0.75-1]
Girl who is feeling up*	B	Abstract	0.59 [0.46-0.72]	0.92 [0.72-1]
TV that is sick*	A	Abstract	0.58 [0.46-0.70]	0.97 [0.77-1]
TV that is healthy*	B	Abstract	0.58 [0.46-0.70]	0.93 [0.73-1]

Note. Asterisks denote questions with low performance that were replaced in Experiment 2.

Data Analysis

The preregistration and all experimental materials, data, and code required to reproduce the analyses for this study are publicly available at https://osf.io/ta8vf. Our final models deviated slightly from the preregistered models in that models with the full random effects structure did not converge, so we used models with a simplified random effects structure. Data cleaning, analyses, and visualizations were performed in R using the tidyverse and lmerTest packages (Kuznetsova et al., 2017; Wickham, 2014).

Results

Main Analyses

To determine how metaphor type and language learning environment influence children's metaphor comprehension, we ran a pre-registered logistic mixed effects model predicting trial accuracy with metaphor type (object or abstract), language learning environment (monolingual or bilingual), the interaction between metaphor type and language group, and question version (A or B) as fixed effects, along with random intercepts for subject and item. We found a borderline effect of metaphor type ($B = 0.49$, $p = .058$) but no significant effect of language environment ($B = -0.32$, $p = .085$) or interaction between metaphor type and language environment ($B = 0.29$, $p = .26$). Children matched the metaphorical phrase to the correct picture at above chance levels for both types of metaphors ($ts > 6.1$, $ps < 0.001$). Performance on object metaphor questions was higher than performance on abstract metaphor questions (object: $M = 0.75$, $SE = 0.02$; abstract: $M = 0.62$, $SE = 0.02$; Figure 2A). In addition, bilingual children demonstrated descriptively higher metaphor comprehension than monolingual children (bilingual: $M = 0.71$, $SE = 0.02$; monolingual: $M = 0.67$, $SE = 0.02$; Figure 2B).

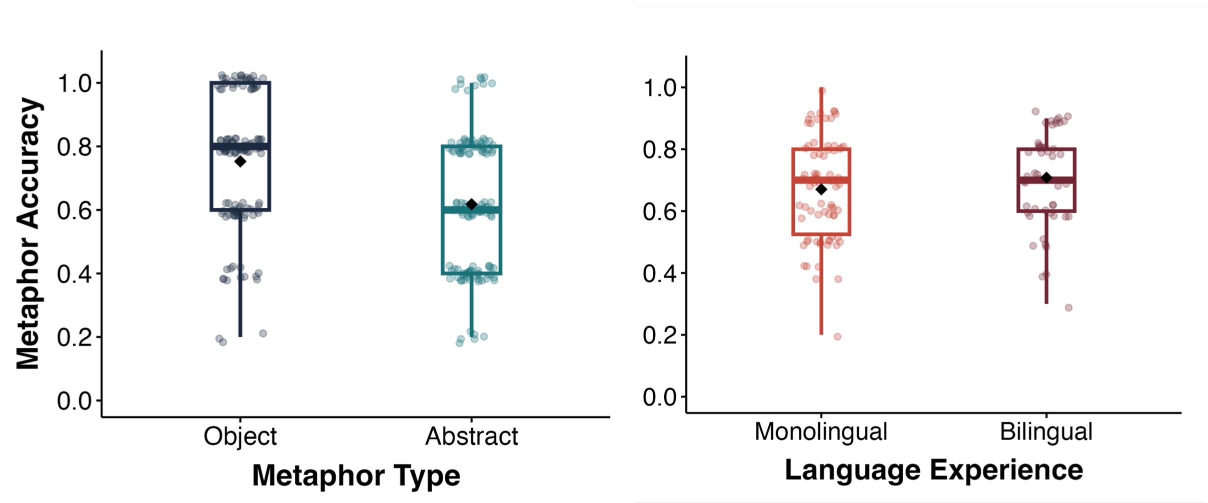


Figure 2. ***Experiment 1 metaphor comprehension accuracy by metaphor type (A) and language experience (B). Diamonds indicate group means.***

Vocabulary Knowledge

Because children needed to score above 70% on their understanding of the literal meaning of the words used in the metaphorical phrases to be included in the final analyses, children's vocabulary comprehension was very strong, and vocabulary questions were answered with an average accuracy of 93%. A linear model predicting vocabulary accuracy found that language learning environment was not a significant predictor of accuracy (B = 0.005, p = 0.74; monolingual children: M = 0.93, SE = 0.01; bilingual children: M = 0.92, SE = 0.01). However, participants scored significantly better on the vocabulary questions that corresponded to object metaphor questions compared to vocabulary questions that corresponded to abstract metaphor questions (B = 0.04, p = 0.007; object: M = 0.95, SE = 0.01; abstract: M = 0.91, SE = 0.01). As planned in our pre-registration, to determine how differences in vocabulary comprehension influence metaphor comprehension, we re-ran our original model predicting metaphor comprehension performance but analyzed only data from metaphor questions for which the corresponding vocabulary question was answered correctly. With this more stringent criterion, we found that there was a main effect of metaphor type (B = 0.62, p = .034), indicating higher performance on object metaphors compared to abstract metaphors, but the effect of language environment was still not significant (B = -0.31, p = .12).

Follow-up Analyses Investigating Age and Second Language Exposure

In a follow-up pre-registered analysis, we explored how metaphor comprehension was affected by age. When we included age as an additional continuous fixed effect, it was a significant positive predictor of performance (B = 0.42, p = .001), though again the effect of metaphor type was borderline (B = 0.49, p = .057), and the effect language learning environment was not significant (B = -0.30, p = .10). With age, children became increasingly successful in matching the metaphors to the correct image. In a second pre-registered follow-up model, we used percent exposure to a second language as a continuous predictor, rather than the dichotomous monolingual versus bilingual categorization. The outcome of this model again revealed a significant effect of age (B = 0.35, p = .007), as well as a significant effect of metaphor type (B = 0.70, p < .001). However, percent exposure to a second language was not a significant predictor (B = 0.11, p = .24). Note that this analysis involves a slightly smaller sample size (n = 112) due to some parents providing ranges of language exposure for their child rather than single values.

In addition to younger children having more difficulty with the metaphor questions, we also found that younger children were more likely to have their data excluded in comparison to older children. Of the 36 participants excluded for performance on vocabulary questions and attention check questions, 29 were under 3.5 years of age. Overall, 53% of all children under 3.5 were cut from further analyses due to poor performance on vocabulary and attention check questions. In a follow-up exploratory analysis, we ran a model using only data from children older than 3.5 years and found that metaphor type was a significant predictor of performance (*B* = 0.77, *p* = .009), but neither language learning environment (*B* = -0.34, *p* = .16) nor age (*B* = 0.45, *p* = .12) were significant predictors. Overall, these patterns of results indicate that children are able to comprehend both object and abstract metaphors, but additional data is needed to determine if abstract metaphors are more difficult. In addition, monolingual and bilingual children are equally proficient in comprehending metaphors.

Experiment 2

In Experiment 1, we tested whether objects metaphors are easier for preschool children to understand in comparison to abstract metaphors and whether children's language learning experience affects their metaphor comprehension. We found that children were successful in matching metaphorical expressions to the target picture for both object and abstract metaphors, but that abstract metaphorical phrases were potentially more difficult. We also found that children's proficiency with metaphor comprehension was not predicted by their language learning environment. In Experiment 2, we tested the robustness of these results with a conceptual replication conducted using Lookit (now called Children Helping Science; Scott & Schulz, 2017; Sheskin et al., 2020). Children Helping Science is a platform that enables families to participate in experiments over the web without an experimenter present. In addition to changing from a moderated to an unmoderated experiment format, we made a few other adjustments. First, we decided to focus on children aged 3.5 to 4.5 years (rather than 2.5 to 4.5 years) because these slightly older children are more likely to know the literal meaning of our metaphorical words and to stay engaged throughout the experimental session. In addition, we replaced the metaphorical expressions for which children performed at chance with novel metaphors and images to determine of the initial results were specific to the metaphorical phrases chosen. Finally, because children also performed better on the object vocabulary questions relative to the abstract vocabulary questions, we adopted a more stringent analysis strategy in which we only analyzed metaphor comprehension data for questions that children answered the corresponding vocabulary question correctly. This enabled us to confirm that differences in metaphor comprehension across metaphor types reflected metaphor comprehension itself versus comprehension of the literal meanings of the words.

Method

Participants

Data from 149 participants aged 3.5-4.5 years old (*M* = 3.89 years, *SD* = 0.32, 79 female) were included in the final analyses. As in Experiment 1, participants were classified as monolingual if they were exposed to English at least 90% of the time (*n* = 93) and classified as bilingual if they were exposed to a second language at least 20% of the time (*n* = 56). We preregistered a sample size of 66 monolingual children and 66 bilingual children. Sixty-six participants per language learning group gives us 90% power to detect a medium effect size for a paired samples *t*-test comparing metaphor accuracy between language experience groups (the effect size for a *t*-test comparing metaphor accuracy between the language experience groups in older children in Experiment 1 was $d = 0.57$). However, due to difficulties with recruiting bilingual participants, continued recruitment resulted in a larger sample of monolingual participants while still not being able to reach our target number of bilingual participants, which limits our statistical power. Caregivers of monolingual participants identified their children as 62% White, 27% multiracial, 10% Asian, and 1% Black or African American. With respect to annual household income, 33% of monolingual caregivers reported less than $100,000, 35% reported $100,000-$200,000, 25% reported more than $200,000, and 7% did not answer. Caregivers of bilingual participants identified their children as 36% Asian, 32% white, 23% multiracial, 5% Hispanic or Latine, and 4% chose not to respond. In terms of annual household income, 30% of bilingual caregivers reported less than 100k, 27% reported $100,000-$200,000, 30% reported more than $200,000, and 13% did not answer. The monolingual and bilingual language groups differed in terms of racial demographics ($X^2(3) = 23.73$, $p < .001$), but not in terms of household income ($X^2(2) = 1.21$, $p = .55$).

Data from an additional 36 participants were excluded for not fitting into our language criteria ($n = 13$), performing below 65% on the vocabulary questions ($n = 8$), not completing the study ($n = 7$), parent-reported developmental disabilities ($n = 3$), caregiver interference ($n = 3$), or performing below 66% on the attention checks ($n = 2$). The University of Washington Institutional Review Board approved of this study. Parents of the participants consented for their child's participation in the study and for video and audio recording of the study. Parents of participants were compensated with a $5 electronic Tango gift card. Data were collected between November 2021 and June 2023.

Procedure

This study was conducted entirely over Lookit (now called Children Helping Science; Scott & Schulz, 2017; Sheskin et al., 2020). Children completed the study with a parent, and there was no experimenter present. Parents read each question aloud to their child and clicked on the picture that their child chose. Because children could now point to their response, the Elmo and Cookie Monster images were removed from the stimuli. The procedure was otherwise identical to Experiment 1.

Materials and Design

The materials were identical to those used in Experiment 1, with the exception that five metaphor questions (one object and four abstract) in Experiment 1 for which children scored below 65% were replaced with new stimuli (Table 1). Because one of our main experimental questions is whether abstract metaphors are more difficult for children to understand than object metaphors, replicating our study with new metaphorical stimuli is critical for determining the robustness of our results. The mean ages of acquisition (Kuperman et al., 2012) for the target object and abstract metaphor words were not statistically different ($t(10.5) = 0.70$, $p = .50$). The new metaphor stimuli were validated in a sample of English-speaking adults (N = 30) who chose the target image 100% of the time. However, initial analyses revealed that children's performance on one version of a new metaphor question ("which wave is weaker?") was significantly below chance, and this item (both the A and B versions) was therefore excluded from all subsequent analyses. The stimuli used in Experiment 2, along with children's accuracy on the metaphor and vocabulary version of the questions, are presented in Table 2.

Table 2

Mean Metaphor and Vocabulary Comprehension Accuracy by Question

Question "Which picture has the..."	**Version**	**Question Type**	**Mean Metaphor Accuracy [95% CI]**	**Mean Vocab Accuracy [95% CI]**
Bottle with a big belly		Object	0.88 [0.75-1]	0.87 [0.75-0.99]
Cup with a sweater		Object	0.80 [0.69-0.91]	0.89 [0.77-1]
Dog with socks		Object	0.91 [0.78-1]	0.95 [0.82-1]
Horse with a backpack		Object	0.81 [0.70-0.92]	0.97 [0.84-1]
Tree with hair		Object	0.78 [0.67-0.89]	0.97 [0.84-1]
Girl who is feeling up	A	Abstract	0.86 [0.69-1]	0.96 [0.77-1]
Girl who is feeling down	B	Abstract	0.66 [0.54-0.78]	1 [0.82-1]
Phone that is healthy	A	Abstract	0.63 [0.50-0.76]	0.99 [0.79-1]
Phone that is sick	B	Abstract	0.87 [0.69-1]	0.75 [0.61-0.89]
Wave that is stronger*	A	Abstract	0.97 [0.77-1]	0.94 [0.75-1]
Wave that is weaker*	B	Abstract	0.36 [0.27-0.45]	0.52 [0.42-0.62]
Boy who is having a bumpy day	A	Abstract	0.85 [0.67-1]	0.90 [0.72-1]
Boy who is having a smooth day	B	Abstract	0.68 [0.54-0.82]	0.75 [0.61-0.89]
Flower that is happy	A	Abstract	0.96 [0.77-1]	0.97 [0.78-1]
Flower that is sad	B	Abstract	0.93 [0.75-1]	0.95 [0.78-1]

Note. Asterisks denote items with mean metaphor comprehension accuracy below 50% for *version B. This question (both the A and B versions) was dropped from all subsequent analyses.*

Data Analysis

The preregistration for this study can be found at https://osf.io/ta8vf. Because we only planned to analyze metaphor questions for which children answered the corresponding vocabulary question correctly, we used a slightly more lenient vocabulary performance threshold of at least 65% correct (rather than 75% in Experiment 1). On average, children contributed 8.3 metaphor questions to the analyses. As in Experiment 1, our final statistical models deviated slightly from the preregistered models, and we used models with the maximal random effects structure that did converge. Data cleaning, analyses, and visualizations were performed in R using the tidyverse and lmerTest packages (Kuznetsova et al., 2017; Wickham et al., 2019).

Results

Main Analyses

As in Experiment 1, we ran a pre-registered logistic mixed effects model predicting trial accuracy with metaphor type (object or abstract), language learning environment (monolingual or bilingual), the interaction between metaphor type and language environment, and question version (A or B) as fixed effects, and random intercepts for subject and item. This model revealed no significant main effects or interactions (metaphor: $B = 0.35$, $p = .43$; language environment: $B = 0.21$, $p = .43$; interaction: $B =-0.38$, $p = .24$). Children matched the metaphorical phrase to the correct picture at above chance levels for both types of metaphors, as they did in Experiment 1 ($ts > 16.28$, $ps < 0.001$). In contrast to the results of Experiment 1, however, performance on object metaphor questions was equivalent to performance on abstract metaphor questions (object: $M = 0.83$, $SE = 0.02$; abstract: $M = 0.80$, $SE = 0.02$; Figure 3A). In addition, bilingual children performed equivalently to monolingual children (bilingual: $M = 0.82$, $SE = 0.02$; monolingual: $M = 0.82$, $SE = 0.01$; Figure 3B).

In a follow-up model, we added age as an additional predictor and found that metaphor performance improved with age ($B = 0.94$, $p = .003$), but no other main effects or interactions were significant (Bs < 0.39, $ps > .23$). In a pre-registered follow-up model, we used percent exposure to a second language as a continuous predictor, rather than the dichotomous monolingual versus bilingual categories. This model used a slightly smaller sample size due to some parents providing ranges of second language exposure (n = 139). The outcome of this model again revealed only a significant effect of age ($B = 0.95$, $p = .006$), and percent exposure to a second language was not a significant predictor ($B = -0.04$, $p = .75$). These patterns of results suggest that children are equally proficient at understanding both object and abstract metaphors and metaphor comprehension improves with age, but language learning environment does not influence metaphor comprehension.

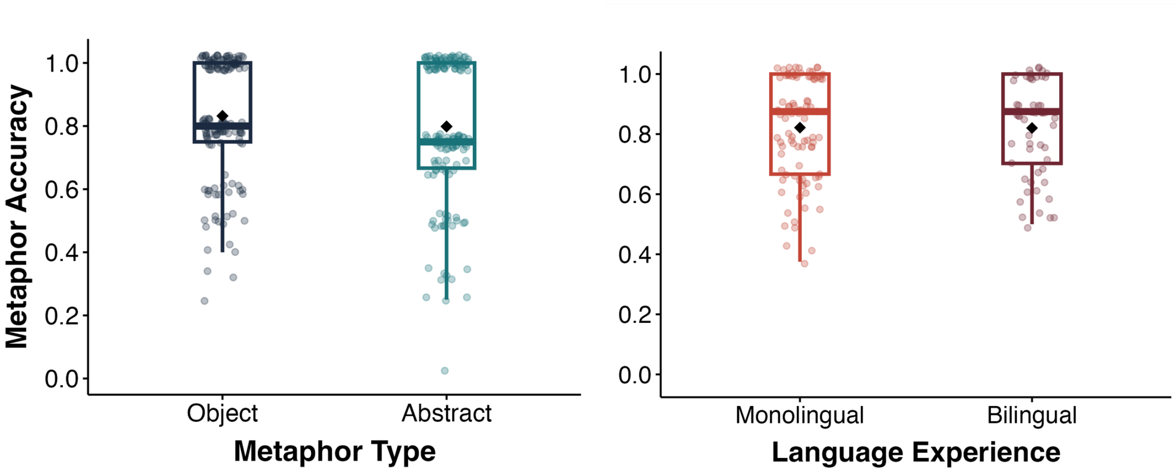


Figure 3. ***Experiment 2 metaphor comprehension accuracy by metaphor type (A) and language experience (B). Diamonds indicate group means.***

Vocabulary Knowledge

Children's vocabulary comprehension performance was slightly better for object metaphor stimuli compared to abstract metaphor stimuli ($B = 0.382$, $p = .068$; object: $M = 0.93$, $SE = 0.01$; abstract: $M = 0.91$, $SE = 0.01$). As a result, slightly more abstract than object metaphor questions were dropped due to missing the corresponding vocabulary questions in the original analysis for the two types of words (56 abstract questions versus 49 object questions). Vocabulary comprehension did not differ between monolingual and bilingual children ($B = 0.32$, $p = .191$; monolingual: $M = 0.93$, $SE = 0.01$; bilingual: $M = 0.91$, $SE = 0.01$).

Combined Analyses Using Data from Experiments 1 and 2

In a final exploratory analysis, we combined the data from Experiments 1 and 2 to see if increasing the sample size might reveal an effect of language learning environment. We used a logistic mixed effects model predicting trial accuracy with metaphor type, language learning environment, age, their interactions, and experiment as fixed effects, along with random intercepts for subject and item. For Experiment 1 data, we used data only from metaphor comprehension questions for which children answered the corresponding vocabulary question correctly and only from children aged 3.5 years and older to make it comparable to Experiment 2 data. The outcome of this model revealed a significant of age ($B = 0.84$, $p < .001$), indicating the metaphor performance improved with age. The effect of experiment was also significant ($B = 0.58$, $p < .001$), indicating the performance was higher in Experiment 2 compared to Experiment 1. No other main effects or interactions were significant (*Bs* < 0.50, *ps* > .13). Taken together, the results of Experiments 1 and 2 suggest that children are equally proficient at understanding object and abstract metaphors, but language learning environment does not influence metaphor comprehension.

General Discussion

The current study was designed to increase our understanding of metaphorical language comprehension in preschoolers in two key ways. First, given preschooler's success in comprehending metaphors based on shared perceptual and functional features of objects (Pouscoulous & Tomasello, 2020; Zhu et al., 2024; Zhu & Gopnik, 2023), we tested whether preschool children are also able to understand metaphors that do not rely on object knowledge. We compared children's performance on metaphors based on shared features of objects, which we term object metaphors, with their performance on abstract metaphors that involve extending words across modalities and conceptual domains. We predicted that children would be able to understand both types of metaphors, but that they would be more successful in understanding object metaphors compared to abstract metaphors. The second goal of the study was to explore how children's language learning environments affect metaphor comprehension, because the majority of work on children's metaphor comprehension has focused only on monolingual children. We predicted that because of their regular experience with mapping more than one label to the same concept, bilingual children might be more successful in understanding metaphorical language relative to monolingual children. We addressed these questions in a primary experiment (Experiment 1) and a conceptual replication (Experiment 2) to increase the strength of our conclusions.

Prior studies have shown that preschool-aged children can understand metaphors that involve mapping between shared perceptual and functional features – in other words, metaphors in which children can employ their knowledge of objects (Gentner, 1988; Pouscoulous & Tomasello, 2020; Vosniadou et al., 1984; Zhu et al., 2024; Zhu & Gopnik, 2023). On the other hand, when faced with abstract metaphors, such as those involving metaphorical motion (e.g., time *flies*), younger preschoolers are less likely to succeed, and comprehension increases between the ages of 3 and 5 years (Özçalişkan, 2005; Özçalışkan, 2007, Sites & Özçalışkan, 2012). These findings indirectly suggest that the ability to understand object metaphors may develop before the ability to understand abstract metaphors.

In the present study, we tested preschool children's comprehension of object and abstract metaphors. Across both experiments, we found that preschoolers were successful in comprehending both types of metaphors. Although in Experiment 1 we found some evidence that children might be more accurate in comprehending object metaphors, in Experiment 2, children were equally accurate in comprehending both types of metaphors. In Experiment 1, there were several metaphorical phrases for which children's comprehension accuracy was under 65%. In Experiment 2, we replaced these phrases. These new metaphorical phrases were apparently easier for children to comprehend, as evidenced by higher performance, which suggests that the disparate results across the two experiments is due to the specific stimuli used. Consistent with prior work (P. Bloom, 2000; Braginsky et al., 2019; Gentner & Boroditsky, 2001), we found that children's comprehension of the literal meaning of the target words was higher for the object metaphor words, which were concrete nouns, relative to their comprehension of the literal meaning of the target abstract metaphor words, which were adjectives. However, when children understood the literal meaning of the target words, they were able to understand metaphors that evoke similarities beyond shared properties and functions of objects. Taken together, the results of Experiments 1 and 2 suggest that preschool-aged children can comprehend not only metaphors based on shared object features and functions, but also metaphors based on more abstract relations.

To resolve object metaphors, children can use their knowledge about the objects involved. For example, when hearing the phrase, "the dog with the white socks," children can use their knowledge that socks go on feet to cue their attention to the dogs' feet in the two image choices. For abstract metaphors, there is no direct physical correlate that supports the metaphor. For example, when hearing the phrase, "the girl is feeling down," the word *down* is unlikely to automatically cue children's attention towards emotions. However, this association between vertical space and emotional valence is common in many cultures (Gibbs, 1994; Lakoff & Johnson, 1980). More broadly, there is a large body of work demonstrating that infants and young children are sensitive to a variety of cross-modal and cross-domain associations (e.g., Casasanto & Henetz, 2012; de Hevia et al., 2014; Dolscheid et al., 2014; Lourenco & Longo, 2010; Srinivasan & Carey, 2010; Starr & Srinivasan, 2018; Walker et al., 2010). Children's intuitions about these associations, for example associating a loud sound with visual brightness, likely guide their abstract metaphor comprehension performance. In addition, as demonstrated by the example "the girl is feeling *down*," many abstract metaphors involve a mapping between space and another domain. In English, for example, we use spatial proximity to describe similarity and personal relationships (e.g., "the two colors were a *close* match," "she is a *distant* relative") and use spatial positions and distances to represent time (e.g., "I'm looking *forward* to a vacation," "the meeting was *short*"). This pattern of using spatial words to describe more abstract phenomena is common across languages and cultures (Clark, 1973; Lakoff & Johnson, 1980; Sweetser, 1990; Talmy, 1988; Xu et al., 2017). With respect to vocabulary development, this pattern of metaphorically extending already-known words for concrete concepts (e.g., spatial words) to more abstract concepts (e.g., time) may help children solve the otherwise difficult mapping problem for learning labels for abstract concepts (Starr et al., 2021). Therefore, abstract metaphors might not only be understandable by young children, they might also help children learn about new concepts, similar to how preschool-aged children can learn from novel metaphors involving objects (Zhu & Gopnik, 2023).

Our abstract metaphors were relatively heterogenous and involved a variety of topics including cross-modal mappings, emotions, and the tenor of a day, and some of the metaphors were easier for children than others. Not surprisingly, previous studies have also found item-level variation in children’s metaphor comprehension across different metaphorical phrases of the same type and structure (e.g., Pouscoulous & Tomasello, 2020; Winner et al., 1980). We aimed to choose metaphors that involved concepts that would be familiar to preschool-aged children, but it is clear that some phrases were more difficult than others. An important direction for future work will therefore be to investigate different types of abstract metaphors. In particular, metaphors involving cross-modal mappings may be easier for children to understand than metaphors involving psychological targets (Winner et al., 1976). Because understanding metaphors requires making inferences about a speaker’s intent, it has been argued that understanding metaphors requires theory of mind (Happé, 1993). However, more recent work that has suggested that not all metaphors equally involve theory of mind, and theory of mind might be particularly relevant for understanding metaphors involving mental states (Lecce et al., 2019) and less involved in understanding metaphors based on object properties (Norbury, 2005). Additional work using a broader range of metaphorical types is therefore needed to better understand factors beyond metaphor type that contribute to the relative difficulty of understanding metaphorical language.

A second focus of the present study was to investigate how children’s early language learning environments influence metaphor comprehension. Although rates of bilingualism are increasing around the world, most studies of metaphor comprehension have focused on monolingual children. We hypothesized that because bilingual children are continuously learning that concepts can have more than one label, they may have an advantage in metaphor comprehension in comparison to their monolingual peers. However, our data do not support this hypothesis: monolingual and bilingual children were equally proficient at understanding both object and abstract metaphors. Although each of our experiments on their own might be underpowered to detect a small effect of language learning experience, we also found no significant effect when we collapsed across the two data sets to double the sample size. Critically, in neither experiment did bilingual children perform worse than their monolingual peers. This pattern of results suggests that even though bilingual preschoolers know fewer English words than their monolingual peers (Hoff et al., 2012), they are able to use the words they do know in complex ways.

For our primary analyses, we characterized children's language status as monolingual or bilingual and used parent report to decide to which category a child was assigned. Although this binary categorization is common in language development studies, it flattens the true experience of bilingualism (Byers-Heinlein et al., 2019; Rocha-Hidalgo & Barr, 2022). In follow-up analyses, we used children's percent exposure to a second language as a continuous predictor rather than using the categorical monolingual versus bilingual distinction. As with the models using the categorical classification, we again did not find that language learning environment significantly predicted metaphor comprehension. However, this continuous measure still fails to capture many meaningful differences in children's language learning environments, such as whether a child is learning two languages simultaneously or sequentially or what the sources of exposure are to each language. We also did not measure children's proficiency in either language beyond children's comprehension of target words in English. It is possible that with more nuanced measures of a child's language learning environment and proficiency, significant effects may be found or that effects may emerge with age as children's linguistic abilities develop further and children encounter other forms of metaphors. In addition, because children in these studies came from primarily middle- and upper-class families in the US, caution must be used in generalizing these findings to children growing up in other contexts.

The current work used a paradigm adopted from Pouscoulous and Tomasello (2020), in which children were asked to choose which of two pictures best matched a given metaphorical phrase. The paradigms used by Zhu and colleagues (2023, 2024) also involved visual depictions of metaphorical language. The use of visuals likely supports early metaphor comprehension and may contribute to the differences in metaphorical understanding found in these studies compared to those conducted by Özçalışkan (2005, 2007), which involved verbal descriptions only. An important direction for future work will be to investigate the contexts that facilitate versus hinder preschooler's metaphorical language comprehension. In addition, we used a two-alternative forced choice paradigm in which the distractor image was either an unrelated object feature (for object metaphors) or the opposite conceptual mapping (for abstract metaphors). To further understand children's early metaphor comprehension, researchers may consider asking children to choose between images depicting metaphorical and literal interpretations or ask them to explain their answers.

Metaphors are a powerful linguistic tool – they enable speakers to highlight similarities and make comparisons by flexibly reusing known words in novel contexts. Although metaphorical language can be difficult for young children, recent work has demonstrated that preschool-aged children can understand and even learn from novel metaphors (Pouscoulous & Tomasello, 2020; Zhu et al., 2024; Zhu & Gopnik, 2023). Here, we found that 3- and 4-year-old children can comprehend not just perceptual and functional metaphors involving objects, but also abstract metaphors that involve less tangible domains. These results contribute to a growing body of work demonstrating that preschool children are able to understand multiple types of non-literal language, including metaphors, metonyms, and irony (Di Paola et al., 2020; Falkum, 2022; Pouscoulous, 2023; Pouscoulous & Tomasello, 2020; Starr et al., 2021; Zhu et al., 2024, 2024). During the preschool years, as children are rapidly learning new words, they are also able to use these words in flexible and sophisticated ways.

References

Andrews, M. L., & Madeira, S. S. (1977). The Assessment of Pitch Discrimination Ability in Young Children. *Journal of Speech and Hearing Disorders, 42*(2), 279–286. https://doi.org/10.1044/jshd.4202.279

Ben-Zeev, S. (1977). The Influence of Bilingualism on Cognitive Strategy and Cognitive Development. *Child Development, 48*(3), 1009–1018. https://doi.org/10.2307/1128353

Bloom, L., Lightbown, P., Hood, L., Bowerman, M., Maratsos, M., & Maratsos, M. P. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development, 40*(2), 1. https://doi.org/10.2307/1165986

Bloom, P. (2000). *How children learn the meanings of words*. MIT Press.

Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review, 112*(1), 193–216. https://doi.org/10.1037/0033-295X.112.1.193

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and Variability in Children's Word Learning Across Languages. *OPEN MIND,* 16. https://doi.org/10.1162/opmi_a_00026

Byers-Heinlein, K. (2017). Bilingualism affects 9-month-old infants' expectations about how words refer to kinds. *Developmental Science, 20*(1), e12486. https://doi.org/10.1111/desc.12486

Byers-Heinlein, K., Esposito, A. G., Winsler, A., Marian, V., Castro, D. C., & Luk, G. (2019). The Case for Measuring and Reporting Bilingualism in Developmental Research. *Collabra: Psychology, 5*(1), 37. https://doi.org/10/gg6xjh

Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and time in the child's mind: Evidence for a cross-dimensional asymmetry. *Cognitive Science, 34*(3), 387–405. https://doi.org/10.1111/j.1551-6709.2010.01094.x

Casasanto, D., & Henetz, T. (2012). Handedness Shapes Children's Abstract

Concepts. *Cognitive Science, 36*(2), 359–372. https://doi.org/10/bp9jg4

Clark, H. H. (1973). Space, time, semantics, and the child. In *Cognitive development and the acquisition of language* (pp. 27–63). Academic Press.

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1323628111

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/gg7p8d

Diaz, V., & Farrar, M. J. (2018). The missing explanation of the false-belief advantage in bilingual children: A longitudinal study. *Developmental Science, 21*(4), e12594. https://doi.org/10.1111/desc.12594

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic Infants Are Sensitive to Space-Pitch Associations Found Across Cultures. *Psychological Science, 25*(6), 1256–1261. https://doi.org/10.1177/0956797614528521

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

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Monographs of the Society for Research in Child Development, 49*(4), 988. https://doi.org/10.2307/1128738

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*, 47–59. https://doi.org/10.2307/1130388

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge University Press. https://doi.org/10.1017/CBO9780511620669.010

Gibbs, R. W. (1994). *The Poetics of Mind*. Cambridge University Press.

Groba, A., De Houwer, A., Obrig, H., & Rossi, S. (2019). Bilingual and Monolingual First Language Acquisition Experience Differentially Shapes Children’s Property Term Learning: Evidence from Behavioral and Neurophysiological Measures. *Brain Sciences, 9*(2), 40. https://doi.org/10/ghv59w

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87*(1), B23–B34. https://doi.org/10/dbxw3d

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language Experience Shapes the Development of the Mutual Exclusivity Bias. *Infancy, 15*(2), 125–150. https://doi.org/10/d42m2d

Jardak, A., & Byers-Heinlein, K. (2019). Labels or Concepts? The Development of Semantic Networks in Bilingual Two-Year-Olds. *Child Development, 90*(2), e212–e229. https://doi.org/10/gjkzwt

Kuhn, T. S. (1993). Metaphor in science. In A. Ortony, *Metaphor and thought* (pp. 533–542). Cambridge University Press. https://doi.org/10.1017/CBO9781139173865.024

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

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., & Johnson, M. (1980). *Metaphors we live by*. The University of Chicago Press.

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

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science, 21*(6), 873–881. https://doi.org/10.1177/0956797610370158

Luk, G. (2017). Bilingualism. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *The Cambridge Encyclopedia of Child Development* (2nd ed., pp. 385–391). Cambridge University Press. https://doi.org/10.1017/9781316216491.062
Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Norbury, C. Frazier. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology, 23*(3), 383–399. https://doi.org/10.1348/026151005X26732

Özçalişkan, Ş. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language, 32*(2), 291–318. https://doi.org/10.1017/S0305000905006884

Özçalışkan, Ş. (2007). Metaphors We Move By: Children's Developing Understanding of Metaphorical Motion in Typologically Distinct Languages. *Metaphor and Symbol, 22*(2), 147–168. https://doi.org/10.1080/10926480701235429

Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics, 25,* 51–79. https://doi.org/10.1075/bjl.25.04pou

Pouscoulous, N. (2023). More than one path to pragmatics? Insights from children's grasp of implicit, figurative and ironical meaning. *Cognition, 240,* 105531. https://doi.org/10.1016/j.cognition.2023.105531

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.pragma.2019.05.021

Repnik, K. M., Chondrogianni, V., & Sorace, A. (2021). Linking disambiguation and retention in a developmental eye-tracking study with monolingual and multilingual children. *Journal of Experimental Child Psychology, 206,* 105072. https://doi.org/10.1016/j.jecp.2020.105072

Rocha-Hidalgo, J., & Barr, R. (2022). Defining bilingualism in infancy and toddlerhood: A scoping review. *International Journal of Bilingualism,* 13670069211069067. https://doi.org/10.1177/13670069211069067

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

Scott, K., & Schulz, L. (2017). Lookit (Part 1): A New Online Platform for Developmental Research. *Open Mind, 1*(1), 4–14. https://doi.org/10/gg84jq

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online Developmental Science to Foster Innovation, Access, and Impact. *Trends in Cognitive Sciences,* S1364661320301455. https://doi.org/10.1016/j.tics.2020.06.004

Silberstein, L., Gardner, H., Phelps, E., & Winner, E. (1982). Autumn leaves and old photographs: The development of metaphor preferences. *Journal of Experimental Child Psychology, 34*(1), 135–150. https://doi.org/10.1016/0022-0965(82)90036-4

Srinivasan, M., & Carey, S. (2010). The long and the short of it: On the nature and origin of functional overlap between representations of space and time. *Cognition, 116*(2), 217–241. https://doi.org/10.1016/j.cognition.2010.05.005

Starr, A., Cirolia, A. J., Tillman, K. A., & Srinivasan, M. (2021). Spatial Metaphor Facilitates Word Learning. *Child Development, 92*(3), e329–e342. https://doi.org/10.1111/cdev.13477

Starr, A., & Srinivasan, M. (2018). Spatial Metaphor and the Development of Cross-Domain Mappings in Early Childhood. *Developmental Psychology, 54*(10), 1822–1832. https://doi.org/10.1037/dev0000573

Stites, L. J., & Özçalışkan, Ş. (2012). Developmental changes in children's comprehension and explanation of spatial metaphors for time. *Journal of Child Language, 40*(05), 1123–1137. https://doi.org/10.1017/S0305000912000384

Sweetser, E. (1990). *From Etymology to Pragmatics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620904

Talmy, L. (1988). Force Dynamics in Language and Cognition. *Cognitive Science, 12*(1), 49–100. https://doi.org/10.1207/s15516709cog1201_2

Vosniadou, S. (1987). Children and Metaphors. *Child Development, 58*(3), 870–885. https://doi.org/10.2307/1130223

Vosniadou, S., Ortony, A., Reynolds, R. E., & Wilson, P. T. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Monographs of the Society for Research in Child Development, 55*(4), 1588.

https://doi.org/10.2307/1130028

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological Science*, *21*(1), 21–25. https://doi.org/10.1177/0956797609354734

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*.

Winner, E., Engel, M., & Gardner, H. (1980). Misunderstanding metaphor: What's the problem? *Journal of Experimental Child Psychology, 30*(1), 22–32. https://doi.org/10.1016/0022-0965(80)90072-7

Winner, E., McCarthy, M., Kleinman, S., & Gardner, H. (1979). First metaphors. *New Directions for Child and Adolescent Development*, *1979*(3), 29–41. https://doi.org/10.1002/cd.23219790305

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

Xu, Y., Malt, B. C., & Srinivasan, M. (2017). Evolution of word meanings through metaphorical mapping: Systematicity over the past millennium. *Cognitive Psychology, 96*, 41–53. https://doi.org/10.1016/j.cogpsych.2017.05.005

Zhu, R., Goddu, M. K., Zhu, L. Z., & Gopnik, A. (2024). Preschoolers' Comprehension of Functional Metaphors. *Open Mind: Discoveries in Cognitive Science*, *8*, 924–949. https://doi.org/10.1162/opmi_a_00152

Zhu, R., & Gopnik, A. (2023). Preschoolers and Adults Learn From Novel Metaphors. *Psychological Science, 34*(6), 696–704. https://doi.org/10.1177/09567976231165267

Data, code and materials availability statement

The data and code necessary to reproduce the analyses presented here are publicly accessible, as are the materials necessary to attempt to replicate the findings. Analyses were also pre-registered. Data, code, materials, and the preregistration for this research are available at the following URL: https://osf.io/ta8vf.

Ethics statement

Ethics approval was obtained from the institutional review board at the University of Washington. Parents of all participants provided informed consent.

Authorship and Contributorship Statement

AS conceptualized the study and wrote the first draft of the manuscript, AS and TP designed the materials, and TP and SS collected the data and contributed to revising the manuscript. All authors contributed to analyzing the data and approved the final version of the manuscript.

Acknowledgements

This research was supported by an award from the NSF Division of Research on Learning (award number 2201033) to AS. Thank you to the participating families and to Wendy Garcia, Elizabeth Li, Sam Seaver, Eun-Sun Shin, Julia Souza, and Yongyan Yue for their assistance with data collection.

License

Novel metaphor processing in young autistic children

Isabel Martín-González
University of the Basque Country (Vitoria-Gasteiz, Basque Country, Spain)

Kristen Schroeder
Ingrid Lossius Falkum
University of Oslo (Oslo, Norway)

Elena Castroviejo
University of the Basque Country (Vitoria-Gasteiz, Basque Country, Spain)

Agustín Vicente
University of the Basque Country (Vitoria-Gasteiz, Basque Country, Spain)
Ikerbasque Foundation for Science (Basque Country, Spain)

Abstract: The purpose of this study is to explore possible differences between autistic and neurotypical (NT) children in novel metaphor comprehension. Much of the recent literature has connected metaphor comprehension difficulties that autistic individuals exhibit to general linguistic abilities. In our design, we carefully pair-matched young autistic children (3.13 to 12.25 years of age) to NT controls (3.69 to 9.04 years of age) on verbal mental age and tested their metaphor interpretation abilities with a picture selection paradigm combined with eye tracking measures. We predicted differences in performance in both types of measures, although we foresaw autistic participants performing above chance in the picture selection task. However, results did not show a difference between groups in the picture selection task, which would favour accounts that relate metaphor interpretation to linguistic abilities in autistic population. Interestingly, the eye tracking observations revealed differences between groups concerning gaze movements in the region corresponding to the processing of the metaphoric vehicle. Such differences replicate those found in previous studies with similar designs, such as Vulchanova et al.'s (2019). On the other hand, the evidence presented and discussed in the paper does not suggest either impairment or delay with respect to metaphor processing. Rather, the evidence only suggests differences. While the source of such processing differences is still unknown, the results of the current study cast some doubts on the idea that the main factor in metaphor processing in the autistic population is their structural language level.

Keywords: metaphor; autism; children; development; eye-tracker.

Corresponding author: Isabel Martín-González, Department of Linguistics and Basque Studies, University of the Basque Country/Euskal Herriko Unibertsitatea, Vitoria-Gasteiz, Basque-Country, Spain. E-mail: isabel.martin@ehu.eus.

ORCID ID: https://orcid.org/0000-0003-4467-7120

Citation: Martín-González, I., Schroeder, K., Falkum, I.L., Castroviejo, E., Vicente, A. (2025). Novel metaphor processing in young autistic children. *Language Development Research, 5*(2), 28-66. http://doi.org/10.34842/im-g

Introduction

Difficulties understanding non-literal uses of language are common across the autistic spectrum in everyday situations (Morra, 2016). In fact, many intervention programs targeting language or social skills include interventions on non-literal uses of language or non-explicit communication (McMahon et al., 2013; Melogno et al., 2017). However, results in the lab concerning non-literal uses of language are typically mixed, depending on what kind of non-literal use is being tested, method of assessment, and matching criteria for neurotypical (NT) controls. For instance, many studies on children, adolescents or young adults report special difficulties understanding irony and sarcasm (Deliens et al., 2017; Happé, 1993; MacKay & Shaw, 2004; Li et al., 2012; Panzeri et al., 2022; Saban-Bezalel & Mashal, 2019; Song et al., 2023), with performance on other aspects of non-literal language being more variable and frequently unaffected in tested samples. For example, many studies report NT-like comprehension of scalar implicatures (Chevallier et al., 2010; Hochstein et al., 2018; Pijnacker et al., 2009; Su & Su, 2015; Van Tiel & Kissine, 2018) (see Mazzaggio et al., 2021; Schaeken et al., 2018 for discrepant results), or on indirect speech acts (Kissine et al., 2015; Marocchini et al., 2022; but see Ozonoff & Miller, 1996; Paul & Cohen, 1985 for opposite results).

Current data on metaphor comprehension in autism is also mixed and the extent to which underlying processing may differ between autism and neurotypical (NT) development is underexplored, especially in younger children. In this study, we specifically evaluate the comprehension and processing of novel metaphoric utterances by autistic children, in comparison to verbal age matched controls, contributing to the wider debate on comprehension of non-literal uses of language in autism.

Theoretical Accounts for Non-literal Language Comprehension in Autism

Discrepant results between types of non-literal uses of language suggest that not all of them impose the same demands for autistic individuals. There have been attempts to theoretically disentangle non-literal uses of language in terms of what might be more costly specifically for autistic people. A distinction that has been proposed differentiates uses of language that require perspective-shifting, and those that arguably can be understood without adopting the perspective of the interlocutor (Andrés-Roqueta & Katsos, 2017, 2020; Kissine, 2016). Andrés-Roqueta & Katsos propose a distinction between *linguistic* and *social* pragmatics: linguistic pragmatics would mainly demand linguistic abilities and sensitivity to pragmatic norms such as pragmatic maxims of informativeness (which, according to them, would involve e.g., scalar implicature derivation). On the other hand, social pragmatics would require additional mind-reading abilities to track a speaker's belief and infer that the speaker's utterance should not contradict that belief (e.g., as in irony, sarcasm).

According to this view, metaphor comprehension would fall under linguistic pragmatics in that it would require sufficient structural language skills and sensitivity to pragmatic norms but arguably not the degree of mentalizing required in irony comprehension. A particularly influential study to this perspective is presented in Norbury (2005). Norbury evaluated the performance of autistic children with and without language impairment in a metaphor[1] comprehension task very similar to one used by Happé (1993), who claimed to have shown that autistic children's low performance in metaphor comprehension was related to their impairment in first-order Theory of Mind (ToM). In Norbury's study neither ToM nor autistic traits emerged as relevant predictors in the regression model, while core language skills did predict metaphor comprehension. Moreover, when matched on structural language abilities, autistic children performed just as well as their NT younger peers on metaphor comprehension. The author therefore concluded that structural language is a stronger predictor of performance in metaphor comprehension than ToM in autism.

Metaphor Comprehension in Autistic Individuals

Since the publication of Norbury (2005), many studies have offered support for the claim that metaphor *comprehension* is more or less spared in autistic individuals (children and adults) when paired on structural language skills, including verbal mental age (VMA) to NT controls (Giora et al., 2012, Kasirer & Mashal, 2014, 2016; Whyte et al., 2014, Chahbouhn et al., 2016; but see Morsanyi et al., 2020, for a meta-review suggesting opposite results). However, several other studies have suggested difficulties in metaphor *processing* in autistic adults, young teens or children whose linguistic and intellectual skills are within the typical range (Chahboun et al., 2017; Chouinard & Cummine, 2016; Gold & Faust, 2010; Vulchanova et al., 2019).

Reviews of metaphor processing in autistic population such as Kalandadze et al. (2019), or Lampri et al. (2023) suggest that discrepancies in results on metaphor between different studies might be due to variability in research methods (i.e., on-line vs. behavioural) and task designs (i.e., the type or form of the metaphor chosen). Concerning methods, on-line measures such as eye tracking (Vulchanova et al., 2019), reaction times (Chouinard & Cummine, 2016), semantic priming (Chahboun et al., 2017) or ERPs (Gold & Faust, 2010) uncover atypicalities in autistic individuals' processing of metaphors that do not emerge in studies using behavioural, multiple-choice paradigms (as in Norbury's own study). Although verbal skills have proven to be relevant to understanding different types of metaphors, the fact that differences persist between autistic and NT controls despite similar linguistic and intellectual skills, suggests that verbal ability may not account for the whole story.

Regarding the choice of materials, a particularly relevant factor of variability can be the *conventionality/novelty* of the metaphors tested. Conventional metaphors are

---

[1] The novelty/conventionality of metaphors in this particular study was not controlled for. Most of the metaphors appear quite conventional (i.e. something is an oven when it is quite hot). Most of the metaphors and similes are adapted from Happé, 1993. This is important because conventional metaphors may be processed differently from novel ones or, at least, require some specific skills, as commented in the following pages.

typically taken to be stored in the lexicon of proficient language users, suggesting a closer relation between structural language and metaphor comprehension. In contrast, understanding novel metaphors seems to require the recruitment of pragmatic (inferential) abilities (at least according to some prominent pragmatic theories; see e.g., Sperber & Wilson, 2015). Generally, evidence that compares directly between conventional and novel metaphors suggests that conventional metaphors are more difficult for autistic children and teens than novel ones (Zheng et al., 2015; Kasirer & Mashal, 2016; or see Melogno et al., 2012 for a review in autistic children; see Pastor-Cerezuela et al., 2020 for opposite results). Moreover, and specifically in children, vocabulary level has been found to relate to performance on conventional metaphors but not significantly for novel metaphors (Kasirer & Mashal, 2014; Olofson et al., 2014). The role of core language skills on novel metaphor comprehension is therefore less clear than in the conventional metaphor case.

Lastly, the results presented to this point include widely different profiles of autistic participants, especially with regard to chronological age. Most studies work with older children from 10 to 12 years old (e.g. Kasirer & Mashal, 2016, Chahboun et al. 2017, Vulchanova et al. 2019), a combination of older children and adults, or adult participants only (see Gold & Faust 2010 or Giora et al. 2012 as examples). A scarce number of studies has focused exclusively on children, and those that have, often include children of ages in which metaphor comprehension could already be more or less established, especially if including autistic profiles with typical IQ and linguistic skills.

Metaphor Comprehension in Autistic Children

Rundblad & Annaz (2010) ran a developmental study on conventional metaphor comprehension by autistic children ranging from 5 to 11 years old in a picture selection task. The results show little improvement with age, as well as a lack of relation between conventional metaphor comprehension and VMA at this point of development, though the sample was relatively small (11 children).

Van Herwegen & Rundblad (2018) compared autistic children and teens (mean age = 16 years) to chronological age matched NT controls on novel metaphor, using a picture selection task. They found that the autistic group performed significantly worse than controls throughout the entire age span included. As part of a second, longitudinal experiment, eight of the younger participants of the previous study were re-selected. Results showed that response accuracy significantly improved with age. This would suggest that novel metaphors may cause special difficulty for autistic children (more than conventional metaphors, which reinforces the finding by Kasirer & Mashal, 2016) but that this difficulty can be overcome— a finding that contrasts with Rundblad and Annaz (2010).

In contrast, Pastor-Cerezuela et al. (2020) found that novel metaphors were more difficult than conventional ones for autistic children, in a study comparing autistic (n=22) and NT (n=22) children aged 6 to 12. They also found that the autistic children were less accurate than the NT children when matched either by chronological age or by verbal mental age.

These studies involved participants whose linguistic and intellectual abilities were, on average, within the typical range. However, neither study matched autistic participants to NT controls on these variables, relying instead on average measures. These findings suggest that average linguistic and intellectual skills alone may not ensure equivalent metaphor comprehension.

Conversely, Zheng et al. (2015) compared the performance of 15 autistic children to the performance of 15 NT children on metonymy and conventional and novel metaphor comprehension. They found that conventional metaphor was more difficult than novel metaphor for the autistic children, confirming again Kasirer & Mashal's (2016) findings. In addition, they found no inter-group differences in the *novel* metaphor condition. Accuracy on this condition was predicted by receptive vocabulary in the autistic group.

In a study combining eye-tracking, mouse-tracking, and picture selection, Vulchanova et al. (2019) investigated novel metaphor comprehension in autistic children aged 10 to 12 years. Participants, matched to NT controls on verbal comprehension using the Wechsler scales (Wechsler, 2005, 2012), achieved 80% accuracy in the picture selection task for novel metaphors and idioms combined—significantly lower than NT children. Eye tracking revealed that autistic participants fixated more and spent longer on literal images during the early stages of processing compared to NT controls. By the end of the trial, however, both groups paid little attention to literal representations. The authors interpreted these results as evidence that autistic participants required more time to consider the literal option and make a decision. Despite this, most autistic participants ultimately selected the correct answer.

Several conclusions can be drawn from these studies. First, obtaining a uniform sample of young autistic children is challenging; most studies involve mixed samples that include teenagers or adults. Second, conventional metaphors appear more difficult for autistic individuals than novel metaphors, as shown by Zheng et al. (2015) and Kasirer and Mashal (2016), regardless of task type (verbal explanation or multiple choice) or age range (5–16 years). Third, comparisons between Van Herwegen and Rundblad (2018) and Vulchanova et al. (2019), which both use a simple picture selection task with autistic children, suggest that matching participants on structural language skills is critical for success in novel metaphor comprehension. Vulchanova et al. (2019) matched participants on verbal and intellectual abilities using Wechsler scales, while Van Herwegen and Rundblad (2018) matched participants only on chronological age. The former group performed significantly better, suggesting that with equivalent verbal and intellectual skills, autistic participants' performance on novel metaphor tasks can approach that of NT individuals, albeit with some differences. In addition, Zheng et al. (2015) show that receptive vocabulary is an important candidate when predicting successful comprehension of novel metaphor in autistic children.

Finally, processing measures, such as those employed in Vulchanova et al. (2019), can reveal atypicalities not captured by accuracy measures alone. Gaze behaviour data indicate that processing may differ in autistic individuals, even when they ultimately achieve high accuracy. While autistic participants in Vulchanova et al. (2019) performed well (80% accuracy), their processing was slower and more focused on literal interpretations, which may have hindered their final performance in the selection task to a minor extent. Such findings highlight the value of integrating processing data with accuracy measures to better understand metaphor comprehension in autism.

The Present Study: Aims and Predictions

In our study we aimed to add to the scientific literature by exploring autistic children's metaphor comprehension in an understudied age range of 5 to 11 years of age. In particular, we wanted to investigate whether autistic children would exhibit more difficulties than NT children of their same VMA (pair-matched) with processing and understanding *novel metaphor*.

VMA-matching was made on the basis of receptive vocabulary scores, as one of the possible measures the Peabody Vocabulary test (PPTV, Dunn & Dunn, 1997) offers. The reasoning behind it is that children who are in a given developmental stage usually know the set of words that is used in the Peabody as a test for each age span. For instance, if typically, 5-year-olds know the meaning of "broom" or "bottle" that would mean that a child who knows those words but not the ones typical of 6 y.o., is more similar to 5 y.o., from a maturational point of view. We acknowledge, however, that receptive vocabulary is but a part of verbal abilities, and so that receptive vocabulary scores can only be considered a proxy for general verbal maturation. We used the Peabody test for receptive vocabulary both for practical reasons (especially in the case of autistic children, since it is easy for them to understand what they need to do, and it is also short) and because most of the studies that we cite also provide VMA measures based on the Peabody.

We obtained offline measures with a picture selection task including a literal competitor and simultaneous online measures by registering gaze movements with an eye tracker, using a Visual World Paradigm[2] (Cooper, 1974; Huettig et al., 2011; Tanenhaus et al., 1995). We had a particular interest in this type of design, since results of both types of measures appear to reveal different aspects of performance, and they can nuance each other. Blending both measures in the same task follows

---

[2] Visual World Paradigm (VWP) experiments involve visual input, usually pictures, and auditory stimuli (like a classical picture selection task). But they also include the registration of the participant's looks to the pictures while hearing the audio, which obviously contains information related to them, in one way or another. VWP results interpretations are based on what is known as a "linking hypothesis" which connects where participants look, their visual attention, to their unfolding comprehension or planification of an utterance (see Zhan, 2018, pp.1).

Vulchanova et al.'s (2019) task design. Our task is structurally very similar to it, although it differs in important aspects that will be detailed in the methods section, such as the type of metaphorical and literal context included, the type of literal competitor, etc.

Taking into account previous literature, we did not expect that autistic participants would fail to understand the novel metaphors included in this experiment across the board. However, we predicted that autistic participants would perform significantly worse than NTs in the picture selection task despite the VMA-matching, as happens in both Van Herwegen et al. (2018) and Vulchanova et al. (2019) with novel metaphors among autistic children with linguistic abilities close to the typical range. We also expected differences between both groups in how they process metaphors. On the other hand, we did not expect differences between groups in the baseline, literal condition, in either picture selection or in eye tracking.

The design of the task aimed at reducing uncertainty as to what the metaphor could mean by providing children with a piece of world knowledge that is relevant in order to figure out what the metaphor means. Utterances of metaphors in real life conversations are typically open ended (Pouscoulous, 2014), which may generate more uncertainty in autistic than in non-autistic individuals (see Vicente et al., 2023, on the role of uncertainty in non-literal language in autism). The paradigm has already been tested on NT children (ages 3 to 9), and results are published in Martin-Gonzalez et al. (2024). There it was found that until 6 years of age, NT children show below or at chance performance in picture selection. Regarding gaze performance, all age groups were above chance in looks to the correct metaphoric image, and such performance strengthened with age. The age span of 3 to 9 years of age proved to be critical to better understand the developmental trajectory of metaphorical abilities in NT children.

Method

Participants

Participants were 29 autistic children and 29 NT children, all native speakers of Spanish. Twenty-five of the autistic participants were recruited from Early Intervention Services and the Association of Autism in Vitoria-Gasteiz, Spain and four participants were recruited from the Association of Autism in Bilbao (APNABI) in Bilbao, Spain. Inclusion criteria were the following:

a) having a non-verbal IQ over 75 points per the Leiter-3 scale (Koch et al., 2019; Roid & Miller, 2013). We had no Leiter data for five participants, but they were reported by their clinicians in the Association to have average intellectual and linguistic skills.
b) being verbal or conversational children according to the Autism Diagnostic Observation Schedule (ADOS-2, Lord et al., 2012), or the report from the experts who either worked with them or performed the diagnosis.
c) having an official diagnosis of autism spectrum disorder (ASD) or of pervasive developmental disorder, in which case they were administered the ADOS-2

observation scale to confirm ASD.

The participants in the NT group were recruited from public schools in Vitoria-Gasteiz. A larger number of participants were part of a previous developmental study by Martín-González et al. (2024) and 29 were selected from the total sample for the purposes of the current study to be individually matched on verbal mental age, measured with the Peabody Picture Vocabulary Test III, (PPTV-III, Dunn & Dunn, 1997) with the participants in the autistic group. See the full details of group matching in Table 1.

Table 1. ***Group Matching.***

		Autistic group	NT group	T test
N		29	29	
Chronological Age (years)	Mean (SD)	7.66 (2.26)	5.97 (1.59)	t = 14.71, df = 1039.9, p < .001
	Age range	3.13 – 12.25	3.69 – 9.04	
Verbal Mental Age (years) - PPTV	Mean (SD)	6.073 (1.99)	6.09 (1.91)	t = -0.15, df= 1155.5, p = .88
	Age range	3.1 – 10.11	3.1 - 10.10	

As can be seen, groups were different with regard to chronological age. As detailed in the analysis section, part of our analyses is dedicated to disentangling whether this variable has an effect in the results, both for picture selection and eye tracking. This was due to the majority of autistic children in our sample, regardless of their intellectual ability, being below their chronological age in receptive vocabulary as measured by the PPTV-III.

Materials and Design

Building on Martín-González et al. (2024), the present experiment collected both behavioural and processing data on novel metaphor comprehension. To ensure all metaphors were novel in Spanish, we conducted a small survey which asked Spanish adults (N =21) to rate how familiar they were with some novel and conventional metaphors on a 1-7 Likert scale, where “one” was the lowest familiarity (see Martín-González et al., 2024 for the results, although novelty scores of the chosen items are included in the list of selected items in Appendix 1). Only metaphors with a mean score under 3 points were selected, or with a median of 2 points or less.

The design manipulated a within-subject and a between-subject independent variable. The former was Group (autistic vs. non-autistic) and the latter was Utterance type (literal vs. metaphorical). Specifically, in a picture-sentence matching task, children were presented with four pictures displayed on a computer screen (target, competitor and two distractor images), as exemplified in Figure 1, and heard a sequence such as (1), in Spanish (see Appendix 1 for a full list of the trial items).

There was a total of 20 items (i.e. 20 sets of four pictures). Participants heard 10 literal and 10 metaphoric utterances, presented in a randomized order. Participants were assigned to one of two lists which counterbalanced which items were presented with a metaphorical or literal statement. The position of the pictures (target, competitor, distractors) was randomized across items.

Figure 1. ***Example of Picture Assembly in a Trial.***

(1) Los saltamontes saltan mucho. {Este animal, Este niño} es un saltamontes. ¿Cuál es?
'Grasshoppers jump a lot. {That animal, That child} is a grasshopper. Which one is it?'

The task consisted in responding to the prompting question by pointing at one of the four pictures. Throughout the task, the child's eye gaze was recorded (see the details in the next subsection).

The structure of the trial was as follows: participants first heard a **context utterance** which described a generic property of the animal or object. The motivation for including this context utterance was twofold: first, it was intended to minimize the effects of differences in world knowledge by providing the piece of world knowledge information required to comprehend the metaphorical utterance; secondly, it was intended to minimize uncertainty, especially in the autistic population, by providing the specific feature to be used in the metaphorical mapping (on uncertainty and non-literal language in autism, see Vicente et al, 2023). The context utterance was followed by a **target utterance**, which presented the sentence containing either the literal or metaphorical condition to be tested (in (1), if the subject was "That animal", the sentence would be interpreted literally, and if it was "That child", it would be interpreted metaphorically). The final component of the trial was the prompting **question**, which forced the child to decide for one of the four alternatives.

For the purposes of eye gaze analysis, the relevant time windows were the following: the **target word**, that is the predicate noun of the target utterance (in (1), "grasshopper"), and the **question** ("which one is it?").
Table 2 schematizes a typical trial with the indication of the two conditions being

studied as well as the relevant windows for eye gaze analysis.

Table 2. ***Trial Structure.***

Condition	**Context sentence**	**Target utterance**		**Question (eye tracker window)**
			Target word (eye tracker window)	
Literal	Grasshoppers jump a lot	That animal	is a *grasshopper*.	Which one is it?
Metaphorical		That child		

Procedure

All participants were tested in a quiet room. Parents of autistic children were allowed to be in the testing room and some children who found it difficult to sit still, sat on their parents' lap, who were instructed to keep their eyes closed during the testing. Participants sat next to one of the experimenters, in front of a screen that was connected to the eye-tracker computer. Eye movements were recorded with a SMI RED250MOBILE portable eye tracker with a sampling rate of 250 hz. The experimenter set the eye tracker and then started the experiment with a 5-point calibration and validation phase. The experiment only continued if deviation of both eyes from the focus point was under 0.5 degrees. Afterwards, participants went through three practice trials to ensure their attention and understanding of the task. The practice trials were similar to critical ones but did not use any figurative language. Participants then continued with the 20 critical trials. The experimenter clicked to advance through the items so as not to rush the child's response. After the experimental session, they were given stickers as compensation. The calibration phase took approximately five minutes followed by an additional five to six minutes to complete the metaphor task.

Analysis

We conducted different analyses for picture selection and eye-gaze data, following a very similar strategy to the one in our previous work (Martín-González et al., 2024). On the one hand, we analysed which picture participants chose in each trial, from among the four options, and compared both groups (autistic vs. NT) in that regard. We focused on the choices for the literal picture compared to the metaphorical one in both conditions. On the other, we analysed where participants were looking at (which of the four pictures) during the unfolding of the auditory stimuli. In order to make statistical comparisons, we divided the auditory stimuli into regions (time-windows, see Table 2) and analysed the critical ones: the TARGET region, when they are hearing the metaphorical vehicle, i.e., the word that is being used with a metaphorical meaning, and the QUESTION region, in which participants were prompted to choose, to make a decision, while hearing 'which one is it?', at the end of the trial.

Our goal was to compare the gaze behaviour of both groups while processing these critical windows of the auditory stimuli, under both levels of the condition variable (literal and metaphorical) and both levels of the group variable (autistic and neurotypical). Since, as seen in Table 1, both groups differed with regard to chronological age (autistic children were older), we also tested the effect of this variable on both picture selection and eye tracking results. We sum-coded both independent variables (condition and group) in all analyses to better understand the main effects of these variables in the results overall (the effect of condition regardless of group and the effect of group regardless of condition; and lastly, whether there was an interaction effect between both variables).

To explore possible interaction effects, first we visualized data with different graphic formulas, all of which will be presented in this paper. In addition, we used the package *emmeans* (Lenth et al., 2023) to apply pairwise comparisons to delve into the differences between autistic and neurotypical participants in both conditions (only when significant interaction effects between the two categorical variables, condition and group, were present in the model output). The analysis script and datasets are available at the project's OSF page:
https://osf.io/ksuwv/?view_only=accde2b9d44b49eea164e75dce89796a

Gaze Behaviour

The analysis of gaze data was performed using the *afex* package in R (Singmann et al., 2022), to model results using a mixed-effects approach. Same strategy was used for both time windows, which were analysed separately. Thus, we built the same model but applied it separately to TARGET region data and QUESTION region data, in order to reduce noise in the analysis. In both models we included GROUP*CONDITION as fixed factors, due to our hypothesis that autistic children would experience more difficulties than NT children in the metaphorical condition, but not in the literal one. We included Item and Participant as random slopes, as well as CONDITION, as the intercept by Participant, and the interaction between CONDITION and GROUP as the intercept by Item. We followed the recommendations for a hypothesis-driven, maximal random factor structure, proposed by Barr et al. (2013).

Lastly, our response variable was *logGaze* (see (2)). We followed the approach taken by Ronderos et al. (2022), which we also followed in our previous study on this matter with neurotypical children (Martín-Gonzalez et al., 2024); that is based on creating a variable comparing the proportion of looks to the metaphorical picture with the proportion of looks to the literal picture, and transforming it to a logarithmic scale with log probability ratios (Arai et al. 2007, see (2) for the explicit formula). The advantage of using this variable is that it allows us to study the preference participants have for one of the critical pictures (the correct one) and compare it with the other one (the competitor), which also resembles the approach taken in picture selection analysis, as will be described below.

(2) *logGaze = log(proportion of looks to the metaphorical + 0.1/proportion of looks to the literal +0.1)*

The 0.1 correction stands in order to avoid divisions of 0/0 which would yield uninterpretable results. This variable will also be used to visualize results in some of the graphs. A summarized interpretation would be: if results are positive, in this case, it means that the preference is inclined towards the metaphorical picture, because it means that a greater proportion of looks are directed towards this area compared to the literal. Conversely, if results are negative, it means that more proportion of looks are directed towards the literal picture area than to the metaphorical. The closer the result is to zero, the less clear the preference is, as it means that both numbers in the division are similar to each other (a graphic example lies in Figure 3, in the results section).

Furthermore, we also wanted to explore the influence of chronological age in results, within both time windows. We coded the same mixed-effects model but adding AGE to the interaction between CONDITION*GROUP, as a numerical variable; thus, including a three-way interaction. Our hypothesis regarding the random variability did not change with the addition of AGE. Therefore, we used the same random factor structure for this model. This model was also sum-coded, since we wanted to explore whether age was a significant predictor overall (whether there was a main effect of chronological age, across groups and conditions), and also its interactions with group and condition.

Picture Selection

To analyse picture selection data, we fitted a mixed-effects model with exactly the same structure as the one for eye tracking data (and also using the same package, *afex*), also sum coded. The only difference being the response variable, which in this case was “response”, and contained 1 and 0 (1 for correct answers, 0 for incorrect). Our response variable was a dichotomic one, even though our picture selection task presented participants four different types of pictures: metaphorical, literal, object distractor and person distractor. All in all, as will be visualized in sections below, the preference for distractors was extremely low in both groups, and the real competition occurred between the literal and the metaphorical picture, in both conditions. Thus, we set the chance level at .50 and coded our response variable in the picture selection model as a dichotomic one, between the metaphorical and the literal picture. In order to run this model, trials in which distractors were chosen were deleted.

To study the influence of AGE in the picture selection results, we fitted the same model as for the eye tracking data (i.e. also with a three-way interaction, and the same random factor structure).

Results

Eye-Tracker

In Table 3 we declare the different patterns of eye tracking data for each group. Generally (and somewhat as expected) the quality of the eye tracking ratio was worse for the autistic participants, although not dramatically (around 15 percentual points, and standard deviation does not differ much from NT's). As said in the Methods section, calibration and validation of gaze for autistic participants was more difficult and interrupted.

Our sample for eye tracking data analysis has one less participant because the quality of eye tracking data was very low for one autistic participant in particular. There was no registration of this participant's saccades to almost any of the Areas of Interests. Therefore, their trials were not taken into account. This data exploration is also available in the data analysis scripts in the OSF repository.

Table 3. ***Characteristics of Eye-Tracking Data per Group and Condition.***

Group	Condition	Nº observations (gaze movements in 16 ms)	Mean (SD) quality of eye tracking ratio	Sample size (N)
Autistic	Literal	23308	67.34 % (22.37)	28
Autistic	Metaphorical	24648	66.14% (23.93)	28
NT	Literal	31904	80.6% (18.26)	29
NT	Metaphorical	32748	80.51% (18.13)	29

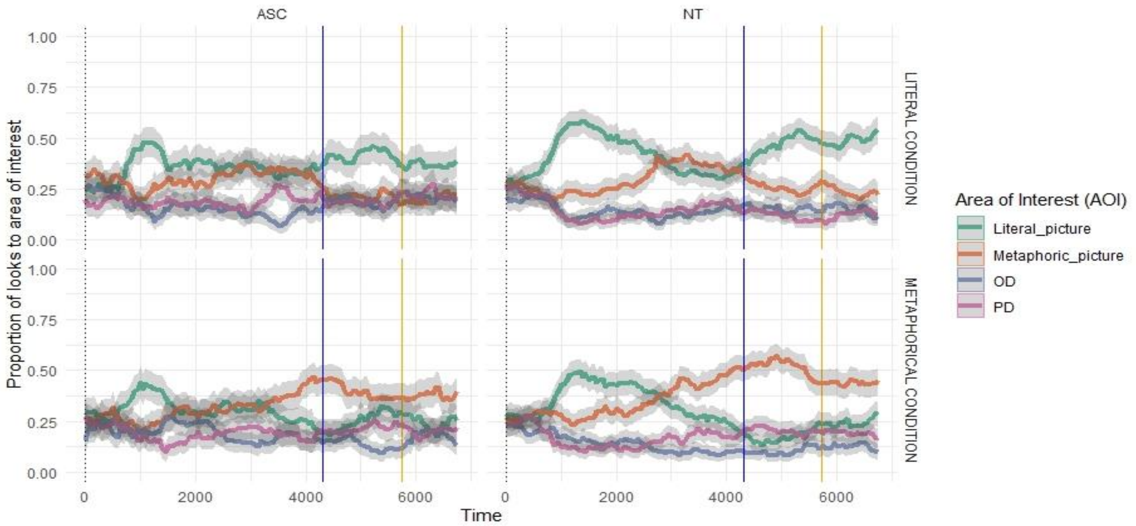


Figure 2. ***Proportion of Looks to each Type of Picture during the Entire Trial.***
Legend: The blue line represents the mean start of the target utterance across trials (that kid/animal is a grasshopper). The yellow line represents the mean start of the metaphorical vehicle (grasshopper). The graph represents the average tendency of looks for all participants and trials. Error ribbons show 95 percent confidence intervals.

However, it seems that when participants start hearing the predicate information, in this case information about jumping (around 2000-3000 ms), they start to look at the metaphorical picture, since it is the only one which depicts jumping. This change happens with more clarity in the NT group than in the autistic group, in which there appears to be a period of mixed looks between the literal and the metaphorical picture around 2000-3000 ms, in both conditions (see orange and green line, in Figure 2). Also, in both groups, after the start of the target utterance, which disambiguates between which picture is going to be the correct one (by mentioning a child, who is a grasshopper, or an animal, which is a grasshopper), the amount of looks for either the literal or the metaphorical picture increases.

Nonetheless, in the autistic group and the metaphorical condition, there is again a period of mixed looks between the literal and the metaphorical picture, if we take a look into the area around the yellow line, which does not happen in the NT group. We will be looking into this effect more deeply in the next sections, when statistically analysing what children prefer in the target word region and the question region.

Lastly, there seems to be an interesting effect in the metaphorical condition (which was also observed in our previous study only with neurotypicals): both groups look to the metaphorical picture more frequently even before hearing the beginning of the target utterance. This does not happen in the literal condition. A possible explanation is that children are anticipating that the correct choice is going to be the jumping child; probably because they are hearing information about jumping, which leads them to look to the jumping child (following the example of the grasshopper trial) and then when they hear “that child” that reinforces the prediction. The real difficulty seems, though, to maintain that prediction when hearing again “grasshopper”, this time used metaphorically, to describe the jumping child. There seems to be a slight drop in the looks to the metaphorical picture especially in the autistic group, around the beginning of the target word (around the yellow line).

Target-word Region Analysis

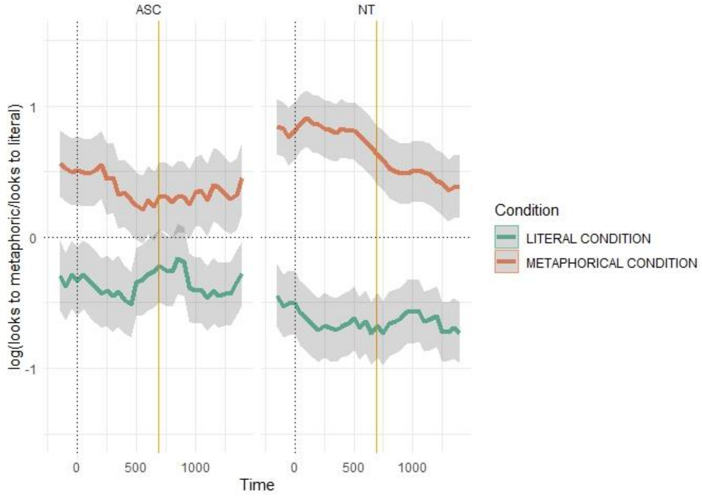


Figure 3. ***Depiction of Preference (logGaze) between Metaphorical and Literal picture at Target Word Region**.*

Legend: Dotted line represents the beginning of the region. Yellow line represents the end of the target word (i.e. grasshopper), which begins at point 0. The graph represents the average tendency of looks for all participants and trials. Error ribbons show 95 percent confidence intervals.

As said, we wanted to explore the specific preference for either the metaphorical or the literal picture in each condition and time-window.

We built a mixed-effects model to test whether the interaction between group and condition was significantly influencing preference, i.e., the above-explained logGaze. Relevant output summary is displayed in Table 4.

Table 4. ***Eye-Tracking Results in the Target Word Region by Group and Condition.***

term	Estimate	Std. error	*df*	*t*	*p*
Intercept	-0.01	0.071	87.53	-0.08	.933
MET v. LIT	-0.95	0.15	86.07	-6.40	< .001***
AUT V. NT	-0.17	0.14	86.283	-1.21	.227
CONDITION X GROUP	0.68	0.29	86.753	2.34	.021 *

The model shows a main effect of condition (t value = 6.4, p<.001), but no main effect of group. The main effect of CONDITION would again speak in favour of both groups preferring the corresponding target image in each condition. The lack of a main effect of GROUP speaks in favour of groups performing similarly within each condition. However, there is an interaction effect between condition and group (t value = -2.33, p<.05). The interaction might be attributable to the fact that the difference between the literal and the metaphorical condition is not the same for both groups (as there is a differing distance between lines in Fig. 3, between both groups). Pairwise comparisons were performed to further explore the interaction. The difference between groups within the literal condition was not significant (p = .826), i.e., they performed similarly, but they did not in the metaphorical condition (p = .05), although the difference was marginal[3]. Thus, it seems that the metaphorical condition was more challenging at the processing level for the autistic group compared to their NT peers, at this time window.

Lastly, we ran a model to test the influence of chronological age in the eye tracker performance. We only found a significant interaction between AGE and CONDITION (t value = 2.53, p<.05), and CONDITION and GROUP (t value = - 3.22, p<.001) which replicates the results found in the general model for gaze movements in the target word region, and no main effect of AGE in results. Specifically, this means that age does not influence the preference for one picture or the other overall, in both conditions and groups; but that it influences preference differentially depending on the condition. The relevant output is depicted in Table 5, and depicted in Figure 5.

Table 5. ***Eye-tracking Results in Target Word Region by Condition, Group and Chronological Age.***

term	Estimate	Std. error	df	*t*	*p*
Intercept	0.01	0.075	85.27	0.19	.848
MET v. LIT	-0.89	0.15	83.39	-5.85	< .001***
AGE (years)	0.05	0.06	52.08	0.83	.413
AUT v. NT	-0.20	0.15	83.58	-1.37	.175
CONDITION X AGE	-0.31	0.12	52.38	-2.53	.014 *
CONDITION X GROUP	0.97	0.30	84.01	3.22	.002**
AGE X GROUP	-0.11	0.12	52.09	-0.91	0.366
CONDITION X AGE X GROUP	-0.19	0.24	52.38	-0.77	0.442

More specifically, it seems that in both groups, age influences the preference for the correct picture. As children in both groups grow older, their probability of looking at the metaphorical picture in the metaphorical condition and the literal picture in the literal condition increases (thus the interaction effect between AGE and CONDITION). However, and in line with the general results for the target word region (Figure 3, Table 4), explained above, the preference is less clear for the autistic group, in most of the age span included. NT children already start with much greater clarity of preference than autistic children (thus, the interaction effect between GROUP and CONDITION).

---

[3] From now on, we will take this difference as significant but marginal, which diminishes the strength in the conclusions of this study; thus, it can be considered as a limitation, that calls for further replication of the design. However, since the difference between groups in the literal condition was far from significant, the significant interaction effect found in the model (Table 4) can likely be attributed to the group differences observed in the metaphorical condition.

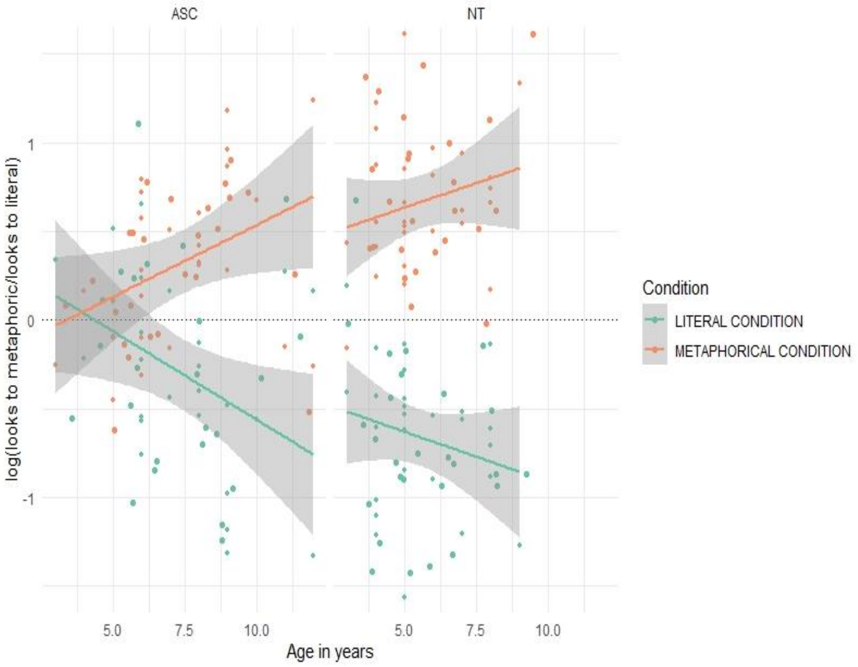


Figure 4. ***Development of Preference for Metaphorical vs Literal in the Target Word Region.***

Legend: Dots represent a single participant; the superimposed regression lines represent the mean tendency.

Question Region Analysis

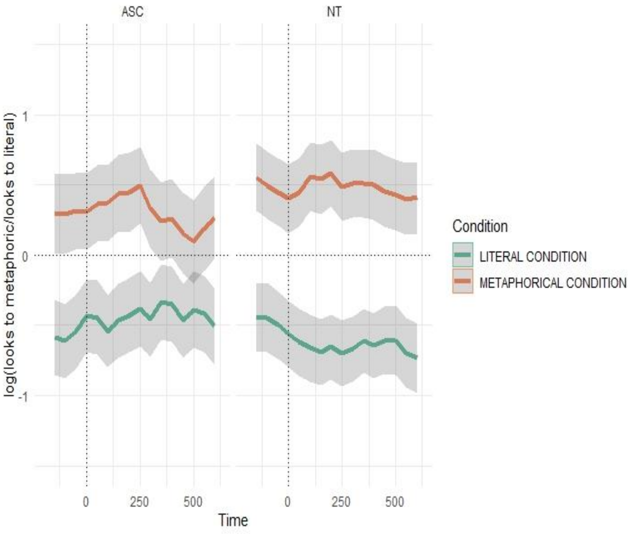


Figure 5. ***Depiction of Preference between Metaphorical and Literal picture at Question Region.***

Legend: Dotted line represents the beginning of the region. The graph represents the average tendency of looks for all participants and trials. Error ribbons show 95 percent confidence intervals.

As explained in the Analysis section, we ran another mixed-effects model with the same structure as the one used for the target time-window but applied to gaze movements during the question region (when children hear "which one is it?", see (1) and Table 2 above). Relevant output summary can be found in Table 6.

Table 6. ***Eye-Tracking Results in the Question Word Region by Group and Condition.***

term	Estimate	Std. error	df	*t*	*p*
Intercept	-0.04	0.09	74.76	-0.43	.670
MET v. LIT	0.89	0.14	81.36	-6.20	< .001***
AUT. V. NT	0.11	0.19	81.74	0.58	.559
CONDITION X GROUP	0.38	0.31	83.66	1.20	.233

Note: Singularity warning for this model.

In this region we only found a significant effect for CONDITION (t value = - 6.02, p <.001), no main effect of GROUP and the interaction between CONDITION and GROUP, which was significant in the target time window, did not remain within the significance threshold. This means both groups performed differently in each condition, i.e., preferring the metaphorical picture in the metaphorical condition, and the literal in the literal one (also according to the graph). When the moment of the decision is approaching, all children seem to have made up their minds with regard to which one is the correct picture, since there is no interaction effect between CONDITION and GROUP, as the one found in the target region analysis.
We also explored the influence of AGE as a linear predictor in the preference during the question region. Results from the model output are summarized in Table 7.

Table 7. ***Eye-tracking Results in Question Word Region by Condition, Group and Chronological Age.***

term	Estimate	Std. error	df	*t*	*p*
Intercept	-0.08	0.07	70.78	-0.82	.415
MET v. LIT	-0.89	0.15	77.97	-5.89	< .001***
AGE (years)	0.16	0.09	52.87	1.80	.077
AUT v. LIT	-0.03	0.20	77.71	-0.17	.865
CONDITION X AGE	-0.26	0.13	52.33	-2.11	.034 *
CONDITION X GROUP	0.69	0.33	82.43	1.84	.079
AGE X GROUP	0.12	0.17	52.83	0.69	.489
CONDITION X AGE X GROUP	0.09	0.25	52.31	0.38	.699

Results replicate the general question region model, as happened with the target region model. Moreover, no main effect of AGE is found, but a significant interaction with CONDITION (t value = -2.12, p<.05), same as happens in the target word region.

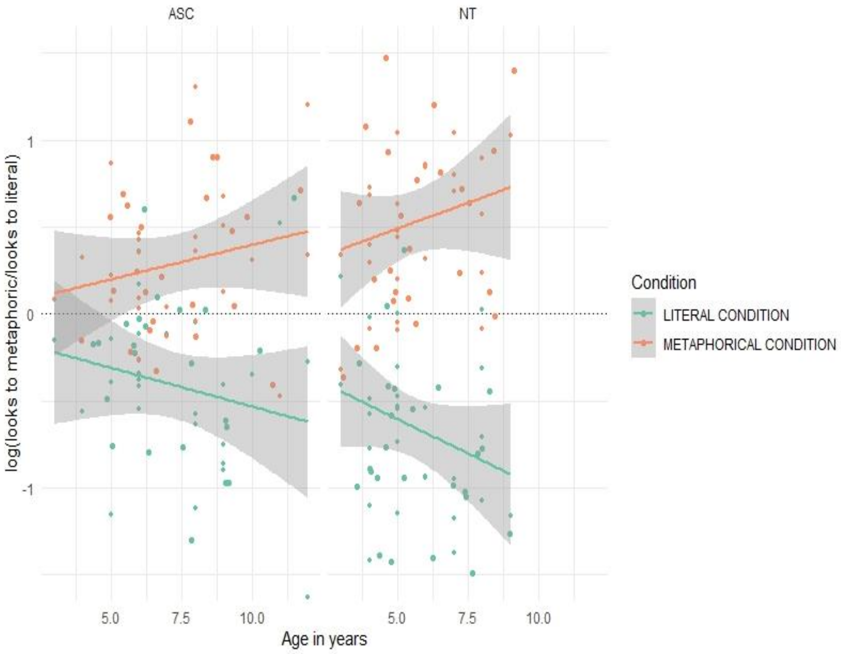


Figure 6. ***Development of Preference to Metaphorical vs. Literal picture in the Question Region.***
Legend: Dots represent a single participant; the superimposed regression lines represent the mean tendency.

Again, as tested in the target word region, the interaction effect found is due to the preference for the literal picture in the literal condition, and conversely, the metaphorical picture in the metaphorical condition, both of them increasing as children get older.

Picture Selection

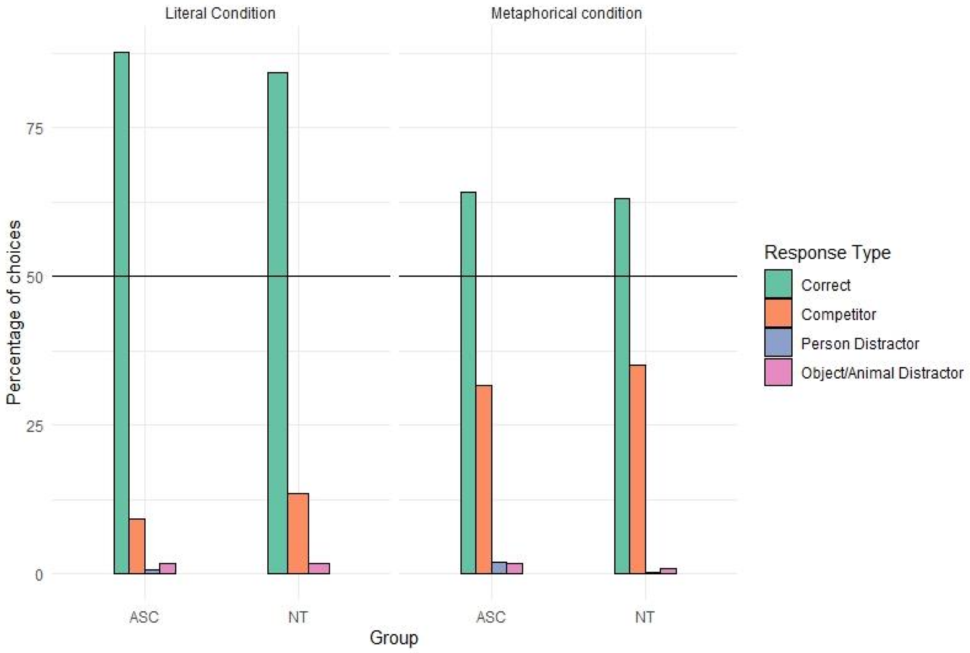


Figure 7. ***Percentage of Choices of each Category of Picture (Correct, Competitor, Person Distractor or Object/Animal distractor).***
Legend: Chance level was set at .50. Choices are depicted by all participants in all critical trials.

A mixed effects model was run to explore a possible interaction effect between CONDITION and GROUP. The relevant output summary can be found in Table 8.

Table 8. ***Picture selection Results by Group and Condition.***

term	Estimate	Std. error	*z*	*p*
Intercept	4.44	0.76	5.86	< .001***
MET v. LIT	3.79	1.05	3.60	< .001***
AUT v. NT	-0.03	0.97	-0.03	.973
CONDITION X GROUP	-0.09	2.07	-0.04	.965

There is a main effect of Condition only, not group (z value = 3.6, p<.001). Moreover, no significant interaction effect was found. The interpretation seems straightforward: both groups are choosing differently in each condition. In light of the graph (Figure 7), both groups exhibit a preference for the literal picture in the literal condition and the metaphorical picture in the metaphorical condition. Also, given that the intercept is significantly above zero, both groups are performing significantly above chance. These results dissociate from what is found in the target region analysis, in which the metaphor condition seemed to involve greater deliberation or doubt for the autistic participants (see Figure 3, Table 4). However, interestingly, they fit with the question region results, in which no specific difficulties with metaphor are found anymore (Table 6).

Regarding our analysis of chronological age effects, the relevant output summary can be found in Table 9.

Table 9. ***Picture selection results for Condition, Group and Chronological Age.***

term	Estimate	Std. error	*z*	*p*
Intercept	4.51	0.75	6.04	< .001***
MET v. LIT	3.09	1.09	2.83	.005 **
AGE (years)	1.52	0.52	2.88	.004**
AUT v. NT	-1.23	0.99	-1.60	.216
CONDITION X AGE	-1.84	1.15	-1.60	.109
CONDITION X GROUP	1.33	2.15	0.62	.537
AGE X GROUP	-0.74	1.01	-0.76	.445
CONDITION X AGE X GROUP	1.88	2.21	0.85	.395

Note: Singularity warning for this model.

A main effect of AGE as a continuous variable was found when testing the effect of the differential chronological age in both groups (z value = 2.88, p<0.001). This means age is a good linear predictor of picture selection results in both groups and since there are no interaction effects between age and condition or age and group, it does not seem that it influences differently in the literal or the metaphorical condition, or in the autistic group, compared to the NT. In our paper studying the development of NT comprehension of novel metaphors we did find an interaction effect between age and condition, that is visible in Figure 8.

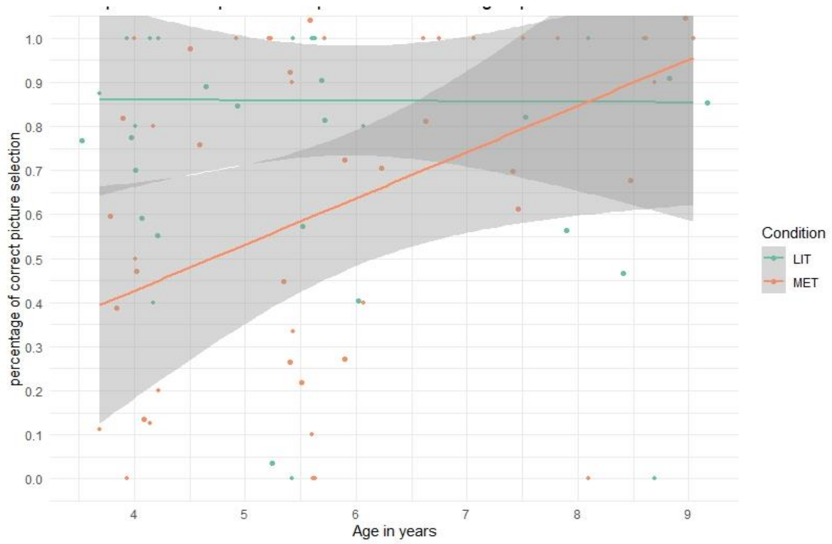


Figure 8. ***Development of Picture Selection Accuracy in the NT Group.***
Legend: Dots represent a single participant; the superimposed regression lines represent the mean tendency.

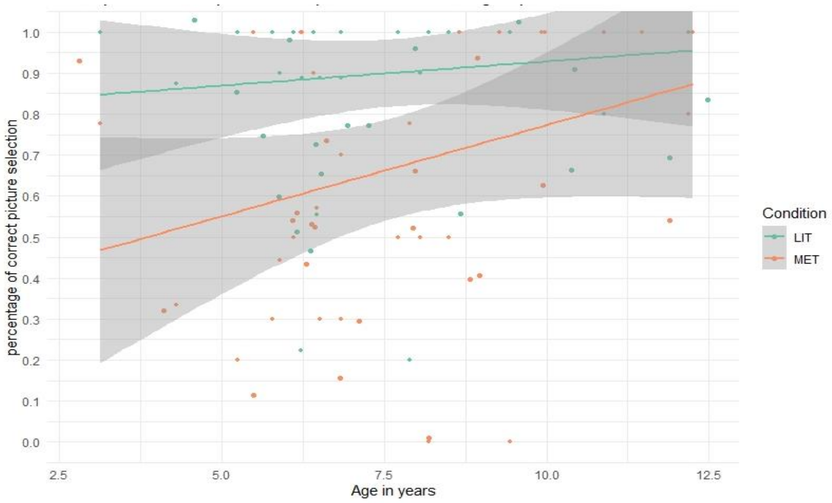


Figure 9. ***Development of Picture Selection Accuracy in the Autistic Group.***
Legend: Dots represent a single participant; the superimposed regression lines represent the mean tendency.

Discussion

In this study we sought to better understand how autistic children understand non-literal uses of language by exploring the case of novel metaphor comprehension. Previous literature suggests that when controlling for structural language maturation, group differences in the comprehension of certain non-literal language uses, such as metaphor, dissipate in off-line measures such as picture selection (Andrés-Roqueta & Katsos, 2017; Kalandadze et al., 2018; Norbury, 2005). However, whether autistic children are utilizing the same interpretive strategies or facing equal processing costs to achieve similar performance as language-matched peers is less clear, as the scarce literature that contains processing data points to (Vulchanova et al., 2019). Further complicating this issue, previous literature on metaphor comprehension in autism has (*i*) often mixed conventional and novel metaphors, which theoretically have differing processing costs, or (*ii*) not consistently controlled for language level – resulting in variability across study outcomes that is difficult to compare.

To this end, we have carefully matched our participants on verbal mental age and developed a task which tests novel metaphoric utterances specifically. By capturing on-line processing data via eye-tracking we hoped to better detect potential differences in autistic children's processing of metaphoric utterances – such as the degree to which a literal interpretation is considered. As we discuss our results, we will reflect on their significance to this wider debate.

Looking first to results from picture selection accuracy, we find that there was no difference between groups. Both groups chose the metaphoric referent above chance in the metaphoric condition and the literal referent above chance in the literal condition and rarely selected distractor pictures. Both groups also performed significantly higher in the literal condition than in the metaphoric condition. Even though our participants were matched only on receptive vocabulary, these results suggest an apparent alignment with accounts that stress the importance of general language skills for metaphor comprehension in autism (Andrés-Roqueta & Katsos, 2017; Kalandadze et al., 2018, 2019; Norbury, 2005) – so that when language level is controlled for, performance is on par with NT peers. Our results contrast with previous studies on novel metaphors such as Van Herwegen and Rundblad (2018), for whom performance of the autistic group was overall low and significantly below NT controls. However, this may relate to differences in matching procedure as their participants were matched on chronological rather than verbal mental age.

In contrast to picture selection results, we observed differences in the pattern of gaze behaviour during the unfolding of the target word in the metaphoric condition. Although both groups attend to the correct metaphoric referent during the target window, there is a significant difference in performance between groups in the metaphoric condition. Taking this result and the gaze trajectories depicted in Figures 2 and 3 into account, it seems that autistic children appear to display a less clear preference in the metaphoric condition as their looks fluctuate more between the competing literal and metaphoric referent pictures. However, as mentioned, the significance of the result is marginal, thus further research is warranted to replicate and strengthen these findings.

Nonetheless, these outcomes also fall in line with gaze performance data seen in Vulchanova et al. (2019), who also found signs of greater deliberation between the correct metaphorical image and the one depicting the literal interpretation in the autistic group. Moreover, as reported in Vulchanova et al. (2019), gaze fluctuations between literal and metaphoric pictures appear to pertain to early moments of processing, and as the decision moment comes closer, these fluctuations dissipate, and the autistic children end up choosing the correct option. This can be seen when analysing our question region results, in which there is no longer an interaction effect of condition and group and only a significant effect of condition remains. In this case, groups perform the same and prefer the literal option in the literal condition and the metaphorical option in the metaphorical condition.

Importantly, these results suggest a refinement to the narrative of the weight of language level in metaphor comprehension in autism: even though we may expect not to see group differences in accuracy, there are still differences only revealed in measures that tap into the processing of metaphoric sentences. Such differences in processing do not hinder comprehension of the metaphorical interpretations, at least in a task like the current one, suggesting processing differences that do not amount to impairment. The observed processing differences may also help us account for why comprehension of metaphor may be more challenging outside of the lab environment (Vicente et al., 2023) - as the experimental set up controls for world knowledge, discourse length, and offers visual support etc., all of which may optimize comprehension among autistic individuals with the required verbal abilities.

As commented in the Methods section, Martín-González et al. (2024) evaluated novel metaphor comprehension among a large sample of NT children. This allows us to descriptively compare the autistic sample not only to their peers, matched on verbal mental age but also to the age cohorts reported in the cross-sectional developmental study, and evaluate the hypothesis that metaphoric comprehension may be delayed in autistic children with respect to NTs. Regarding this hypothesis, we can say that none of the age groups studied in Martín-González et al. exhibited the profile of autistic children in the current study, when considering picture selection and gaze data together: overall, above chance accuracy in picture selection combined with a reduced difference between conditions in gaze behaviour. In fact, gaze behaviour for the autistic group may appear most similar to the younger NT children in the 3-4 age group who also showed a reduced difference between conditions, while still looking to the novel metaphoric image significantly more than the literal image in the target region. However, the 3- to 4-year-old NT children selected the literal option in the metaphoric condition more often than the metaphorical option. Regarding picture selection, the autistic group appears more akin to the older age ranges – 6 and older -- on par with the autistic groups' average VMA. This descriptive comparison to results from Martín-González et al. (2024) suggests that the overall profile of performance in the autistic group is not indicative of delay. The results contribute to the wider debate that we should not only consider the question of impairment or delay when conceiving of group differences. Metaphor processing could be different in the autistic population without necessarily implying an impairment or a delayed comprehension ability.

What could be the nature of this difference in processing between the autistic and the NT groups? Although we can infer that there may be differences in underlying processing, we cannot describe with any certainty the nature of how processing differs between groups, which is a limitation of the current study. It is possible that autistic participants have a reduced degree of certainty regarding non-literal interpretations of linguistic utterances (Vicente et al. 2023). According to some views, autistic individuals exhibit a higher degree of uncertainty than NT individuals across the board (Bervoets et al., 2021). Language processing should not be an exception to such a difference, especially when non-explicit linguistic communication is involved. Alternatively, these results are also compatible with an actual difference in processing. Theories of compensation effects in autism suggest that autistic individuals may resort to alternative or compensatory strategies to achieve the same means (Livingston & Happé, 2017). For example, we may find relations between language performance in autism and non-linguistic general cognitive skills, which would be unexpected for typical development. Standardized measures of non-verbal reasoning correlate with performance on standardized language assessments in autism (Faerman et al., 2023). Furthermore, non-verbal reasoning can have a predictive effect in autism performance in linguistic tasks, but not in NT teens, suggesting autistic participants may be relying on alternative cognitive resources to achieve performance similar to NT peers (Jensen de López et al., 2018). This points to the importance of considering implicit and explicit results in tandem and that conclusions on what is or what is not affected in autism may not necessarily align with what is or is not impaired in autism.

Conclusions

The purpose of this paper was to discuss empirical data concerning the comprehension of novel metaphor in verbal-age matched autistic and non-autistic children, collected through a paradigm that combined a picture-selection task and eye-tracking measures. In a nutshell, we encountered no differences in picture selection between the two developmental groups; nevertheless, we did encounter a marginal difference in processing profile, specifically in the target word region for the metaphorical condition. Our study brings to the table how possible differences in processing styles might be concealed by some methodologies. Differences can be found when analysing gaze movements that might reveal different processing profiles in autistic children when facing innovative metaphorical uses of language, even though in the end their accuracy does not differ from verbal age matched controls. While the results seem to rule out impairment for participants of the profile tested here, they suggest that autistic children might still be different from their NT peers in novel metaphor processing. However, the nature of such differences remains an open question, and this can be seen as a limitation of the current study.

Our design was focused on exploring potential differences in both processing and picture selection between autistic children and verbal age matched NT peers; rather than on unveiling the source of differences, if there were any. In spite of this, we may speculate two potential explanations of our findings, in line with relevant theoretical and experimental literature. We have suggested that autistic children may experience more uncertainty than NT children, which would make them waver between the literal and the metaphorical interpretation in a way that NT children do not. The compensation alternative, which we also consider, is that autistic children process metaphors differently. What seems to be “easy” in the case of NT children may require from autistic children more reasoning or effortful processing in any case. Such effortful processing may involve going back and forth between the literal and the metaphorical interpretation.

Be that as it may, the observed processing differences can help us achieve a more nuanced conception of how autistic traits impact everyday communication, and connect with testimonies from outside the lab, in which many autistic individuals report experiencing challenges with different kinds of figurative language. All this offers interesting paths for investigation on the subject to move forward.

References

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology, 54*(3), 218–250. https://doi.org/10.1016/j.cogpsych.2006.07.001

Andrés-Roqueta, C., & Katsos, N. (2017). The Contribution of Grammar, Vocabulary and Theory of Mind in Pragmatic Language Competence in Children with Autistic Spectrum Disorders. *Frontiers in Psychology, 8*, 996. https://doi.org/10.3389/fpsyg.2017.00996

Andrés-Roqueta, C., & Katsos, N. (2020). A distinction between linguistic and social pragmatics helps the precise characterization of pragmatic challenges in children with autism spectrum disorders and developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(5), 1494, 1598. https://doi.org/10.1044/2020_JSLHR-19-00263

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bervoets, J., Milton, D., & Van de Cruys, S. (2021). Autism and intolerance of uncertainty: An ill-fitting pair. *Trends in Cognitive Sciences, 25*(12), 1009-1010. https://doi.org/10.1016/j.tics.2021.08.006

Chahboun, S., Vulchanov, V., Saldana, D., Eshuis, H., & Vulchanova, M. (2017). Can you tell it by the prime? A study of metaphorical priming in high-functioning autism in comparison with matched controls. *International Journal of Language & Communication Disorders, 52*(6), 766-785. https://doi.org/10.1111/1460-6984.12314

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar Inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 40*(9), 1104-1117. https://doi.org/10.1007/s10803-010-0960-8

Chouinard, B., & Cummine, J. (2016). All the world'sa stage: Evaluation of two stages of metaphor comprehension in people with autism spectrum disorder. *Research in Autism Spectrum Disorders, 23*, 107-121. https://doi.org/10.1016/j.rasd.2015.12.008

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Deliens, G., Antoniou, K., Clin, E., & Kissine, M. (2017). Perspective-taking and frugal strategies: Evidence from sarcasm detection. *Journal of Pragmatics, 119,* 33-45. https://doi.org/10.1016/j.pragma.2017.08.002

Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test (3rd ed.). American Guidance Service.

Faerman, A., Sakallah, A., Skiba, S., Kansara, S., Kopald, B. E., Lewine, J. D., & Demopoulos, C. (2023). Language Abilities are Associated with Both Verbal and Nonverbal Intelligence in Children on the Autism Spectrum. *Developmental Neuropsychology, 0* 101-119. https://doi.org/10.1080/87565641.2023.2225663

Gold, R., & Faust, M. (2010). Right hemisphere dysfunction and metaphor comprehension in young adults with Asperger syndrome. *Journal of Autism and Developmental Disorders, 40,* 800-811. https://doi.org/10.1007/s10803-009-0930-1

Gold, R., Faust, M., & Goldstein, A. (2010). Semantic integration during metaphor comprehension in Asperger syndrome. *Brain and Language, 113,* 124-134. https://doi.org/10.1016/j.bandl.2010.03.002

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101-119. https://doi.org/10.1016/0010-0277(93)90026-r

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar Implicature in Absence of Epistemic Reasoning? The Case of Autism Spectrum Disorder. *Language Learning and Development, 14*(3), 224-240. https://doi.org/10.1080/15475441.2017.1343670

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151–171. doi: 10.1016/j.actpsy.2010.11.003

Jensen de López, K., Schroeder, K., & Gavarró, A. (2018). Successful passive sentence comprehension among Danish adolescents with autism spectrum disorders. *Autism & Developmental Language Impairments, 3.* https://doi.org/10.1177/2396941518761239

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

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism, 22*(2). https://doi.org/10.1177/1362361316668652

Kissine, M. (2016). Pragmatics as Metacognitive Control. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.02057

Kissine, M., Cano Chervel, J., Carlier, S., De Brabanter, P., Ducenne, L., Pairon, M.-C., Deconinck, N., Delvenne, V., & Leybaert, J. (2015). Children with Autism Understand Indirect Speech Acts: Evidence from a Semi-Structured Act-Out Task. *PloS One, 10*(11). https://doi.org/10.1371/journal.pone.0142191

Koch, C., Miller, L. J., Pomplun, M., & Roid, G. H. (2019). *LEITER-3. Escala de Inteligencia No Verbal.* Giunti Psychometrics.

Lampri, S., Peristeri, E., Marinis, T., & Andreou, M. (2023). Figurative language processing in autism spectrum disorders: A review. *Autism Research, n/a* https://doi.org/10.1002/aur.3069

Lenth, R. V., Love, J., Herve, M. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.0. Retrieved from https://CRAN.R-project.org/package=emmeans

Li, J. P., Law, T., Lam, G. Y., & To, C. K. (2012). Role of sentence-final particles and prosody in irony comprehension in Cantonese-speaking children with and without autism spectrum disorders. *Clinical Linguistics & Phonetics, 27*(1), 18–32. https://doi.org/10.3109/02699206.2012.734893

Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience & Biobehavioural Reviews 80*, 729-742. https://doi.org/10.1016/j.neubiorev.2017.06.005

Lord, C., Rutter, M., DiLavore, P. S., Risi, S., Gotham, K., & Bishop, S. (2012). *(ADOS®-2) Autism Diagnostic Observation Schedule™, Second Edition* https://www.wpspublish.com/ados-2-autism-diagnostic-observation-schedule-second-edition

MacKay, G., & Shaw, A. (2004). A comparative study of figurative language in children with autistic spectrum disorders. *Child Language Teaching and Therapy, 20*(1), 13-32. https://doi.org/10.1191/0265659004ct261oa

Marocchini, E., Di Paola, S., Mazzaggio, G., & Domaneschi, F. (2022). Understanding indirect requests for information in high-functioning autism. *Cognitive Processing, 23*(1), 129-153.

https://doi.org/10.1007/s10339-021-01056-z

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

Mazzaggio, G., Foppolo, F., Job, R., & Surian, L. (2021). Ad-hoc and scalar implicatures in children with autism spectrum disorder. *Journal of Communication Disorders, 90*. https://doi.org/10.1016/j.jcomdis.2021.106089

McMahon, C. M., Lerner, M. D., & Britton, N. (2013). Group-based social skills interventions for adolescents with higher-functioning autism spectrum disorder: A review and looking to the future. *Adolescent Health, Medicine and Therapeutics*, 23–38. https://doi.org/10.2147/AHMT.S25402

Melogno, S., Pinto, M. A., & Orsolini, M. (2017). Novel Metaphors Comprehension in a Child with High-Functioning Autism Spectrum Disorder: A Study on Assessment and Treatment. *Frontiers in Psychology, 7* https://doi.org/10.3389/fpsyg.2016.02004

Morra, L. (2016). Raising awareness of how Asperger persons perceive their capacity to use Metaphors. *Medicina e Stor*ia, *16*, 129–146.

Morsanyi, K., Stamenković, D., & Holyoak, K. J. (2020). Metaphor processing in autism: A systematic review and meta-analysis. *Developmental Review, 57*(August), 100925. https://doi.org/10.1016/j.dr.2020.100925

Norbury, C. Frazier. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology, 23*(3), 383-399. https://doi.org/10.1348/026151005X26732

Ozonoff, S., & Miller, J. N. (1996). An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language, 52*(3), 411-434. https://doi.org/10.1006/brln.1996.0022

Panzeri, F., Mazzaggio, G., Giustolisi, B., Silleresi, S., & Surian, L. (2022). The atypical pattern of irony comprehension in autistic children. *Applied Psycholinguistics, 43*(4), 757-784. https://doi.org/10.1017/S0142716422000091

Pastor-Cerezuela, G., Fernández-Andrés, M.-I., Tordera-Yllescas, J. C., & González-Sala, F. (2020). Metaphor comprehension in children with and without autism spectrum disorder. *Research in Autism Spectrum Disorders, 76,* 101588. https://doi.org/10.1016/j.rasd.2020.101588

Paul, R., & Cohen, D. J. (1985). Comprehension of indirect requests in adults with autistic disorders and mental retardation. *Journal of Speech, Language, and Hearing Research, 28*(4), 475-479. https://doi.org/10.1044/jshr.2804.475

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.-P., & Geurts, B. (2009). Pragmatic Inferences in High-Functioning Adults with Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders, 39*(4), 607-618. https://doi.org/10.1007/s10803-008-0661-8

Pouscoulous, N. (2014). The elevator's buttocks. *Metaphorical Abilities in Children. In D. Matthews (Ed.), Pragmatic Development in First Language Acquisition*, 239–259.

Roid, G. H., & Miller, L. J. (2013). *(Leiter-3) Leiter International Performance Scale, Third Edition*. WPS.

Ronderos, C. R., Guerra, E., & Knoeferle, P. (2022). When sequence matters: The processing of contextually biased German verb–object metaphors. *Language and Cognition*, 1–28.
doi:10.1017/langcog.2022.22

Rundblad, G., & Annaz, D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism, 14*(1), 29–46. https://doi.org/10.1177/1362361309340667

Saban-Bezalel, R., & Mashal, N. (2019). Different Factors Predict Idiom Comprehension in Children and Adolescents with ASD and Typical Development. *Journal of Autism and Developmental Disorders, 49*(12), 4440, 4450. https://doi.org/10.1007/s10803-019-04193-9

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The Understanding of Scalar Implicatures in Children With Autism Spectrum Disorder: Dichotomized Responses to Violations of Informativeness. *Frontiers in Psychology, 9,* 1266. https://doi.org/10.3389/fpsyg.2018.01266

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2022). afex: Analysis of Factorial Experiments. R package version 1.3.0. Retrieved from https://CRAN.R-project.org/package=afex

Song, Y., Nie, Z., & Shan, J. (2023). Comprehension of irony in autistic children: the role of theory of mind and executive function. *Autism Research, 17*(1), 109–124. https://doi.org/10.1002/aur.3051

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science, 268*(5217), 1632–1634. https://doi.org/10.1007/s13164-023-00704-x

Su, Y., & Su, L.-Y. (2015). Interpretation of logical words in Mandarin-speaking children with autism spectrum disorders: Uncovering knowledge of semantics and pragmatics. *Journal of Autism and Developmental Disorders, 45*, 1938-1950. https://doi.org/10.1007/s10803-014-2350-0

Van Herwegen, J., & Rundblad, G. (2018). A Cross-Sectional and Longitudinal Study of Novel Metaphor and Metonymy Comprehension in Children, Adolescents, and Adults With Autism Spectrum Disorder. *Frontiers in Psychology, 9*, 945. https://doi.org/10.3389/fpsyg.2018.00945

Van Tiel, B., & Kissine, M. (2018). Quantity-based reasoning in the broader autism phenotype: A web-based study. *Applied Psycholinguistics, 39*(6), 1373-1403. https://doi.org/10.1017/S014271641800036X

Vicente, A., Michel, C., & Petrolini, V. (2023). Literalism in Autistic People: A Predictive Processing Proposal. *Review of Philosophy and Psychology*, 1–24. https://doi.org/10.1007/s13164-023-00704-x

Vulchanova, M., Chahboun, S., Galindo-Prieto, B., & Vulchanov, V. (2019). Gaze and Motor Traces of Language Processing: Evidence from Autism Spectrum Disorders in Comparison to Typical Controls. *Cognitive Neuropsychology, 36*(7-8), 383-409. https://doi.org/10.1080/02643294.2019.1652155

Whyte, E. M., Nelson, K. E., & Scherf, K. S. (2014). Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 57*(5), 1839–1851. https://doi.org/10.1044/1092-4388(2013/12-0308)

Zhan, L. (2018). Using eye movements recorded in the visual world paradigm to explore the online processing of spoken language. *Journal of Visualized Experiments, 140*, e58086. doi:10.3791/58086

Zheng, Q., Jia, Z., & Liang, D. (2015). Metaphor and metonymy comprehension in Chinese-speaking children with high-functioning autism. *Research in Autism Spectrum Disorders, 10*, 51–58. https://doi.org/10.1016/j.rasd.2014.11.007

Data, code and materials availability statement

The relevant datasets with information (anonymized) about participants, the scripts used for the analysis, the materials for building the task and instructions on how to use all these files (ReadMe) are available on:
https://osf.io/ksuwv/?view_only=accde2b9d44b49eea164e75dce89796a
The experimental materials are also detailed in Appendix 1 in this manuscript.

Ethics statement

Ethical approval was issued by the University of the Basque Country's Ethics Committee for research with human beings. All parents/guardians gave signed consent prior to the participation in the investigation.

Authorship and Contributorship Statement

IMG conceived and designed the task. All authors contributed to designing the actual experiment and its implementation. KS, ILF and EC gave useful input about methodology and AV contributed the theoretical framework. AV helped in selecting participants and reached out for public schools to recruit NT children. IMG & KS collected the data. IMG & KS cleaned and reviewed the data. ILF provided tools and strategy to analyse the data. IMG analysed the data. All authors contributed to interpret the results. All authors drafted the paper. ILF, EC and AV contributed to fundraising. A special contribution is that of Dr. Camilo Rodríguez Ronderos who participated in the work with NT children and supervised and helped IMG in the data analysis for this particular study. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

We would like to warmly thank Adurtza Ikastola and Arantzabela Ikastola, two schools from Vitoria-Gasteiz (Basque Country, Spain), for their invaluable collaboration in organizing the testing of typically developing children. We are equally grateful to the families and the autistic children who participated in the study at our lab—their involvement was essential to understanding the specific needs of each child and ensuring a calm environment during eye-tracking sessions.
Special thanks go to Dr. Camilo Rodríguez-Ronderos, whose close collaboration with the first author was instrumental in the development of this work. We are also grateful to the DEVCOM group at the University of Oslo for their constructive feedback throughout the project.

We thank Dr. Mikhail Kissine and Dr. Inge-Marie Eigsti for their critical insights, which greatly contributed to the early stages of this research.
We also appreciate the feedback received during conference presentations, especially at the International Pragmatics Association Conference (IPRA 2023) and the

International Association for Child Language Conference (IASCL 2024)—in particular, the audience and organizers of the “Language and Communication in Children with Neurodiverse Profiles” symposium. Finally, we acknowledge the helpful critiques shared at DEVCOM’s closing conference, *The Developing Communicator* (Oslo, May 2024), which helped refine the final version of this work.

We are also grateful to the anonymous reviewers of this manuscript for their thoughtful and constructive comments.

Funding

This research was partially supported by the project grant PID2021-122233OB-I00 and the pre-doctoral grant PRE2019-089758, funded by the Spanish Ministry of Science and Innovation and the Spanish Research Agency (MICIU/AEI/10.13039/501100011033), with co-funding from the European Regional Development Fund (ERDF/EU), and by the IT1537-22 Research Group (Basque Government).

Additional support was provided by the BBVA Foundation Grants for Scientific Research Projects 2021 (RILITEA). The BBVA Foundation bears no responsibility for the opinions, statements, or content of this publication, which are solely those of the authors.

This study is part of a project funded by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 853211).

Finally, the first author received support from the Centre for Philosophy and the Sciences, University of Oslo, through a Visiting Scholar Scholarship (June 2022).

License

Appendix 1

Experimental materials (with mean metaphor novelty scores in parenthesis)

1. Squirrels (2.27)

1 met. Squirrels climb up trees. That boy is a squirrel. Which one is it?
Las ardillas suben a los árboles. Ese niño es una ardilla. ¿Cuál es?
1 lit. Squirrels climb up trees. That animal is a squirrel. Which one is it?
Las ardillas suben árboles. Ese animal es una ardilla. ¿Cuál es?

Literal option: a squirrel climbing up a tree.	Distractor-literal: a cat jumping to the top of a wall.
Distractor-metaphorical: a boy climbing up a metal ladder	Metaphorical option: a boy climbing up a tree.

2. Swordfish (1.12)

2 met. Swordfish swim fast. That boy is a swordfish. Which one is it?
Los peces espada nadan rápido. Ese niño es un pez espada. ¿Cuál es?
2 lit. Swordfish swim fast. That animal is a swordfish. Which one?
Los peces espada nadan rápido. Ese animal es un pez espada. ¿Cuál es?

Literal option: a swordfish	Distractor-literal: a tiger
Distractor-metaphorical: a boy at the beach	Metaphorical option: a boy swimming

3. *Grasshoppers* (2.69)

3 met. Grasshoppers jump a lot. That boy is a grasshopper. Which one is it?
Los saltamontes saltan mucho. Ese niño es un saltamontes. ¿Cuál es?
3 lit. Grasshoppers jump a lot. That boy is a grasshopper. Which one is it?
Los saltamontes saltan mucho. Ese animal es un saltamontes. ¿Cuál es?

Literal option: a grasshopper jumping	Distractor - literal. a beatle jumping
Distractor-metaphorical: a boy running	Metaphorical option: a boy jumping

4. *Rhinos* (1.81)

4 met.Rhinos are strong. That boy is a rhino. Which one is it?

Los rinocerontes son fuertes. Ese niño es un rinoceronte. ¿Cuál es?
4 lit. Rhinos are strong. That animal is a rhino. Which one is it?
Los rinocerontes son fuertes. Ese animal es un rinoceronte. ¿Cuál es?

Literal option: a rhino	Distractor - literal. A giraffe
Distractor-metaphorical: a boy playing football with friends	Metaphorical option: a very big boy with wide shoulders and strong, flexed arms. He is standing between two slim children, ensuring a noticeable contrast

5. *Crocodiles* (1.42)

5 met. Crocodiles have big teeth. That boy is a crocodile. Which one is it?
Los cocodrilos tienen los dientes grandes. Ese niño es un cocodrilo. ¿Cuál es?
5 lit. Crocodiles have big teeth. That animal is a crocodile. Which one is it?
Los cocodrilos tienen los dientes grandes. Ese animal es un cocodrilo. ¿Cuál es?

Literal option: a crocodile with open mouth wide (showing big, sharp teeth) trying to bite a beach ball	Distractor-literal: a tiger with big teeth trying to bite a beach ball
Distractor-metaphorical: the face of a boy with long hair	Metaphorical option: the face of a boy with a wide smiling grin

6. *Sheep* (2.31)

6 met. Sheep have a lot of wool. That boy is a sheep. Which one is it?
Las ovejas tienen mucha lana. Ese niño es una oveja. ¿Cuál es?
6 lit. Sheep have a lot of wool. That animal is a sheep. Which one is it?
Las ovejas tienen mucha lana. Ese animal es una oveja. ¿Cuál es?

Literal option: a sheep with a thick wool fleece.	Distractor-literal: a cow
Distractor-metaphorical: a boy wearing lea jacket and cowboy boots.	Metaphorical option: A boy wearing thick white, fleece clothing (fuzzy white fleece sweater, hat and mittens).

7. *Leopards* (1.65)

7 met. Leopards have a lot of dots. That boy is a leopard! Which one is it?
Los leopardos tienen muchos puntos. Ese niño es un leopardo. ¿Cuál es?
7Lit. Leopards have a lot of dots. That boy is a leopard!
Los leopardos tienen muchos puntos. Ese animal es un leopardo. ¿Cuál es?

Literal option: a leopard	Distractor-literal: a zebra
Distractor-metaphorical: the face of a boy wearing sunglasses.	Metaphorical option: a boy with noticeable freckles on his face

8. *Hippos* (2.35)

8 met. Hippos have a very big mouth. That kid is a hippo! Which one is it?
Los hipopótamos tienen la boca muy grande. Ese niño es un hipopótamo! ¿Cuál es?
8 lit. Hippos have a very big mouth. That animal is a hippo! Which one is it?
Los hipopótamos tienen la boca muy grande. Ese animal es un hipopótamo! ¿Cuál es?

28. Literal option: a hippopotamus with mouth wide open	29. Distractor-literal: a pelican with mouth wide open.
30. Distractor-metaphorical: The face of a boy whistling	31. Metaphorical option: The face of boy with mouth wide open, looking like he's laughing really hard

9. *Chicks* (2.85)

9a. Chicks always follow their mum. That boy is a chick! Which one is it?
Los pollitos siempre van con su mamá. Ese niño es un pollito. ¿Cuál es?
9b. Chicks always follow their mum. That animal is a chick! Which one is it?
Los pollitos siempre van con su mamá. Ese animal es un pollito. ¿Cuál es?

Literal option: Chick walking next to her mother hen	Distractor-literal: two adult cats sleeping together
Distractor-metaphorical: a boy and a girl facing each other	Metaphorical option: A little boy walking next to his mother, following her by the hand, resembling the hen and chicken picture

10. *Flamingos* (2.38)

10 met. Flamingos are pink. That kid is a flamingo! Which one is it?
Los flamencos son rosa. Ese niño es un flamenco. ¿Cuál es?

10lit. Flamingos are pink. That animal is a flamingo! Which one is it?
Los flamencos son rosa. Ese animal es un flamenco. ¿Cuál es?

Literal option: a flamingo	Distractor-literal: a yellow duck
Distractor-metaphorical: a blond-haired boy who is very pale, lips bluish from the cold	Metaphorical option: a boy with very pink-tone skin.

11. *Caterpillars* (1.42)

11 met. Caterpillars inch on the ground. That kid is a caterpillar! Which one is it?
Las orugas van por el suelo. Ese niño es una oruga. ¿Cuál es?
11 lit. Caterpillars inch on the ground. That animal is a caterpillar! Which one is it?
Las orugas van por el suelo. Ese animal es una oruga. ¿Cuál es?

Literal option: a caterpillar inching on the ground	Distractor-literal: a frog jumping
Distractor-metaphorical: a boy stretching	Metaphorical option: a boy 'inching' like a caterpillar on the ground

12. *Vacuum cleaners* (2.35)

12 met.Vacuum cleaners absorb everything. That kid is a vacuum cleaner! Which one is it?
Las aspiradoras absorben todo. Ese niño es una aspiradora. ¿Cuál es?
12 lit. Vacuum cleaners absorb everything. That thing is a vacuum cleaner! Which one is it?
Las aspiradoras absorben todo. Esa cosa es una aspiradora. ¿Cuál es?

Literal option: a man vacuuming some dirt.	Distractor-literal: a man pushing a shopping cart.

Distractor-metaphorical: a boy eating a salad with a fork.	47. Metaphorical option: A boy eating a plate of spaghetti, slurping a plate of noodles into his mouth.

13. *Skyscrapers* (2.35)
13 met. Skyscrapers are tall. That kid is a skyscraper! Which one is it?
Los rascacielos son altos. Ese niño es un rascacielos. ¿Cuál es?
13 lit. Skyscrapers are tall. That thing is a skyscraper! Which one is it?
Los rascacielos son altos. Esa cosa es un rascacielos. ¿Cuál es?

Literal option: a skyscraper next to several small houses	Distractor-literal: a fence
Distractor-metaphorical: three kids sitting in chairs talking	Metaphorical option: A very tall boy standing between two shorter kids

14. *Balloons* (1.65)
14 met. Balloons are round. That kid is a balloon. Which one is it?
Los globos son redondos. Ese niño es un globo. ¿Cuál es?
14 lit. Balloons are round. That thing is a balloon. Which one is it?
Los globos son redondos. Esa cosa es un globo. ¿Cuál es?

Literal option: An orange balloon	Distractor-literal: a soap bubble
Distractor-metaphorical: An average size boy, wearing normal clothes.	Metaphorical option: a boy in a puffy snowsuit

15. *Computers* (2.15)
15 met. Computers do math very fast. That kid is a computer! Which one is it?
Los ordenadores hacen mates muy rápido. Ese niño es un ordenador. ¿Cuál es?
15 lit. Computers do math very fast. That thing is a computer! Which one is it?
Los ordenadores hacen mates muy rápido. Esa cosa es un ordenador. ¿Cuál es?

Literal option: a desktop computer with spreadsheet on the screen	Distractor-literal: a blender
Distractor-metaphorical: a boy sitting at a desk, painting	Metaphorical option: a boy sitting at a desk, working on an assignment. He is depicted having a thought bubble with an arithmetic equation.

16. *Strawberries* (1.35)

16 met. Strawberries are red. That kid is a strawberry. Which one is it?
las fresas son rojas. Ese niño es una fresa. ¿Cuál es?
16 lit. Strawberries are red. That thing is a strawberry. Which one is it?
Las fresas son rojas. Esa cosa es una fresa. ¿Cuál es?

Literal option: a strawberry	Distractor-literal: a green pepper
Distractor-metaphorical: The face of a boy with normal pale cheeks.	Metaphorical option: The face of a boy who has very red cheeks

17. *Rivers* (1.27)

17 met. Rivers have tons of water. That kid is a river. Which one is it?
Los ríos tienen mucha agua. Ese niño es un río. ¿Cuál es?
17 lit. Rivers have tons of water. That thing is a river. Which one is it?
Los ríos tienen mucha agua. Esa cosa es un río. ¿Cuál es?

Literal option: a river	Distractor-literal: a road
Distractor-metaphorical: a boy	Metaphorical option: a boy wearing soaking-wet clothes, standing in a puddle with drops of water falling to the floor

18. *Spinning tops* (2.46)

18 met. Spinning tops spin a lot. That kid is a spinning top. Which one is it?
Las peonzas dan muchas vueltas. Ese niño es una peonza. ¿Cuál es?
18 lit. Spinning tops spin a lot. That thing is a spinning top. Which one is it?
Las peonzas dan muchas vueltas. Esa cosa es una peonza. ¿Cuál es?

Literal option: a spinning top	Distractor-literal: a boy going down a slide

Distractor-metaphorical: a spinning sewing wheel	Metaphorical option: a boy doing a cartwheel.

19. *Scissors* (1.08)

19 met. Scissors cut stuff. That kid is a scissor! Which is it?
Las tijeras cortan cosas. Ese niño es una tijera. ¿Cuál es?
19 lit. Scissors cut stuff. That thing is a scissor! Which is it?
Las tijeras cortan cosas. Esa cosa es una tijera. ¿Cuál es?

Literal option: a pair of scissors cutting a big piece of paper	Distractor-literal: a hammer, hammering a nail
Distractor-metaphorical: a child carrying some boxes	Metaphorical option: a boy tearing a big piece of paper with his hands.

20. *Race cars* (1.92)

20 met. Race cars go very fast. That kid is a racing car! Which one is it?
Los coches de carreras van muy rápido. Ese niño es un coche de carreras. ¿Cuál es?
20 lit. Race cars go very fast. That thing is a racing car! Which one is it?
Los coches de carreras van muy rápido. Esa cosa es un coche de carreras. ¿Cuál es?

Literal option: a race car	Distractor-literal: a skate-board
Distractor-metaphorical: a boy doing a handstand	Metaphorical option: a boy running with a sprinting gait.

Not early, not late, but developing: Children's “good-enough” understanding of metaphors

Sarah Ferrara
Université de Reims Champagne-Ardenne, C2S, Reims, France

Marc Aguert
Normandie Univ, UNICAEN, LPCN, France

Christelle Declercq
Université de Reims Champagne-Ardenne, C2S, Reims, France

Abstract: To date, the debate over the age at which children begin to understand metaphors remains unresolved. Do children begin to acquire comprehension early, around age 3 to 4, or later, around age 8? One way to answer this question is to use the notion of "good-enough" comprehension proposed by Ferreira et al. (2002) and to hypothesize that young children understand metaphors in a “good enough” manner while older children understand them in a more precise and accurate manner. This hypothesis was tested using a task where children were asked to assess the extent to which more or less precise and relevant rephrasings resembled nominal metaphors. We therefore sought to (1) differentiate between "good-enough" and "good" (precise and accurate) understanding in school-age children and (2) show that the former appears earlier than the latter during development. Data collected from 300 children aged 5 to 11 suggest that both goals were reached. These results suggest that, while metaphorical abilities emerge early, comprehension processes then evolve during childhood, with a refinement of understanding between 7 and 9 years old. These results may open a path to reconcile the proponents of early acquisition with those of later acquisition.

Keywords: cognitive development; metaphor comprehension; language development

Corresponding author(s): Sarah Ferrara, Laboratoire C2S (Cognition, Santé, Société), Université de Reims Champagne-Ardenne, 57 rue Pierre Taittinger, 51571 Reims Cedex, France. Email: sarah.ferrara@univ-reims.fr

ORCID ID(s): Sarah Ferrara: https://orcid.org/0000-0002-3953-0661; Marc Aguert: https://orcid.org/0000-0002-0484-0328; Christelle Declercq: https://orcid.org/0000-0002-3242-644X

Citation: Ferrara, S., Aguert, M., & Declercq, C. (2025). Not early, not late, but developing: Children's “good-enough” understanding of metaphors. *Language Development Research*, *5*(2), 67–92. http://doi.org/10.34842/ldr2025-806

Introduction

Metaphor comprehension in children has been the subject of research for decades, but the field lacks a coherent account of how metaphor comprehension abilities develop throughout childhood due to the diversity of theoretical and methodological approaches. This study aimed to examine the development of metaphor understanding in children aged 5 to 11 years old using the theory of "good-enough" comprehension (Ferreira et al., 2002). We hypothesised that understanding could fall outside of the dichotomy of being either "correct" or "incorrect": some understanding could be correct "enough" to ensure the conversational flow, while still being incomplete and imprecise.

Metaphors are widely used utterances describing one thing using another element that is conceptually different. For instance, in the metaphor "cold air is a needle", the speaker communicates something about the topic (e.g., cold air) by means of the vehicle (e.g., a needle) in a non-literal way to describe the hurtful nature of coldness (Di Paola et al., 2019). In the developmental literature, interest has long been focused on the age at which children begin to understand metaphors, with mixed findings. On one hand, some studies have concluded that children begin to understand metaphors rather late, around 8 to 10 years of age, with possible further developments in adolescence, thus supporting the notion that metaphors are complex statements to comprehend (Billow, 1977; Carriedo et al., 2016; Cometa & Eson, 1978; Deckert et al., 2019; Willinger et al., 2019). On the other hand, alternative studies have argued that metaphors are not inherently complex statements to understand. These studies have empirically demonstrated an earlier comprehension, around 3 to 4 years of age, with a methodology adapted to young children, including facilitating contexts and age-appropriate vocabulary (Vosniadou, 1987 for a review; Pouscoulous et al., 2011).

This divergence has long been explained with reference to the method used to assess understanding. In particular, the commonly used method of asking children to verbally explain the meaning of the metaphor has been criticised as being too cognitively costly, falling more within the realm of metalinguistic abilities (Vosniadou, 1987; Pouscoulous, 2011). Other methods that do not involve verbalisation have been developed to address this criticism, including selecting from multiple rephrasings of the metaphor (Nippold et al., 1984; Nippold & Sullivan, 1987; Winner et al., 1980) matching sentences with metaphorically related pictures (Kogan, 1980) , choosing appropriate metaphorical completions for sentences (Gardner et al., 1975), or performing an action such as an enactment (Vosniadou & Orthony, 1986). Recently, Pouscoulous and Tomasello (2020) found that children as young as 4 years old showed an understanding of referential metaphors by being able to choose the right puppet when asked questions such as, "Show me the dog with brown socks" (e.g. the dog with brown feet, while also being presented with a dog with a brown bow).

The variety of methodologies in the literature likely reflects different conceptions of what it means to understand a metaphor, which extends from describing the elementary processes enabling the understanding of a basic metaphor to portraying what children understand about the variety of metaphors that surround them in their natural environment. In addition, multiple research aims are associated with this question, such as the description, mostly through correlational studies, of factors influencing children's understanding of metaphors. Some factors have been considered as specific to metaphorical abilities (e.g., classificatory skills or analogical reasoning are precisely what enable metaphor comprehension, Billow, 1977; Johnson & Pascual-Leone, 1989; Gentner, 1988), while others have been thought to be more broadly related to cognitive development (e.g. vocabulary, executive functions or theory of mind abilities, Białecka-Pikul, 2010; Del Sette et al., 2020; Huang et al., 2015; Pierce et al., 2010; Rundblad & Annaz, 2010; Tonini et al., 2023; Willinger et al., 2019). Overall, these results depict metaphors as utterances that are most often intricate, and understanding them involves multiple social cognitive constructs.

Based on the previous data, one could consider that metaphor comprehension emerges early in some simple situations but continues to progress during childhood. This would offer a means to reconcile findings that demonstrate early metaphorical abilities with those indicating the influence of various factors associated with cognitive development. However, there is an intensive body of research on the development of metaphor comprehension and conclusions often refers to understanding as a process ending in an understanding or an incorrect one. Understanding is then considered as something perfect or described as attained by some children if not all of an age group. Some authors have implied an intermediate level without really seeking theoretical implications (Deckert, 2019 ; Lecce at al., 2019). An alternative stance would be to consider that, between an undeniable understanding or an understanding in progress, one might find a formalized intermediate level of comprehension. Children could strive to grasp the overall meaning of what is being said by relying on the whole situation, rather than undertaking a detailed analysis of the metaphor's meaning. It is this stance that we aimed to explore further, assuming that metaphor comprehension falls on a spectrum rather than a binary categorisation of good or bad. This hypothesis is grounded in the “good-enough” theory introduced by Ferreira et al. (2002).

Ferreira et al. (2002) posit the existence of different levels of comprehension within their "good-enough" hypothesis. They define “good-enough comprehension” as the creation of a vague and superficial yet largely satisfactory understanding of the meaning of a statement. An illustrative example is the Moses illusion (Erickson & Mattson, 1981, cited by Ferreira et al., 2002). When people are asked, "How many animals of each sort did Moses put on the ark?" they usually answer "two" instead of the correct answer, "zero," as it was Noah who placed the animals on the ark. This example demonstrates how adults, despite their capacity for precise comprehension, can construct an imprecise and shallow representation of a statement. Ferreira et al. (2002) proposed that good-enough comprehension is commonplace, as it reduces cognitive effort in everyday interactions where people communicate without investing the time needed for precise comprehension. In this view, relying on a good-enough mode of comprehension could be more efficient than exerting the effort to achieve precise comprehension.

Good-enough comprehension could be especially effective in processing complex language forms such as metaphors by economizing resources. Metaphor processing requires linguistic skills (knowing which semantic features of the topic and the vehicle are salient), pragmatic skills (using language in an unconventional way which is contextually relevant), executive resources (inhibiting literal meaning, switching meanings with flexibility), and conceptual and classificatory skills (finding common ground between the topic and the vehicle). Adults possess advanced linguistic skills, but children experience a considerable evolution in their abilities through childhood and even adolescence, often presenting them with language-related difficulties (see Gervain, 2020 for a review). Considering these inherent difficulties and the importance of verbal communication in daily life, good-enough understanding could be a significant advantage and even a preferred path to more precise and complete "good" understanding. Moreover, school-aged children also possess limited cognitive resources, and their executive functions (i.e. working memory, flexibility and inhibition) are still in development (Diamond, 2020). Good-enough understanding could also be advantageous by reducing working memory load and thus providing a possible economy in cognitive resources. In addition, studies have shown a complementary role played by vocabulary skills and working memory on understanding, thus supporting the complex relationships between these skills (Chiappe & Chiappe, 2007; Mashal, 2013; Stamenković et al., 2019). Good-enough understanding could therefore represent an advantage for children regarding these limitations.

The concept of good-enough understanding could also help make sense of conflicting findings in the literature. In regards of the theoretical and methodological divergences, we could hypothesize that authors had accounted for different levels of understanding without intending it. For instance, studies asking children to explain precisely what they understood by the metaphor could be seen as demonstrating a "good" (i.e. precise and complete) understanding of the metaphor, while studies using an action criterion for comprehension may show what could be defined as good-enough understanding. Thus, the difference observed in the results, especially regarding the emergence of comprehension, could depend not only on the nature of their methodologies but also on the type of comprehension authors were actually accounting for. Good-enough understanding could have been considered "good" by early-understanding proponents (even if this entailed glossing over certain nuances in understanding) while it would have been considered "bad" by later-understanding proponents (even if this meant overlooking early signs of understanding), in accordance with a dichotomic consideration of comprehension. At that stage, authors did not explore the theory of different levels of understanding, as their existence had not yet been hypothesised. However, certain studies have operationalised intermediate levels of understanding. For instance, in a verbal explication task of metaphors, Lecce et al. (2019) assigned an intermediate score to children who gave an "incomplete" response ("They are very good" for the metaphor "Soldiers are lions") or referred to "a non-salient feature of the metaphor vehicle" ("They run as fast as lions"). However, these intermediate levels were not correlated with any clear theoretical framework as is the case with Ferreira's good-enough understanding.

Following Ferreira et al. (2002), three levels of understanding of metaphors were distinguished in the current study: good understanding and poor understanding, as traditionally described in the literature, with the addition of an in-between level, good-enough understanding. Good understanding is defined as accurate and comprehensive, incorporating the specific reasons why the vehicle has been chosen to qualify the topic. Good-enough understanding relies on the whole situation, avoiding the cost of a fine-grained construction of the metaphor’s meaning: it enables the listener to understand the intended meaning of the metaphor, that is the communicative intent relative to the topic, in coherence with the context, but not necessarily the specific reasons why the vehicle has been chosen to represent the topic. Poor understanding does not account for the meaning of the metaphor; this type of understanding could be diverse, as children can produce multiple misinterpretations of a metaphor’s meaning, using a variety of clues from the material presented within and outside of the metaphor. Previous studies have for example presented children with literal interpretations or irrelevant context-based interpretations in multiple choice tasks (Declercq et al., 2010).

The current study aimed to test whether children showed evidence of good-enough understanding of metaphors with an innovative task involving metaphor rephrasing judgement. In each trial, participants were presented with a metaphor and a rephrasing of this metaphor varying in quality. They evaluated the quality of this rephrasing compared to the metaphor on a 5-point scale. Their evaluation of the different rephrasings would determine whether they had a good, good-enough, or poor understanding of the presented metaphor. Based on previous literature, we expected that by the age of 5, a good-enough understanding would be observed, and supplemented later in development by a good understanding. Our study aimed to (1) differentiate between good-enough understanding and good understanding in school-aged children and (2) establish a developmental pattern of both levels of understanding.

Method

Participants

We planned to include 300 children in the study, aiming to recruit between 30 and 50 participants of each age. The experimental design, which involved three crossed random variables (participants, schools, and items, analyzed using a linear mixed-effects model), did not allow for a conventional a priori power analysis (e.g., using G*Power) or a simulation-based approach (e.g., using R, *MixedPower* package). Rather than providing a power analysis that would lack interpretability, we opted for a large sample size (N = 300), which is substantial relative to the standards observed in studies on children's metaphor comprehension (Almohammadi et al., 2024 ; Pouscoulous & Tomasello, 2020).

Three hundred participants (52.7% girls), whose mean age was 8 years and 3 months (SD = 2 years, range [5 years 1 month; 11 years 12 months]) were recruited (See Figure 1 for the distribution depending on age). Participants were recruited from nine French public schools, situated in both rural and urban areas to ensure a high level of representativeness. All children were met with after their caregiver had given informed consent to their participation. They spoke fluent French and had not been diagnosed with any language impairment. The experiment began after the child also gave their informed consent.

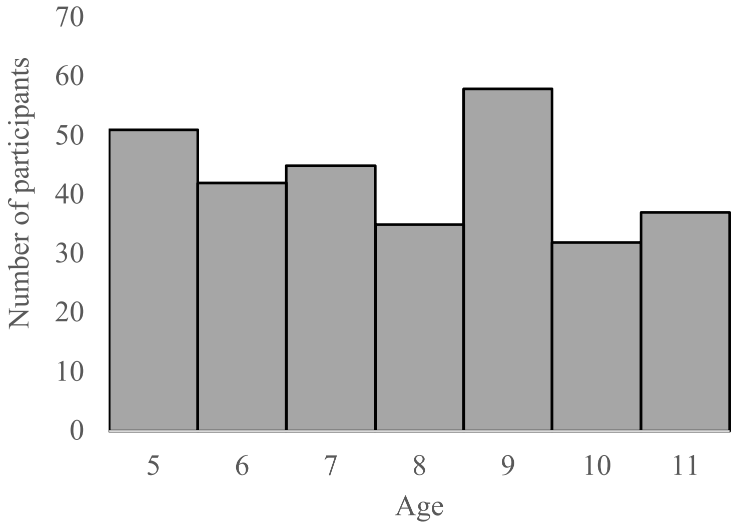


Figure 1. ***Distribution of Participants by Age***

Material

Metaphor understanding task

In the task used to assess metaphor understanding, children were presented with 10 metaphors, each preceded by a short context. Each metaphor was followed by a rephrasing to be judged on a 5-point scale for resembling quality, that is, representing how closely the rephrasing resembled the metaphor. A similarity judgment on a scale between the metaphor and the reformulation was used, rather than selecting among multiple rephrasings, due to the large number of rephrasings proposed.

Answers were compiled as a resembling score, from –2 (poor meaning resemblance) to 2 (very good meaning resemblance). Rephrasings were randomly selected from one of 4 types (See Table 1 for an example). The complete set of material (contexts, metaphors, and rephrasings) is available on OSF. Hereafter, we refer to an item as a metaphor within its context and a rephrasing of this metaphor.

Table 1. ***Level of understanding reflected by each type of rephrasing when judged as highly resembling the metaphor "Victor collects photos of butterflies in an album. For him, butterflies are rainbows".***

Good understanding	**Good-enough understanding**	**Poor understanding**			**Absence of understanding**
Metaphorical	Situational	Contextual	Literal	Vehicle	Incongruous
Victor loves butterflies because they are full of colours.	Victor thinks that butterflies are very beautiful insects.	Victor cuts out pictures of butterflies from magazines.	A butterfly is a colourful half-circle that we see in the sky after it rains.	After it rains, Victor looks for rainbows in the sky.	Victor's father bought a new car.

The different rephrasings were created to reflect different levels of understanding. The first type of rephrasings was labelled as metaphorical rephrasings. When judged as highly resembling the target metaphor, this type reflected a nuanced and thorough comprehension of the metaphor by precisely mirroring the characteristics of the vehicle in relation to the topic. The second type of rephrasings, situational rephrasings, demonstrated children's good-enough understanding of a metaphor when they received high ratings. Situational rephrasings were relevant in the communication context but did not precisely convey the link between the relevant semantic features of the vehicle and the topic. In the example presented in the Table 1, the situational rephrasing "Victor thinks that butterflies are very beautiful insects" convey the positive, esthetical aspect of butterfly without linking it to their colorful aspect. The third type of rephrasings, poor rephrasings, reflected a poor understanding of the metaphor when they were rated as highly resembling the metaphor. Given the various ways one can misunderstand a metaphor, we chose to present participants with different subtypes of poor rephrasings within this category. Specifically, for each metaphor, a participant could choose (1) the literal subtype which literally translated the metaphor, without regard for the context, resulting in a semantic incongruity; (2) the vehicle-oriented subtype which used the vehicle in its literal sense within a non-semantically incongruous sentence; (3) the contextual subtype, which was compatible with the context but did not convey the metaphor's figurative meaning. Finally, the fourth type of rephrasings, incongruous rephrasings, had no connection with the context of the metaphor, its figurative meaning, or the literal meaning of the words that made it up. When judged as highly resembling the target metaphor, this type of rephrasing would reflect a complete absence of understanding or consideration of the material, allowing for a methodological check to ensure that children correctly used the 5-point scale.

Metaphors, their embedding context and their rephrasings were pretested on multiple criteria, starting with an initial 28 items. First, we conducted two pretests on the metaphors themselves with adults to establish their comprehensibility and metaphoricity. We then conducted a pretest on the rephrasings with 295 adults (via an online questionnaire) to confirm that they did not all reformulate the metaphor equally well. Specifically, this pretest allowed us to retain only the items for which the metaphorical rephrasing was judged significantly better than the situational rephrasing, as intended in the construction of the material. A fourth pretest was conducted with 43 adults on the contexts accompanying the metaphors to ensure that situational rephrasings matched the context better than the contextual subtype of poor rephrasings, which in turn matched the context better than incongruous rephrasings. Finally, the 143 words from the initial pool of 28 items judged most difficult by the three authors of this article were included in two questionnaires designed to assess the knowledge of these words by 5- to 7-year-old children. These questionnaires consisted of 104 questions each and were administered to 67 young children ($M_{age} = 6.5$; $SD_{age} = 0.89$), allowing us to identify problematic words, i.e., those known by less than 70% of the children in the sample. When self-evident, a simpler word replaced a problematic one. More often, items containing problematic words were discarded. After these pretests, the initial number of items was reduced from 28 to 10 (refer to OSF documents for a detailed presentation of all pretests and results).

The final metaphor comprehension task was computerized with OpenSesame software (Mathôt et al., 2012). The interface allowed children to use a Likert scale represented by five coloured circles, blue to green, from left to right: non resembling, slightly resembling, moderately resembling, very resembling, highly resembling. Pictures of two similar cartoon characters were associated with the resembling side of the scale while two different characters were associated with the non-resembling side of the scale (see Figure 2). Once the whole item was heard, children could ask to replay either the context, metaphor or rephrasing to hear them again if needed. The children participated in a short training session on 6 items (5 literal and one metaphoric) before moving on to the test items to ensure their understanding of the task and answering scale. This training session was not the object of any analysis as its objective was primarily to allow the experimenter to explain the scale use to the participants. On each of the 6 items, the experimenter precise the meaning of each circle by rephrasing the choice of the participants (e.g. when the participant chose the “-2” circle: “so you mean that [target sentence] is not at all resembling to the [rephrasing]?”).

The selection of rephrasings was pseudo-randomised by OpenSesame, ensuring that, on average, participants encountered an equal number of acceptable rephrasings (i.e., metaphoric and situational rephrasings) and unacceptable ones (i.e., literal, contextual, and vehicle-oriented subtypes of poor rephrasings and incongruous rephrasings), approximately 5 of each. Specifically, the probability of selecting a metaphorical rephrasing was 2 out of 8, that of a situational rephrasing was also 2 out of 8, that of a poor rephrasing was 3 out of 8 (one for each subtype), and that of an incongruent rephrasing was 1 out of 8. OpenSesame also made it possible to randomly reverse the response scale.

The final task with 10 items was presented to 50 adults ($M_{age} = 36.10$; $SD_{age} = 11.74$; 29 women; 21 men) before being used with children to ensure that the task yielded consistent results in adults. A robust mixed model analysis showed that, as expected, participants judged metaphorical rephrasings as more resembling to the metaphor compared to situational ones ($b = 0.848$, CI_{95} [0.437, 1.258], $t(9.8) = 4.050$, $p = .002$, $\eta^2 = 0.626$), those two better than poor rephrasings ($b = 2.716$, $CI95$ [2.346, 3.086], $t(9) = 14.382$, $p < .001$, $\eta^2 = 0.958$) and finally incongruous as less resembling compared to all other rephrasings ($b = 2.114$, CI_{95} [1.833, 2.395], $t(28.78) = 14.76$, $p < .001$, $\eta^2 = 0.883$) (See data and complete analyses on OSF)

In summary, the computerized metaphor comprehension task consisted of 10 items, each comprising a metaphor, a brief context, and a rephrasing, pseudo-randomly chosen from 4 possible types. For each item, the participant’s task was to judge the extent to which the meaning of the rephrasing accurately conveyed the presented metaphor. Participants responded on a 5-point scale, allowing OpenSesame to collect a resembling score. Like the traditional forced-choice tasks of selecting the best metaphor rephrasing, the present task did not require participants to verbalise anything. However, unlike forced-choice tasks, participants were presented with only a single rephrasing for each metaphor, which (1) reduced the cognitive cost of the task by sparing them from having to maintain multiple competing rephrasings in working memory and (2) minimised the possibility of choosing a rephrasing by eliminating other options.

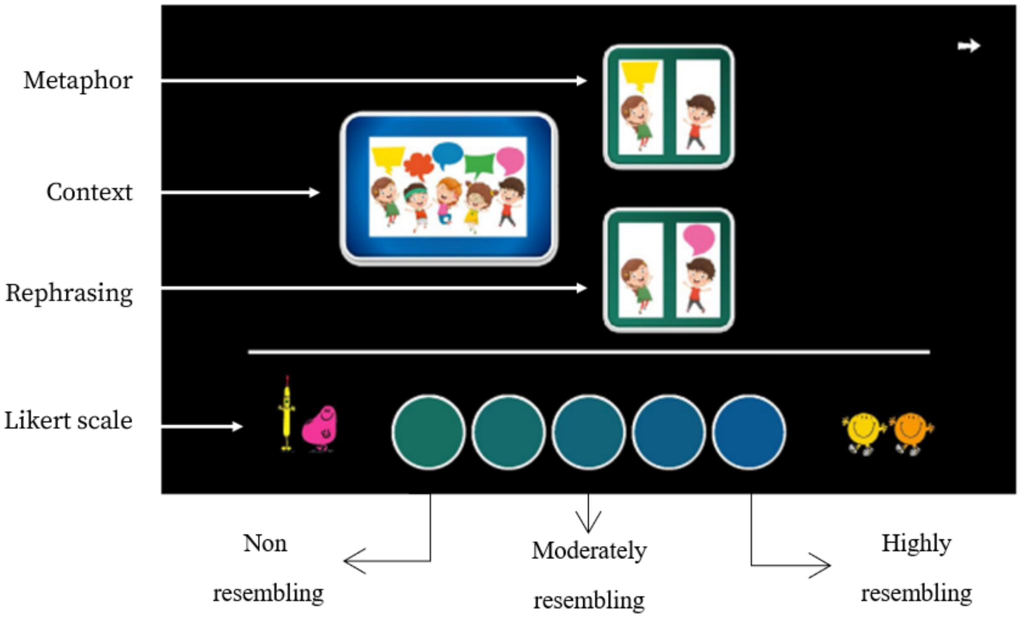


Figure 2. ***Opensesame Interface***

Peabody Picture Vocabulary Test (French Version)

The Peabody Picture Vocabulary Test (French version) (Dunn & Dunn, 1997) measured the receptive vocabulary level of French children by presenting four images and asking the participant to find the one representing a word. An age-adjusted standard score was calculated for each child, and this standard score was used for the statistical analyses (see the results section). Given the strong correlation between age and raw vocabulary score ($r = .79$), no young participant had a high raw vocabulary score, just as no older participant had a low raw vocabulary score, which prevented any satisfactory statistical modelling that included both age and raw vocabulary score.

Procedure

Children were met individually in a quiet room in their schools after their informed consent was obtained. They were first given the Peabody Picture Vocabulary Test. This test lasted approximately 10 minutes and was directly followed by the metaphor understanding task. This second task took around 15 minutes to be completed by participants.

Results

Overall, we collected 3000 observations (300 participants * 10 items), but 38 observations (1.27%) were discarded because we established that the participants did not hear either the metaphor or its rephrasing. All analyses and visualizations were conducted using the following R packages: lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), dplyr (Wickham et al., 2021), ggplot2 (Wickham et al., 2016), and robustlmm (Koller, 2016). All scripts and data can be found on OSF. We used a linear mixed model to analyse the resembling scores provided by participants depending on (1) their age, a continuous variable ranging from 5 to 11 years old, centred around the age 5; (2) the type of rephrasing presented, a four-level factor coded with a Helmert contrast: metaphorical rephrasings, situational rephrasings, poor rephrasings and incongruous rephrasings; and, most importantly, (3) the interaction between these two variables. We hypothesised that the resembling score would vary with age, depending on the type of rephrasing: it would increase with age for the metaphorical and situational rephrasing and decrease with age for poor rephrasings. Furthermore, we expected the resembling score for metaphorical rephrasings to show a greater increase with age compared to situational rephrasings. The participants' standard score on the Peabody Picture Vocabulary Test (a continuous variable centred around its mean) was included in the model as a covariate (4) to account for between-participant variability in vocabulary within each age bracket, as the standard score of vocabulary was adjusted for age. Additionally (5), since the interaction between the vocabulary score and the type of rephrasing was found to be significant, this interaction was also included in the model. This model was selected as the best fit for the data, as determined by the "step" function in the lmerTest package. The model's random structure, also determined with the "step" function, included by-schools, by-participants and by-items as random intercepts. The variance covariance structure was assumed to be unstructured. The degrees of freedom were calculated using the Satterthwaite method. The model assumptions were checked, and as the residuals deviated slightly from normal distribution, a robust analysis was conducted with the robustlmm package. Effect sizes were computed using the "tback" method described in Corell et al. (2022). Note that according to these authors, usual rule of thumb does not apply, η^2 being structurally smaller in mixed models.

Analysis of the Resembling Scores

As the variable age was centred around 5, this analysis indicates how children judged the different rephrasings at exactly 5 years old, independently of their vocabulary score (see Table 3 in Appendix A). Five-year-old children did manage to grasp the task, as evidenced by the fact that the resembling score they assigned to incongruous rephrasings was significantly lower than that assigned to other rephrasings (b = 1.69, *Wald* CI_{95} [1.33, 2.05], $t(2821) = 9.14$, $p < .001$, $\eta^2 = 0.029$); see Figure 3. However, no significant difference was found, either between the metaphorical and the situational rephrasings (b = -0.15, *Wald* CI_{95} [-0.49, 0.19], t(2830) = -0.87, p = .38), or between these two acceptable rephrasings taken together and the poor rephrasings (b = 0.08, *Wald* CI_{95} [-0.17, 0.34], t(2824) = 0.68, p = .50), which prevents us from concluding that 5-year-old children understand metaphors.

The pattern observed at age 5 evolved over time, as shown by significant partial interactions between age and the type of rephrasing; see Figure 3. First, children's ability to differentiate between poor rephrasings on the one hand and situational and metaphorical (i.e., acceptable rephrasings) on the other hand increased significantly and rapidly with age (b = 0.28, *Wald* CI_{95} [0.21, 0.34], t(2807) = 8.31, $p < .001$, = 0.024). The resembling score for poor rephrasings decreased while it increased for acceptable rephrasings, indicating an improving comprehension of metaphors over time. An examination of Figure 3 suggests that by age 6, children were able to distinguish between these two types of rephrasings. More crucially and as expected, over time, participants increasingly judged metaphorical rephrasings as resembling the metaphor better than situational rephrasings. Accordingly, the difference in resembling scores between metaphorical and situational rephrasings became significantly more pronounced with age (b = 0.12, *Wald* CI_{95} [0.04, 0.21], t(2824) = 2.76, $p < .001$, η^2 = 0.0027). An examination of Figure 3 suggests that the difference was well established around age 9. In summary, between the ages of 5 and 11, the children not only advanced in their understanding of metaphors by rejecting poor rephrasings, but they also appeared to evolve from good-enough understanding (with no clear difference between situational and metaphorical rephrasings) to fine understanding, where they rated metaphorical rephrasings higher than situational ones.

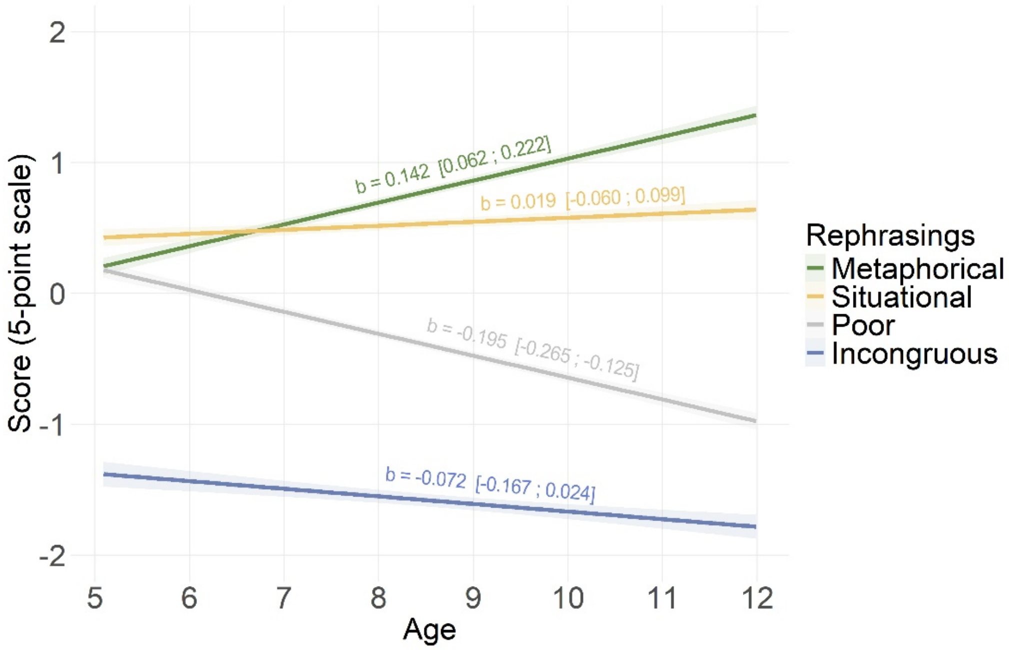


Figure 3. ***Resembling score for the four types of rephrasings as a function of age. Note: The translucent coloured areas around the lines represent the standard error. The b values reported inside the figure indicate the slope of each line [and its 95% confidence interval].***

In addition, a partial interaction between the standard score of vocabulary and the type of rephrasing was also observed. Specifically, as vocabulary level increased, the difference in resembling scores between metaphorical and situational rephrasings on the one hand and poor rephrasings on the other became more pronounced ($b = 0.28$, *Wald* CI_{95} [0.21, 0.34], $t(2807) = 8.31$, $p < .001$, $\eta^2 = 0.024$). In other words, at all brackets of age, children with the highest vocabulary proficiency were those who distinguished the best between acceptable and poor rephrasings. No other comparisons involving the standard score of vocabulary were found to be significant.

Exploratory Analyses on Poor Rephrasings

An original aspect of this study was to employ various subtypes of poor rephrasings of the metaphors: literal, contextual, and vehicle-oriented subtypes. We grouped these three subtypes of poor rephrasings together in the main analysis since our main hypotheses focused on the comparison between metaphorical, situational, and poor rephrasings. However, a post-hoc analysis of the resembling score differences between these different subtypes seemed valuable given the limited data in the literature comparing different forms of poor comprehension. This analysis, conducted using robust mixed-effects models with standard score of vocabulary as a covariate, included three pairwise comparisons between literal, contextual, and vehicle-oriented rephrasings at age 5 and their three interactions with age. The Holm method was used to adjust for multiple comparisons (see Table 2).

Examination of Figure 4 reveals that at age 5, the contextual subtype was easier to identify as a poor rephrasing than the literal and vehicle subtypes, which shared a literal interpretation of the metaphor's vehicle. Over development, the resemblance score assigned to all three subtypes decreased, more notably for the vehicle subtype than the other two. Consequently, it was the literal subtype that was the most challenging to identify as a poor rephrasing for older children. However, with the exception of the difference between the contextual and vehicle subtypes at age 5, Holm's correction applied to limit the risk of type I error prevents concluding significant statistical differences (see Table 2).

Table 2. ***Post-hoc comparisons between the three subtypes of poor rephrasings at age 5 and throughout development***

Comparisons	**Estimate**	***SE***	**Lower 95% CI**	**Upper 95% CI**	***t***	***df***	**Holm *p* (adjusted)**	**η^2**
Age	-0.180	0.043	-0.264	-0.096	-4.200	14.64	0.006*	0.547
Literal vs. Vehicle	-0.155	0.261	-0.667	0.357	-0.594	1085	1	0.000
Literal vs. Contextual	0.595	0.258	0.090	1.100	2.308	1090	0.085	0.005
Vehicle vs. Contextual	-0.750	0.259	-1.258	-0.241	-2.891	1079	0.023*	0.008
Literal vs. Vehicle *Age	0.158	0.066	0.030	0.286	2.412	1085	0.080	0.005
Literal vs. Contextual *Age	0.006	0.068	-0.127	0.140	0.090	1087	0.929	0.000
Vehicle vs. Contextual*Age	0.152	0.066	0.022	0.282	2.285	1073	0.068	0.005

Note: * indicates statistical significance

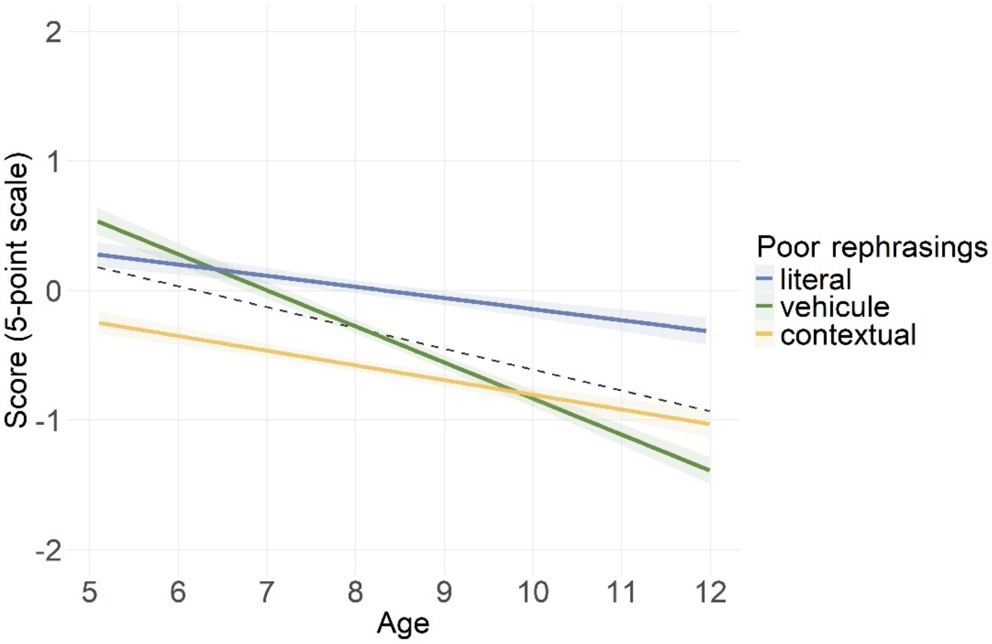


Figure 4. ***Resembling score for the three subtypes of poor rephrasings as a function of age. Note: the translucent coloured areas around the lines represent the standard error; the dashed ghost-line represents the resembling score for all three poor rephrasings, see also Figure 3.***

Discussion

The current study was designed to investigate the development of metaphor understanding in 5- to 11-year-old children within the theory of "good-enough" comprehension. An innovative metaphor rephrasing judgement task was used which was suitable for young children and did not involve verbal explanation or forced choices between several rephrasings of the metaphor. Our goals were to discover whether children demonstrated "good-enough" comprehension of metaphors, as opposed to "good" (precise and accurate) comprehension, and to test the hypothesis that good-enough comprehension precedes good understanding in development. We thus expected children to judge metaphorical rephrasings better than situational ones, and that their ability would become increasingly accurate with age.

Firstly, our data showed a vigorous and continuous development of metaphor comprehension between the ages of 5 and 11. This result was shown by a continuous increase in the resembling score assigned to the acceptable rephrasings, the metaphorical and situational rephrasings, and, simultaneously, a continuous decrease in the score assigned to the poor rephrasings. Our data thus support a substantial development in the understanding of metaphors throughout childhood which aligns with most of the previous literature (Di Paola et al, 2019; Deckert al., 2019; Ortony et al., 1978; Pouscoulous et al., 2011; Vosniadou, 1987; Willinger et al., 2019). The most original contribution of our work is to show not only that the understanding of metaphors evolves but also how it evolves. Our results show that children around the age of 7 understand metaphors, as they clearly distinguish between poor and acceptable rephrasings. However, they do not differentiate between metaphorical rephrasings, which formulate the link between the topic and the vehicle, and situational rephrasings, which do not. Their lack of distinction between situational and metaphorical rephrasings likely reflects the children's good-enough understanding at that stage that is vague yet satisfactory. Later in development, around the age of 9, children are able to distinguish between metaphorical and situational rephrasings, judging the former to be more accurate than the latter. This result suggests that their understanding has evolved to become more precise and accurate – good understanding – by taking into account the very precise link between the topic and the vehicle.

Karimi and Ferreira (2016) argue that two routes can lead to these types of understandings: 1) a heuristic route, applying straightforward guidelines that can generate a rapid, general representation of the information being processed, providing a benefit in efficiency in terms of cognitive exertion; 2) an algorithmic route, processing with precise and unambiguous syntactical procedures to compute accurate representations from all the provided linguistic input. Our data support the idea that children primarily access the meaning of metaphors through the heuristic route around the age of 7. When processing complex statements such as metaphors, it is likely that the algorithmic route is not sufficiently effective or is too costly at this age. As development progresses, the algorithmic route develops to become children's preferred route for understanding metaphors by around 9 years old in the context of experimental tasks like those used in this study. Consistent with Ferreira's good-enough model, good comprehension via the algorithmic route does not replace good-enough comprehension via the heuristic route at some point between 7 and 9 years old. Instead, children seem to expand their range of strategies to cope with complex statements to understand as metaphors. In a real-life situation with low stakes or limited integration time, 9- to 10-year-olds can settle for good-enough understanding, just like adults. However, in the tasks typically used to measure metaphor comprehension in children, good comprehension is favoured over good-enough comprehension: Indeed, elements such as the setting of the experiment in a laboratory or school, the presence of the experimenter, or the academic formatting of the task can make it difficult to measure good-enough comprehension in an experimental setting as it leads children to produce a good understanding. This should be a point of concern for future research: if children have different treatment modes, based on different route, resulting in different levels of understanding, does the task used bias responses towards one treatment mode rather than another?

In addition to the observation of vigorous and continuous development of metaphor comprehension between the ages of 5 and 11, the current study found that 5-year-old children struggled to distinguish between good and poor rephrasings, in particular those reflecting a literal interpretation of the metaphor's vehicle (literal and vehicle-oriented subtypes). We interpret the inability to distinguish between these rephrasings as a difficulty in understanding the meaning of the presented metaphors. This difficulty certainly cannot be attributed to a task unsuitable for children of this age, as they managed to judge incongruous rephrasings accurately. In addition, the vocabulary used in the task was carefully chosen to be understood by 5-year-old children. How do we explain the difficulties encountered by 5-year-old children in our study when previous research has shown that children can understand metaphors as early as at age 3 (e.g., Pouscoulous & Tomasello, 2020)? An explanation for this divergence in findings could lie in the methodology employed. The notion of metaphor corresponds to a large variety of utterance types that can also occur in a variety of situations. For instance, while Pouscoulous and Tomasello exclusively used simple metaphors (e.g., "The tower with the hat") referring to a presented object, the current study employed classic "an X is a Y" nominal metaphors embedded in a linguistic context and presented outside any situational context. Studying the development of metaphor comprehension in these different frameworks is valuable since it is the reality of what children experience every day.

Regarding the study of the evolution of metaphor comprehension, a contribution of the current study is related to the methodological choice to include measurements of poor understanding and not require the child to choose between different rephrasings of the metaphor. Asking children to evaluate each rephrasing on a scale provided a genuine method to assess the evolution of various potential interpretations of metaphors throughout the entire developmental period under consideration. In particular, our data suggest that literal rephrasings may pose difficulties even for older children. Although this exploratory observation needs to be confirmed by future research, such a tendency to retain the literal meaning has already been found in the comprehension of other forms of figurative language, such as idioms, metonyms or irony, and termed “literal bias” (Köder & Falkum, 2021). The two aspects presented, better assessing the impact of the intrinsic characteristics of metaphors on their comprehension throughout development, and better evaluating the developmental impact of literal bias on comprehension, appear to be important avenues for future research.

Although different from tasks that ask the child to explain the metaphors presented to them or tasks that ask them to choose among several options the one that best paraphrases the metaphor, our task of judging the similarity between the metaphor and a rephrasing on a 5-point scale necessarily involves metalinguistic skills. This is a clear limitation for many experiments that assess metaphor comprehension in children, because the development of metalinguistic skills after 5-year-old (Melogno et al., 2022) may explain part of the children's success or failure in the task. Measuring children's metalinguistic skills, with CELF-V subtests per example (Coret & McCrimmon, 2015) and using this measure as a covariate in analyses is a direction to explore for future research.

In addition, future research should be aware of the risk of bias in the use of Likert scales with young children. Although scales have been used in experimental tasks with children as young as five years old, they have been criticized for a potential tendency among younger children to favor the extreme ends of the scale—a tendency that decreases with age (Chambers & Craig, 1998; Chambers & Johnstone, 2002; von Baeyer et al., 1997). This bias toward the scale’s endpoints was observed in the present data. However, this bias alone cannot fully explain our results. To explore the impact of the bias, we performed a mixed logistic regression after recoding the original 5-point scale into a binary response system: scores of -2 and -1 were recoded as "no" (i.e., the metaphor and its rephrasing do not convey the same meaning), 9.8% of "0" scores were treated as N/A, and scores of +1 and +2 were recoded as "yes" (i.e., the two sentences convey the same meaning). This binary recoding of participants’ responses eliminated both the scale’s gradation and the potential bias. The regression yielded the same results as those presented in this article: we found an identical pattern of changes in resemblance scores for each type of rephrasing across age groups (see the "dichotomous answers" analysis and results on OSF). These additional analyses support the conclusion that, although a bias toward the scale’s endpoints was present in young children, it does not account for the primary findings of this study. Nonetheless, future research using Likert scales with young children should remain mindful of this potential bias and exercise caution when employing similar methodologies.

These findings allow for a reconsideration of the classical debate between early vs. later understanding. Metaphor understanding appears to be best described through a developmental lens, beginning with good-enough understanding that allows children to process metaphors they hear in everyday contexts in a manner that can be difficult to measure in demanding, experimental settings. Precise and accurate understanding was shown here to emerge later, between the ages of 7 and 9 years old, which aligns with previous literature showing a late understanding of metaphors (Billow, 1977; Deckert et al., 2019; Cometa & Eson, 1978). Metaphor comprehension may even undergo further refinement during adolescence (Carriedo et al., 2016; Nippold, 2006). According to our statistical model, we observed a gap of 0.71 between the resembling scores assigned to situational and metaphorical rephrasings at precisely 12-years-old. Data from the pilot study with adults (see Methodology) indicated a gap of 0.85. Although merely descriptive, this statistic suggests that development in understanding might continue into adolescence. Future work would benefit from extending the developmental period both into preschool as well as into adolescence.

In conclusion, this study conducted with an original task over a significant developmental period paves the way for a new conceptualisation of metaphor understanding.

We have shown that good (precise and accurate) understanding of metaphors is a slow process extending throughout childhood and possibly beyond, and that before being able to fully understand metaphors, children go through a phase where their understanding is “good enough”. New theoretical questions arise for the future: how does good-enough understanding emerge? In what way is good-enough understanding a lever for the development of good understanding? How and under what conditions do older children, adolescents, and adults transition from one type of understanding to another? There are also methodological questions, since it has become apparent that the tasks used to measure metaphor understanding inherently induce one level of understanding or another. Ultimately, considering and describing these different levels of comprehension can be used to support children in learning metaphors and perhaps other forms of language as well, both figurative and literal.

References

Almohammadi, A., Gaskins, D. K., & Rundblad, G. (2025). Metaphor comprehension in the acquisition of Arabic. *Journal of Child Language, 52*(2), 334–354. https://doi.org/10.1017/S0305000923000740

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Białecka-Pikul, M. (2010). Teaching children to understand metaphors as a path leading to theory of mind development. *European Journal of Developmental Psychology, 7*(5), 529-544. https://doi.org/10.1080/17405620802643692

Billow, R. M. (1977). Metaphor: A review of the psychological literature. *Psychological Bulletin, 84*(1), 81-92. https://doi.org/10.1037/0033-2909.84.1.81

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), e0150289. https://doi.org/10.1371/journal.pone.0150289

Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain, 78*(1), 27-37. https://doi.org/10.1016/S0304-3959(98)00112-2

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children’s use of rating scales. *Journal of Pediatric Psychology, 27*, 27–36. https://doi.org/10.1093/jpepsy/27.1.27

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

Cometa, M. S., & Eson, M. E. (1978). Logical operations and metaphor interpretation: A Piagetian model. *Child Development, 49*(3), 649-659. https://doi.org/10.2307/1128232

Coombes, L., Bristowe, K., Ellis-Smith, C., Aworinde, J., Fraser, L. K., Downing, J., … & Harding, R. (2021). Enhancing validity, reliability and participation in self-reported health outcome measurement for children and young people: a systematic review of recall period, response scale format, and administration modality. *Quality of Life Research, 30,* 1803-1832. https://doi.org/10.1007/s11136-021-02814-4

Coret, M. C., & McCrimmon, A. W. (2015). Test Review: Wiig, E. H., Semel, E., & Secord, W. A. (2013). Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5). https://doi.org/10.1177/0734282914557616

Correll, J., Mellinger, C., & Pedersen, E. J. (2022). Flexible approaches for estimating partial eta squared in mixed-effects models with crossed random factors. *Behavior Research Methods, 54*(4), 1626–1642. https://doi.org/10.3758/s13428-021-01687-2

Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language, 46*(2), 334-367. https://doi.org/10.1017/S0305000918000491

Declercq, C., Baltazart, V., & Didon, D. (2010). Les nuages sont un troupeau de moutons: À quel âge les enfants comprennent-ils qu'il n'y pas de moutons dans le ciel? *CanadianJournal of Experimental Psychology, 64*(2), 142–152. https://doi.org/10.1037/a0020175

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

Diamond, A. (2020). Executive functions. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), *Handbook of Clinical Neurology* (Vol. 173, pp. 225-240). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00020-4

Di Paola, S., Domaneschi, F., & Pouscoulous, N. (2019). 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

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (4th ed.)*. [Measurement instrument]. NCS Pearson.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*(5), 540-551. https://doi.org/10.1016/S0022-5371(81)90165-1

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

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

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

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*(1), 47-59. https://doi.org/10.2307/1130388

Gervain, J. (2020). Typical language development. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), *Handbook of Clinical Neurology* (Vol. 173, pp. 171-183). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00016-2

Huang, S. F., Oi, M., & Taguchi, A. (2015). Comprehension of figurative language in Taiwanese children with autism: The role of theory of mind and receptive vocabulary. *Clinical Linguistics & Phonetics, 29*(8-10), 764-775. https://doi.org/10.3109/02699206.2015.1027833

Johnson, J., & Pascual-Leone, J. (1989). Developmental levels of processing in metaphor interpretation. *Journal of Experimental Child Psychology, 48*(1), 1-31. https://doi.org/10.1016/0022-0965(89)90038-6

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology, 69*(5), 1013-1040. https://doi.org/10.1080/17470218.2015.1053951

Köder, F., & Falkum, I. L. (2021). Irony and perspective-taking in children: The roles of norm violations and tone of voice. *Frontiers in Psychology, 12,* 624604. https://doi.org/10.3389/fpsyg.2021.624604

Kogan, N., Connor, K., Gross, A., & Fava, D. (1980). Understanding visual metaphor: Developmental and individual differences. *Monographs of the Society for Research in Child Development, 45*(1), 1-78. https://doi.org/10.2307/1165832

Koller, M. (2016). robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models. *Journal of Statistical Software, 75*(5), 1-24. https

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

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

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

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

Nippold, M. A. (2006). Language development in school-age children, adolescents, and adults. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 368-373). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00852-X

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

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior, 17*(4), 465-477. https://doi.org/10.1016/S0022-5371(78)90283-9

Pierce, R. S., MacLaren, R., & Chiappe, D. L. (2010). The role of working memory in the metaphor interference effect. *Psychonomic Bulletin & Review, 17,* 400-404. https://doi.org/10.3758/PBR.17.3.400

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 51*(3), 460–477. https://psycnet.apa.org/doi/10.1037/xlm0001381

Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics, 25*(1), 51-79. https://doi.org/10.1075/bjl.25.04pou

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.pragma.2019.05.021

R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2022-01-01).

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

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

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 von Baeyer, C., Carlson, G., & Webb, L. (1997). Underprediction of pain in children undergoing ear piercing. *Behavioral Research and Therapy, 35*, 399–404. https://doi.org/10.1016/S0005-7967(96)00127-1

Vosniadou, S. (1987). Contextual and linguistic factors in children's comprehension of nonliteral language. *Metaphor and Symbolic Activity, 2*(1), 1-11. https://doi.org/10.1207/s15327868ms0201_1

Vosniadou, S., & Ortony, A. (1986). Testing metaphoric competence of the young child: Paraphrase vs. enactment. *Human Development, 29*, 226–230. https://www.jstor.org/stable/26764522

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., & Woo, K. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A grammar of data manipulation. R package version 1.0.7. https://doi.org/10.32614/CRAN.package.dplyr

Willinger, U., Deckert, M., Schmöger, M., Schaunig-Busch, I., Formann, A. K., & Auff, E. (2019). Developmental steps in metaphorical language abilities: The influence of age, gender, cognitive flexibility, information processing speed, and analogical reasoning. *Language and Speech, 62*(2), 207–228. https://doi.org/10.1177/0023830917746552

Winner, E., Engel, M., & Gardner, H. (1980). Misunderstanding metaphor: What's the problem? *Journal of Experimental Child Psychology, 30*(1), 22-32. https://doi.org/10.1016/0022-0965(80)90072-7

Data, Code and Materials Availability Statement

The material, data and analysis syntaxes that support the findings of this study are openly available on OSF at: https://osf.io/9pw7u/?view_only=None.

Ethics Statement

All the procedures contributing to this study complied with the ethical standards of the relevant national and institutional committees on human experimentation and with the 1975 Declaration of Helsinki, as revised in 2008. No information other than age, gender, first language, grade and oral consent were collected on behalf of the children. All children were met with after their caregiver had given informed consent to their participation. The experiment began after the child also gave their informed consent.

Authorship and Contributorship Statement

Sarah Ferrara, **Marc Aguera** and **Christelle Declercq** designed the study and created the material. **Sarah Ferrara** collected and analysed data and wrote the first draft of the manuscript. **Marc Aguera** collected and analysed data and revised the manuscript along with **Christelle Declercq**. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

The authors would like to express their gratitude to all participants and their families, as well as the students and schools involved in the data collection process. Additionally, we extend our thanks to colleagues who supported this work with their valuable advice, especially on analysis and statistical software.

License

Appendix A

Table 3: ***Fixed Effects Parameter Estimates***

	Estimate	***SE***	***t***	***p***	**η2**
(Intercept)	-0.063	0.141	-0.449	0.658	0.0099
Age	-0.026	0.031	-0.863	0.395	0.0251
Vocabulary	-0.003	0.003	-0.817	0.415	0.0025
Incongruous - (Poor, Situational, Metaphorical)	**1.688**	**0.185**	**9.140**	**0.000**	**0.0288**
Poor - (Situational, Metaphorical)	0.087	0.129	0.676	0.499	0.0002
Situational - (Metaphorical)	-0.150	0.172	-0.871	0.384	0.0003
Incongruous - (Poor, Situational, Metaphorical) * Age	0.061	0.045	1.354	0.176	0.0006
Poor - (Situational, Metaphorical) * Age	**0.276**	**0.033**	**8.314**	**0.000**	**0.0240**
Situational - Metaphorical * Age	**0.123**	**0.044**	**2.761**	**0.006**	**0.0027**
Incongruous - (Poor, Situational, Metaphorical) * Vocabulary	0.012	0.007	1.880	0.060	0.0013
Poor - (Situational, Metaphorical) * Vocabulary	**0.020**	**0.005**	**4.113**	**0.000**	**0.0060**
Situational - Metaphorical * Vocabulary	0.005	0.006	0.722	0.470	0.0002

Note: Significant results are presented in bold.

Making beds and dying of boredom literally: A developmental study on the comprehension of nonliteral uses of language in autism

Marta Ponciano
University of the Basque Country (UPV/EHU), Spain

Agustín Vicente
Ikerbasque, Basque Foundation for Science
University of the Basque Country (UPV/EHU), Spain

José V. Hernández-Conde
University of Valladolid, Spain

Elena Castroviejo
University of the Basque Country (UPV/EHU), Spain

Abstract: Comprehension of simple nonliteral uses of language was investigated in three- to nine-year-old autistic and linguistically matched typically developing (TD) children, by assessing their understanding of nonliteral uses of language with potential literal senses. Children were tested on conventional metaphors, idioms, hyperboles, and light verb constructions. The aim of the study was to determine whether autistic children showed a genuinely strong tendency to interpret nonliteral uses of language literally across development. A total of 166 children (N = 42 Autistic children; N= 124 TD children) were tested using a paradigm with online (response times) and offline (picture selection) measures. Overall, there were no significant group differences on the picture selection task, but autistic children were slower in spite of increasing verbal age. Both groups showed continuous improvement of their understanding of literal and nonliteral senses with increasing verbal mental age. The results, nevertheless, call for a reflection on the (possible) literalist behavior in autism, indicating that it is important to take into account individual variation, as we observed different kinds of performance within the autistic group.

Keywords: nonliteral uses of language; autism; literalism; development; heterogeneity

Corresponding author(s): Marta Ponciano, Department of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU), Centro de Investigación Micaela Portilla Ikergunea (UPV/EHU), Justo Vélez de Elorriaga, 1, 01006 Vitoria-Gasteiz, Spain. Email: marta.ponciano@ehu.eus.

ORCID ID(s): Marta Ponciano: https://orcid.org/0000-0002-7062-0518; Agustín Vicente: https://orcid.org/0000-0002-5331-5546; José V. Hernández-Conde: https://orcid.org/0000-0002-8502-6570; Elena Castroviejo: https://orcid.org/0000-0003-0371-1214

Citation: Ponciano, M., Vicente, A., Hernández-Conde, J.V. & Castroviejo, E. (2025). Making beds and dying of boredom literally: A developmental study on the comprehension of nonliteral uses of language in autism. *Language Development Research, 5*(2), 93–129. https://doi.org/10.34842/ldr2025-826

Introduction

Nonliteral Language in Autism

Difficulties understanding nonliteral uses of language by interpreting them literally are common across the autistic spectrum in everyday situations (Morra, 2016). Literalism has been taken to be characteristic of autism since the pioneering investigations of Kanner (1943), who observed that autistic individuals exhibited an unusual tendency to interpret non-literal language literally, and Asperger (1991), who noted that the children he studied exhibited characteristic difficulties in understanding jokes (Geurts et al., 2019).

However, previous studies in laboratory settings have typically yielded contradictory results (Gernsbacher & Pripas-Kapit, 2012; Kalandadze et al., 2018; Lampri et al., 2023; Vicente et al., 2024). Some studies report systematic difficulties, often showing tendencies to interpret nonliteral language in a literal manner, while others indicate no significant differences from typically developing controls. The discrepancies between studies' findings may depend on the type of nonliteral use that is being tested or the methodology used (Kalaldadze et al., 2018). Concerning types of nonliteral uses tested, many studies on autistic children, adolescents or young adults report special difficulties understanding irony and sarcasm (Deliens et al., 2018; Happé, 1993; MacKay & Shaw, 2004; Saban-Bezalel et al., 2019), while several studies targeting scalar implicatures or indirect speech acts do not find similar differences (on scalar implicatures: Chevallier et al., 2010; Su & Su, 2015; Hochstein et al., 2018; Pijnacker et al., 2009; Van Tiel & Kissine, 2018; and Schaeken et al., 2018, and Mazzaggio et al., 2021 for discrepant results; on indirect speech acts: Kissine et al., 2015, Marocchini et al., 2022; and Paul & Cohen, 1985, and Ozonoff & Miller, 1996 for opposite results).

Focusing on figurative language, results on autistic individuals' comprehension of metaphors and idioms are mixed (Kalandadze et al., 2018, 2019; Lampri et al., 2024). Since Norbury's (2005) seminal work, many researchers hold that differences in metaphor comprehension are related to differences in general linguistic abilities, such that if autistic and non-autistic participants are matched on linguistic abilities, no significant differences should emerge in metaphor comprehension (Andrés-Roqueta & Katsos, 2017). However, this "structural language hypothesis" does not account for all the variability observed: some studies that match autistic and non-autistic individuals on linguistic abilities have found that autistic people exhibit more difficulties than neurotypicals with metaphor interpretation. In particular, several of these studies find that autistic participants exhibit a stronger tendency than neurotypicals to understand metaphors, as well as idiomatic expressions, literally (Chahboun et al., 2016, 2017; Vulchanova et al., 2012; Walenski & Love, 2017). Even if autistic participants in these studies end up selecting the nonliteral interpretation in a forced-choice task, online measures such as eye-tracking reveal a more significant interference of literal options in the autistic population than in neurotypicals (O Shea et al., 2024, Martin-González et al., 2025).

There is not much work on the development of figurative language comprehension (or on nonliteral language comprehension in general) in autism research. Rundblad & Annaz (2010) tested *conventional* metaphor comprehension in autistic children (n=11), with ages ranging from 4 to 11, using a picture selection task. Results showed difficulties in all ages, as well as a lack of relation between verbal mental age (VMA) and accuracy, which in principle is surprising since other studies with older children (10-12 yr.olds) have found such a relation (Kasirer & Mashal, 2014, 2016; see also Olofson et al., 2014, and Pastor-Cerezuela et al., 2020, who, in contrast with Kasirer & Mashal, 2016, found novel metaphors to be more difficult than conventional metaphors for autistic children). In a cross-sectional similar study on *novel* metaphors with autistic children and teens, Van Herwegen & Rundblad (2018) found more difficulties in the autistic group than in a control group matched on chronological age. Van Herwegen & Rundblad selected some of the youngest participants of their initial sample (children of around 9 years) for a longitudinal study, which showed improvements on metaphor comprehension related to age. In a recent study, also on novel metaphors, Martin-González et al. (2025) tested two groups of children, autistic and non-autistic, matched on VMA (chronological ages ranging from 3 to 12), on a picture selection task. The gaze patterns of the children were recorded. The results did not find differences between the autistic and the non-autistic groups in the picture selection task, but there were significantly more looks at literal competitors in the autistic group in the metaphorical condition than in the TD group.

In sum, several decades of research on nonliteral uses of language in autism has thrown mixed results, with some evidence suggesting that autistic individuals may interpret nonliteral language literally more often than neurotypicals. In this paper we add to the developmental research on literalism in autism: we explored whether autistic children have easier access than their TD peers to the literal interpretations of ambiguous expressions, and whether TD children have easier access than their autistic peers to the nonliteral interpretations of such expressions.

Our main research question is about a particular kind of literalism. There are several different notions of literal meaning at the level of utterances (Allott & Textor, 2022). Depending on the notion at stake, literalism can be understood in one way or another. For the purposes of this paper, we propose to understand a literal interpretation of any given utterance as a thoroughly compositional interpretation, i.e., one generated on the basis of fixed rules of grammar that take as input stable and concrete meaning assignments. Correspondingly, by nonliteral uses of language, or nonliteral language, we mean any use of language that is intended not to be processed compositionally. According to this notion of *the literal*, the metaphor *Juliet is the Sun* counts as a nonliteral use of language, but also do hyperboles such as *to die of boredom,* conventional metaphors such as *to be a pig* (meaning: to be greedy and selfish), and even light verb constructions such as *to take the bus* or *to make the bed*. Our main research question, thus, is whether autistic individuals at some point in development exhibit this kind of literalism, i.e., whether they exhibit a more significant tendency than neurotypicals to understand utterances in a word-by-word, compositional manner.

As we are aware that this notion of the literal may be controversial, the following paragraphs argue why we consider it a legitimate notion. We take it that the most controversial point concerns light verb constructions (LVCs). We follow Fleischer (1997), Fellbaum et al. (2006) and Wittenberg (2016) in considering that light verbs are "light" because they have lost most of their semantic weight, such a loss representing a shift away from the literal. According to this view, when a light verb appears in a LVC, the verb is no longer being used in its original, literal sense, and instead, takes on a quasi-auxiliary role within the construction. However, LVCs do also retain a literal interpretation in the sense here characterized (Wittenberg, 2016). Such literal interpretation is given by applying rules of grammar to meanings that represent specific rather than unspecific notions (Taylor, 2006). Light verb constructions are not dissimilar from other idiomatic expressions, where the meaning of the whole cannot be derived from the literal meanings of its parts. For this reason, we think that LVCs can be considered part of the spectrum of nonliteral language (Fleischer, 1997, Fellbaum et al., 2006). Yet, we do not need to endorse such a view; what we are after is the contrast between a literal interpretation of a LVC, as we have defined it, and an interpretation that is not literal in that same sense.

The most habitual interpretation of a LVC is its nonliteral interpretation. *To make the bed* is almost never understood as "to create or generate a bed". Actually, the nonliteral meaning of the phrase is arguably its *conventional* meaning. On most occasions, if someone is told to make a bed, they are expected to lay down sheets, etc.; otherwise, they would be taken as not complying with the directive. This suggests that what we call the *nonliteral meaning* of a LVC may actually be taken to be its literal meaning. While this is correct, the issue is terminological: we use a notion of literal meaning such that the conventional and the literal do not have to coincide. Actually, we are interested in finding out whether autistic children at some point of development may go for a compositional meaning of a phrase instead of going for its conventional and habitual meaning.

The metaphors, hyperboles and idioms that we test in our study, together with items displaying LVCs, are instances of nonliteral conventional language. However, to say that they are instances of *conventional* nonliteral language does not mean that the nonliteral meaning is, or may be, the conventional meaning. In the case of metaphors and hyperboles, *conventional* is opposed to *novel*. That is, to say that a metaphor is conventional is, roughly, to say that it has been lexicalized, and has become another meaning of the word (see Bowdle & Gentner, 2005). However, this does not entail that such a meaning has become conventional in the sense of being dominant.

These expressions are also part of utterances whose more frequent meaning is their nonliteral meaning. For example, *to die of boredom* rarely, if ever, means to be killed by boredom. In cases like this hyperbolic use of *die*, however, it is arguably less controversial to hold that the most frequent meaning is not the literal meaning of the expression. This is because at the word level speakers would have it clear that *die* has the literal meaning it has. That is, even though at the phrase level the nonliteral meaning of *to die of boredom* is clearly more frequent than its compositional meaning, the verb *die* is still dominantly used in its literal sense. However, despite there being such a difference between e.g., *to take the bus* and *to die of boredom* with respect to verb meanings, we take it that it is a difference in degree: *take* is used in its nonliteral, modified, sense more often than *die*, such that *take*'s literal, concrete, sense is more difficult to access than the literal meaning of *die* in nonliteral constructions.

In sum, the purpose of the study was to explore whether autistic children exhibit more facility than TDs in interpreting expressions whose more frequent meaning is their nonliteral meaning in a literal, compositional way. We also wanted to investigate how difficult it is for autistic children to acquire nonliteral meanings, even when such meanings are very familiar and salient. Understanding these differences is important, as it can provide insights into the cognitive and linguistic processing styles of autistic individuals. The study thus seeks to contribute to broader discussions in autism research and pragmatic theory, ultimately fostering more effective communication strategies and educational approaches tailored to the needs of autistic children.

Current Study

In this study, we tested autistic children ranging in age from four to nine, with verbal mental ages (VMA) ranging from three to nine, on their understanding of familiar nonliteral expressions in their L1, Spanish. Autistic children and TD children matched on VMA were compared. The (main) interest of the study was to compare the two groups' development of nonliteral language comprehension based on linguistic development, focusing particularly on response times. This focus allowed us to evaluate how easily each group processed literal versus nonliteral interpretations of familiar expressions.

In Castroviejo et al. (2024), we conducted the typical developmental study using offline (picture selection) and online (response time) data collection from the picture selection task used in this paper. Unlike the present study, which treats VMA as a continuous variable, the aforementioned study grouped participants into discrete age categories. We observed a linear progression across ages in *nonliteral* readings. The progression started with around 68% of correct responses at age three and reached *ceiling* performance in the oldest group (age nine). The development of *literal* interpretations was less straightforward: three-year-olds exhibited an accuracy of around 75%, and performance remained at that level until they were five. From five to eight they became more accurate. Concerning reaction times (RT), it was observed that access to literal interpretations was easier than access to nonliteral interpretations for three- and four-year-olds, that both kinds of interpretations began to be equally accessible at around five years, and that the pattern may reverse again at nine.

The results indicated distinct patterns in children's performance between two groups of experimental items: those categorized as Light Verb Constructions (LVC) and those classified under (what we called) the Metaphor-Hyperbole-and-Idiom continuum (MHI). For example, LVC items involved expressions like *hacer la cama* 'to make the bed', where the verb itself contributes with little semantic content. In these LVC items, children were more accurate in the nonliteral than the literal condition. In contrast, items in the MHI category, which included metaphorical expressions like *ser una tortuga* (*lit.* 'to be a turtle'; meaning: 'to be slow'), idioms like *partirse de risa* (*lit.* 'to split oneself with laughter'; meaning: 'to burst out laughing'), and hyperboles like *morirse de aburrimiento* ('to die of boredom') showed that younger children (ages three to five) were more accurate in the literal condition than in the nonliteral one.

The task was a forced-choice task including no competitors (i.e., in the literal condition, a representation of the nonliteral meaning; and in the nonliteral condition, a representation of the literal meaning). Children had to decide between a representation of either the literal or the nonliteral meaning, and two distractors. This paradigm was chosen primarily because the study aimed to compare the accuracy and ease of access to literal meanings *versus* the accuracy and ease of access to nonliteral meanings in both groups, rather than to examine whether children preferred one interpretation over the other. We also thought that the inclusion of competitors might confound the results, particularly when the research question focuses on how accurately and rapidly children access one meaning versus the other. Children may be able to understand a nonliteral expression and do so quickly, but this might not be the case if a literal competitor is explicitly presented. For instance, Köder and Falkum (2020) reported that four- and five-year old TD children exhibited a stronger tendency than younger and older children to choose a representation of the literal meaning of a metonymy, even though their gaze patterns were similar to those of other ages. What Köder and Falkum observed may mean that four-and five-year-old TDs understand metonymies, but that, when prompted to choose between competing representations in a picture selection task, they chose the literal one. Finally, we also considered that including competitors could affect the autistic group more than the non-autistic group, since autistic children may experience more uncertainty in general than TD children (Vicente et al., 2024).

Research Questions and Hypotheses

As said, the main goal of this study is to test the literalism trend in autistic and non-autistic children matched on verbal mental age across development. By assessing performance on familiar expressions with a more frequent or dominant nonliteral use, we can observe (a) whether autistic children experience more difficulties accessing nonliteral meanings across development, and (b) whether autistic children have a comparably easier access to literal meanings, even when such literal meanings are unusual. Our research questions are spelled out below.

RQ1: Are there any differences in the performance of TD and autistic children when interpreting nonliteral expressions across development (from 3 to 9 years of verbal mental age)?

RQ2: Do autistic children exhibit an easier access to literal, compositional, meanings at different developmental stages?

We hypothesized that autistic children would have more difficulties understanding nonliteral uses of language across the board than TD peers. At the same time, in keeping with the idea that autistic individuals are more literalist than non-autistic individuals (Vicente & Falkum, 2023), we expected autistic children to experience less difficulties than TD children understanding the literal but uncommon use of the expressions in our sample. We thus hypothesized that autistic children would perform better than TDs in the literal condition. Concerning latencies, we hypothesized that the said differences between autistic and TD children should have an effect on response times (RT): in general, we thought we could observe autistic children accessing literal meanings faster than nonliteral meanings. However, we were more confident about observing the reverse pattern in the TD group: TD children should access nonliteral interpretations faster than literal, word-by-word interpretations. We considered the prediction about autistic children's faster access to literal interpretations to be just one possibility, given the nature of the items with highly frequent nonliteral meanings and the age range of the participants. In any event, we did expect to find this kind of difference in access in the youngest autistic children.

We did not have specific predictions regarding the classification of items (LVC category *vs* MHI category) for autistic participants. The investigation of this dimension was more exploratory in nature. We did not predict anything in development either regarding those two item categories. Previous literature has focused extensively on metaphors, metonymies, and irony, but not so much on conventionalized expressions such as LVCs. For this reason and the observed bimodal distribution in typically developing children's performance, we included those two categories into the analyses.

According to the literalist hypothesis, in atypical development we expected to observe a more prolonged literalist stage, where children would perform better in the literal than in the nonliteral condition. With increasing verbal mental age, autistic children should improve on the nonliteral interpretation, since there should be an effect of frequency corresponding to an improvement on linguistic skills generally. We also expected them to be more sensitive to frequency effects as their VMA increased, resulting in more difficulties in accessing literal interpretations.

Method

Participants

We tested a total of 166 Spanish speaking children aged three to nine–years. Children were divided into two groups based on whether they were typically developing (TD) (N=124) or had a diagnosis of autism (N=42) (see Table 1 for the participants' descriptive data). The control group was recruited from participation in the previous study (Castroviejo et al., 2024).

For this experiment, 42 Spanish-speaking children (9 girls and 33 boys) diagnosed with autism were recruited (a) through the Early Intervention program (Atención Temprana, Álava), Araba, Spain; (b) from the APNABI Autism association (Asociación de familias de personas con un Trastorno del Espectro del Autismo), Bizkaia, Spain; (c) from the Ilargia Intervention Center, in Logroño, Spain, and (d) from mainstream public schools, Bizkaia, Spain. Most of the children had received a formal diagnosis of autism in conformity to the criteria of DSM-5 (and/or ICD11) by a team of professional neuropsychologists, pediatricians and speech therapists specialized in ASD diagnosis. For those who had not received a diagnosis prior to the study, diagnosis of autism was confirmed by a licensed clinician on the basis of DSM-criteria, using The Autism Diagnostic Observation Schedule (ADOS-2) (Lord et al., 2012) and sometimes additionally Autism Diagnostic Interview – Revised (ADI-R) (Rutter et al. 2003). All autistic participants were attending mainstream classrooms following inclusive education criteria, and the majority of them were receiving intervention (either group intervention, individual intervention, or both, via public or private services).

We did not conduct a formal power analysis before starting the study, so the sample size was not determined based on statistical criteria. Instead, we recruited as many eligible children as we could within the age range of interest, guided by previous literature and the design of the broader study. In the end, practical factors like time constraints and participant availability largely shaped the final sample size.

The chronological age (CA) of autistic children ranged from four to nine years (M_{age}= 80.69 months, *SD*= 18.29 months). Autistic children's nonverbal intellectual abilities, as measured by the Leiter-3 scale (Roid, Miller et al., 2013), ranged from 73 to 125 (*M*=98.88; *SD*=11.87), with average or above-average non-verbal intelligence. Two children's cognitive measures were absent, and mean imputation was used to replace those missing values. There was no evidence to believe that they were below the threshold on non-verbal abilities. Also, there was no mention of any intellectual disability in their diagnoses.

A total of 143 Spanish-speaking TD children were initially recruited from various mainstream public schools in Bizkaia, North of Spain; 19 children were eventually excluded from analyses (due to one of the following reasons: i.a. being in the process of receiving a diagnosis of Generalized/Pervasive developmental disorder (PDD-NOS)/Autism Spectrum Disorder (ASD), poor performance (below chance levels) in the filler items or being absent from the testing sessions). The final group consisted of 124 Spanish-speaking children (55 boys and 69 girls) between the ages of 2;11 and 9;11 (year, month) (M_{age}=72.05 months; *SD*=22.88 months, among the TD children whose results were included within the analyses, one was slightly below the age of three, 2;11;29 (years; months; days)).

Groups were matched on receptive vocabulary as measured by the Peabody Picture Vocabulary Scale/Test (PPVT-3, Dunn et al., 2006; ASD: $M_{age\ equivalent}$ = 67.45 months, *SD* = 20.65 (34-119); TD: $M_{age\ equivalent}$ = 72.05 months, *SD* =22.88 (35-119)), t(77) = -1.21 , p = .22, although TD children were not tested with the PPVT. Their receptive vocabulary measures were assumed to be in line with their chronological age (CA), meaning their age equivalents were considered to be similar or equal to their chronological age. The assumption was necessary because the TD group was originally part of a developmental study where VMA was not a variable of interest. Autistic children were significantly older (*t*(87)= 2.47, *p* = .01) than TD children, and had a significant difference between their CA and their VMA (*t*(41) =4.71, *p* = 2.753 e-05).

Table 1. *Participants' descriptive data.*

	ASD		**TD**	
N (*M*:*F*)	42 (33:9)		124 (55:69)	
Variable	Mean (*SD*)	Range	Mean (SD)	Range
Age in months	80.69 (18.29)	49-118	72.05 (22.88)	35-119
VMA in months	67.45 (20.65)	34-119	72.05 (22.88)	35-119
Non-verbal IQ	98.88 (11.87)	73-125	-	-

Notes: Matching measures are in boldface. VMA (Verbal Mental Age) age equivalence (standardized score) is the verbal mental age measure obtained from PPVT-III. It is an age-scaled score from the direct score. Non-verbal IQ test (standard scores) from LEITER-3 test, composite scores have a mean of 100 and an SD of 15. In the case of TD children CA and VMA are assumed to be the same. TD, typically developing; ASD, autism spectrum disorders.

Materials

Stimuli consisted of Spanish expressions whose nonliteral interpretation is highly frequent, but whose literal reading is plausible though awkward. Frequency was tested by means of a norming study (Castroviejo et al., 2024). Participants were tested on the interpretation of expressions including hyperboles (*morir de aburrimiento* 'to die of boredom'), metaphors (*ser una tortuga*; *lit*. 'to be a turtle'; *nonlit*. 'to be slow'), idioms (*partirse de risa*; *lit*. 'to split oneself with laughter'; *nonlit*. 'to burst out laughing'), and light verb constructions (*hacer la cama* 'to make the bed'), as in (1).

(1) a. Sergio se muere de aburrimiento.
Sergio SE dies of boredom
'Sergio is dying of boredom.'

b. Unax es una tortuga.
Unax is a turtle
'Unax is a turtle.'

c. Tania se parte de risa.
Tania SE breaks of laughter
'Tania bursts out laughing.'

d. Juan hace la cama.
Juan makes the bed
'Juan makes the bed.'

All stimuli were a combination of a visual array of three pictures and an orally presented sentence. Only one picture reflected the target meaning of the sentence, and the other two were distractors. We created 16 critical sentences that contained the aforementioned linguistic expressions (see Appendix, Table A1.). There were also filler items. Each sentence had a literal and nonliteral version/ visual representation, in which the same linguistic expression is used to refer to the literal or nonliteral interpretation (see Figure 1). Participants saw 8 nonliteral and 8 literal versions of the sentences. The purpose of this design was to assess participants' ability to access each meaning (frequent non-compositional or infrequent but compositional) independently. Each participant saw a total of 32 experimental trials, 16 critical sentences (8 nonliteral and 8 literal) and 16 filler items.

Figure 1. ***Example of the visual display that participants saw on a computer screen showing the literal and nonliteral conditions of the critical item quedarse helado ('to be shocked, lit. to get frozen'), respectively. Participants saw one of the two visual grids.***

Filler trials consisted of simple, common sentences in which the sentence referred to only one picture, such as *María anda en bici* ('Maria is riding her bike') or *Tomás está dormido* ('Tomás is asleep') or *Uxue come un helado* ('Uxue is eating an icecream'), similar in structure to experimental items. They were meant to be understood compositionally (i.e., literally). Filler sentences were included to keep children's attention and ensure that they were actively participating in the task. There were four pre-test trials very similar to filler sentences (e.g. *Andrés bebe agua* 'Andrés is drinking water'; *Hace sol* 'It is sunny').

Going back to the stimuli, the alternative images for two experimental items (the correct response in each condition is underlined) are illustrated in (2) and (3).

(2) Juan hace la cama. 'Juan makes the bed.'	(3) Pedro está hecho polvo Pedro is done dust 'Pedro is exhausted.'
a. Juan is lying on his bed.	a. Pedro is watching TV.
b. Juan is looking at his bed.	b. Pedro is smiling.
c. Juan is building a bed (literal).	c. Pedro has turned into dust (literal).
d. Juan is arranging the linen (nonliteral).	d. Pedro is exhausted (nonliteral).

Two counterbalanced lists were created, to make sure that participants did not see the same trial twice – once in the literal, once in the non-literal interpretation – and we assigned them randomly to one of two lists. Thus, participants did not see both target meanings in the same visual array. For each participant, order of presentation of the stimuli was randomized. E-prime 3.0 stimulus presentation software (Psychology Software Tools) was used to build and run the experiment.

The total list of critical items was split into two categories, following the distribution of performances found in the analyses of TD children (Castroviejo et al., 2024): light verb constructions on the one hand (i.e., *hacer la cama* 'to make the bed') and a continuum of metaphors, hyperboles and idioms on the other (i.e., *partirse de risa* 'to burst into laughter') (remember, LVC and MHI, respectively). The continuum included several phenomena since it seemed to be difficult to disentangle hyperboles from metaphors (see Wilson & Carston, 2007) and idioms based on metaphors from metaphors (Vega-Moreno, 2005). Each category contained an unequal number of items (five in the LVC category and nine in the MHI category: three hyperboles, three metaphors and three idioms). Accordingly, the effect of item category was taken into account in the following analyses.

Comprehension of these expressions undoubtedly depends on frequency factors and degree of exposure to these linguistic inputs. Since there is no easy access to corpora made up from transcripts of children's linguistic input and child-ambient speech in Spanish, specifically what the child incidentally learns from adult speech, we could not include those estimates for the critical items. We opted instead for a subjective measure. We ran a small norming study by means of an online questionnaire with adult native speakers of Peninsular Spanish (*N*= 63, ages ranging from 18 to 66 years old) to confirm that the stimuli were highly frequent in their nonliteral, dominant sense. As explained in Castroviejo et al., (2024), the results of the norming study confirmed that the expressions were highly frequent in their nonliteral sense (see a detailed presentation and discussion of the study in Castroviejo et al., 2024). On the other hand, the results reported in Castroviejo et al. show that the expressions used were familiar enough to TD children: even the youngest 3-year-olds were able to select the picture representing the nonliteral interpretation with above 60% of accuracy).

Procedure

The study was designed as a sentence-picture matching task. Participants were asked to help a friend, a cartoon character, to learn some things from the target language. First participants saw the cartoon character appearing on the computer screen together with our target sentences. The stimuli sentences were presented orthographically and auditorily twice (via a pre-recorded female voice). The reason for hearing the stimuli twice was motivated by the need to ensure that the youngest children had enough opportunities to engage with the task. We also wanted to make sure that they did not have trouble with their inattention and/or impulsivity. After listening to the stimulus, participants saw a visual display with three pictures (see

Figure 1). The position for presenting the images was counterbalanced between participants and items and order of the expressions. The participants' task was to select the visual target which matched the sentence including the expression they had previously heard, by clicking on the screen. The computer-mediated interface was especially appealing for children.

TD children were tested individually at their school in a quiet room away from their class. Autistic children either came to the laboratory to participate in the experiment, accompanied by their caregivers; or those who belonged to APNABI Autism Association/Ilargia Intervention center were individually tested at their center during one of the intervention sessions. PPVT-III and LEITER-3 tests were administered either at the laboratory or at their respective centers in a different videotaped session within a few weeks of the first session, to avoid participants becoming tired, distracted or restless. Breaks were given as often as were required. This second testing session lasted about one hour. Before testing took place, children verbally assented to participating in the experiment.

Each child was shown first four pretest trials and, if they could answer those appropriately, they were then given the complete experimental package which consisted of 32 items. The task took participants around 10 minutes to complete. Measures of accuracy and response times (RT) were collected to determine ease of processing. Reaction times (RTs, in ms) were calculated from the time the visual display appeared, just after the sentence was played for the second time, until the moment the participant tapped for a quick selection on the touch-screen display.

Analysis

Prior to the analysis, we intended to remove participants who scored less than 50% accuracy on filler items, but none was below this threshold. All statistical analyses were conducted in R (R Core Team, 2022), using the lme4 and lmerTest packages (Bates et al. 2007; Kuznetsova et al., 2017). We fitted generalized linear mixed models on accuracy (correct versus incorrect responses / rates of successful target responses) and linear mixed-effects models on response times, including the four factors, verbal mental age (VMA) as a centered continuous variable, group (ASD, TD), condition (literal, nonliteral) and category of item (LVC, MHI) together with their interactions. We coded all the variables using sum-contrast coding. The random effects structure included random intercepts by items and by participants. We intended to include a maximal random effects structure (Barr et al., 2013) but simplified it due to convergence issues. Non-convergence issues were handled by using the 'bobyqa' optimizer as well. We report relevant t-values and p-values for reaction time and z-values and p-values for accuracy.

We took two different approaches. We ran two analyses. The first one, the correct response analysis, provided us with a general overview of the abilities to identify the correct target image or response (either literal or nonliteral interpretation) compared to the non-target images. Second, the response time analysis explored the ease with which participants chose the target response and provided us with an overview of the different experimental manipulations of the study. Finally, an additional analysis involved a descriptive examination of autistic children's performance given some variability observed in the data. These analyses are presented in detail in the following sections, along with the data that emerged from them.

Results

Of the 16 initial critical items one was an exploratory item, and another one was removed due to high rates of incorrect responses across the board. The latter was disproportionately difficult for the majority of children. The final number of trials that were analyzed for each participant was 14.

For both groups, we collected participants' chronological age (in months). For the autistic group, we also gathered standardized measures, including VMA (from PPVT-III), which serve as matching variables and potential explanatory factors.

Correct Response Analyses

In terms of task performance, both groups performed above chance levels, with overall high picture selection accuracy. For critical items, the ASD group had a mean accuracy of $M = 77.21$ ($SD = 41.98$) and the TD group $M = 83.58$ ($SD = 37.05$). For filler items, the ASD group's mean accuracy was $M = 88.83$ ($SD = 31.51$), and the TD group's was $M = 93.69$ ($SD = 24.30$).

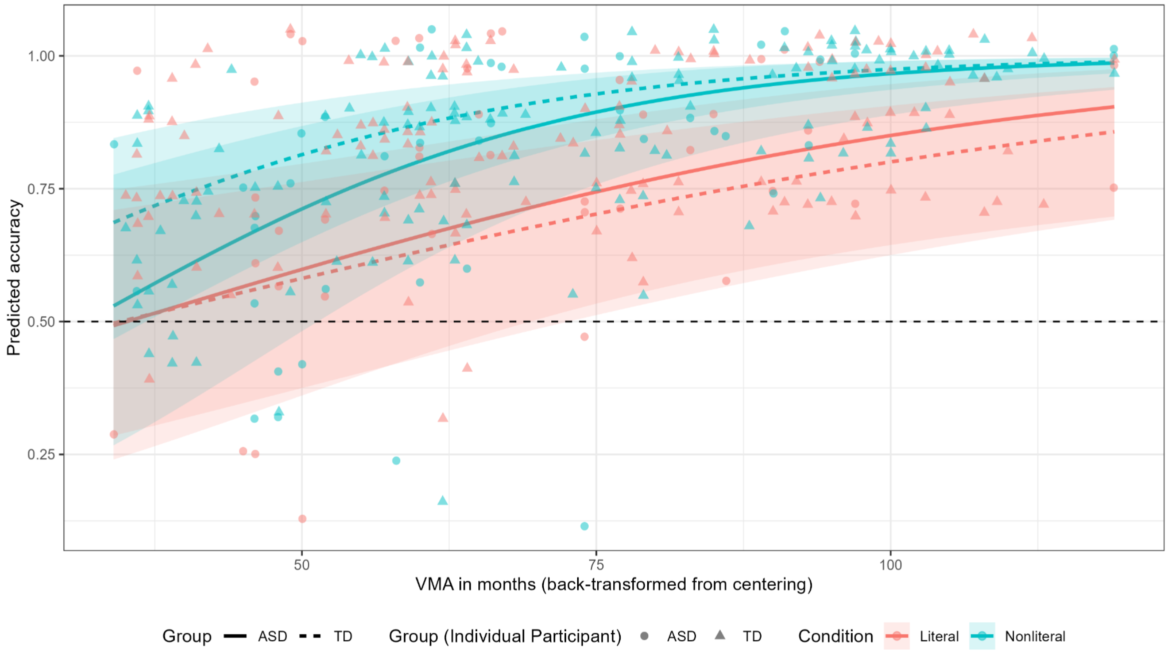
To examine the effects of group, condition, category, and verbal mental age, we fitted a generalized linear mixed model (estimated using *bobyqa* optimizer, binomial distribution, Logit link) on correct responses (accurate performance or correct choice of the target picture, i.e., choice of literal picture in the literal interpretation trials, choice of nonliteral picture in the nonliteral interpretation trials). As anticipated in the Analysis section, the model included verbal mental age as a continuous variable (centered), group, condition and type of item as fixed effects (sum-contrast coded) together with a three-way interaction between VMA, group and condition and another three-way interaction between group, condition and category. Random intercepts for participants and items were included in the model as random effects. The results of the model are summarized in Table 2. From visual inspection of the data (Figure 2), nearly all children appeared to perform above chance (chance = .50), consistently selecting the target picture with substantial accuracy.

The model revealed differences in accuracy. There is a significant interaction between VMA and CONDITION (z = –3.02, p = .002). Specifically, in the nonliteral condition, performance tends to increase as VMA increases. There is also a three-way interaction between GROUP, CONDITION and CATEGORY (z = 2.56, p = .011). We failed to find a significant difference between groups (ASD *versus* TD) across VMA for each of our conditions (The results are shown in Table 2, while the pattern of results is shown in Figure 2). The first interaction was not analyzed further because we were mainly interested in the differences between groups.

Table 2. ***Logistic regression model output for correct responses.***

Term	**β**	**95% CI**	**z-value**	**p-value**
Intercept	1.93	[1.42, 2.45]	7.34	< .001
VMA (centered)	0.04	[0.03, 0.04]	7.03	< .001
GROUP	−0.40	[−0.80, 0.00]	−1.94	.052
CONDITION	−0.25	[−0.55, 0.05]	−1.61	.108
CATEGORY	−0.84	[−1.82, 0.14]	−1.69	.091
VMA × GROUP	0.01	[−0.01, 0.03]	0.57	.566
VMA × CONDITION	−0.02	[−0.04, −0.01]	−3.02	.002
GROUP × CONDITION	−0.08	[−0.69, 0.52]	−0.27	.790
GROUP × CATEGORY	0.53	[−0.02, 1.08]	1.90	.057
CONDITION × CATEGORY	−2.09	[−2.65, −1.53]	−7.34	< .001
VMA × GROUP × CONDITION	0.00	[−0.03, 0.03]	−0.02	.986
GROUP × CONDITION × CATEGORY	1.44	[0.34, 2.54]	2.56	.011

In order to establish the origin of the three-way interaction, additional multiple comparisons with Tukey contrasts were run. For the autistic group in the MHI category, no significant differences were observed between conditions. However, for the LVC category, there was a significant difference between conditions (p = .01) with more accurate performance in the nonliteral condition. For the control group, significant differences were observed between conditions in both item categories (i.e., MHI and LVC) (p < .001), with better performance in the nonliteral condition for LVC items and slightly better performance in the literal condition for MHI category of items. We also explored differences between groups for each item category and condition. No significant differences were observed except for the literal condition in the MHI category of items (p = .001) in which autistic participants performed below their TD verbal mental age matched-peers. Now, while for the control group there is differential performance depending on the category of the items in different conditions or interpretation trials, this pattern is partially absent in autistic participants.



Note: The ribbon around the lines indicates 95% confidence intervals

Figure 2. ***Results of correct responses with verbal mental age as a continuous predictor and a superimposed fitted accuracy slope by age. Individual dots and triangles show participant averages. Error ribbons show 95-percent confidence intervals.***

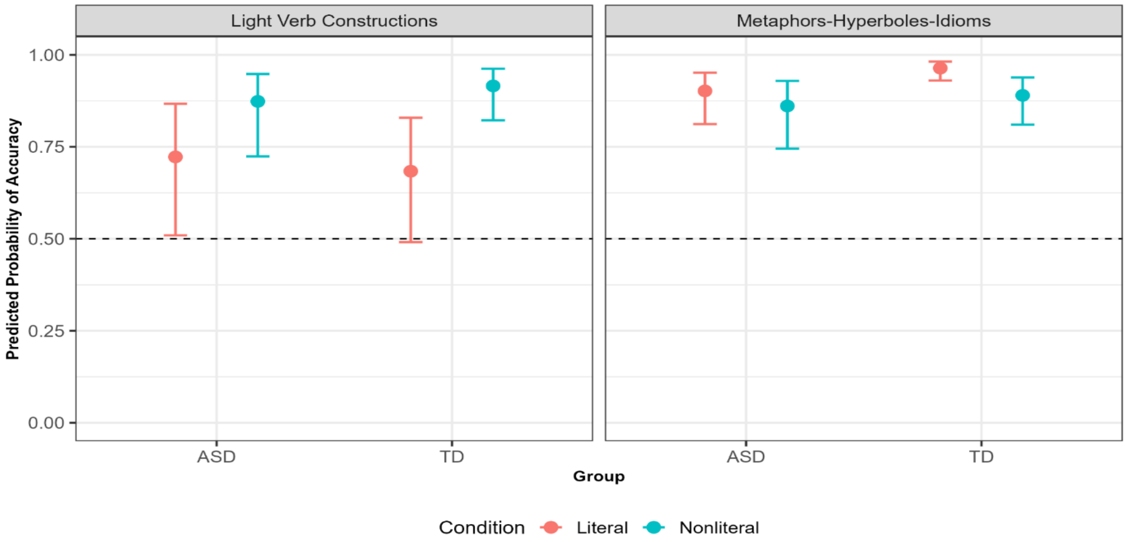


Figure 3. ***Results of correct responses per group on type of item x condition. Error bars denote one standard error of the mean.***

For autistic children, additional analyses confirmed that performance was related to verbal mental age, and it was not predicted by chronological age. Two models for accuracy were run for autistic participants only, with either chronological age or verbal mental age (centered) as continuous predictor variables (see Appendix Tables A2–A3). The model showed a main effect of VMA (z-value = 3.74, $p < .001$) as compared to no effect in the model with CA as an independent variable (z-value = 1.54, $p = .123$). CA was not a good predictor of performance in the autistic group.

Reaction Time Analyses

Excluding incorrect responses, both groups showed relatively fast response times in filler trials. The ASD group had a mean RT of $M = 2929.73$ ms ($SD = 1384.63$), while the TD group responded more quickly, with a mean RT of $M = 2413.80$ ms ($SD = 1120.50$). This intergroup difference was statistically significant, with autistic participants responding more slowly than their typically developing peers ($t(836) = 8.17$, $p < .001$). These values reflect the relative ease of target selection in filler items and establish a baseline for comparing response times in critical trials. For critical trials, the mean RT for the ASD group was $M = 3913.83$ ms ($SD = 1899.36$), compared to $M = 3127.03$ ms ($SD = 1637.42$) for the TD group. Comparing critical to filler trials provides context for understanding how each group responded to the task. Within-group comparisons showed significant differences in RTs between critical and filler trials for both groups (ASD: $t(762) = -9.15$, $p < .001$; TD: $t(2353) = -13.97$, $p < .001$), indicating that critical trials were more demanding.

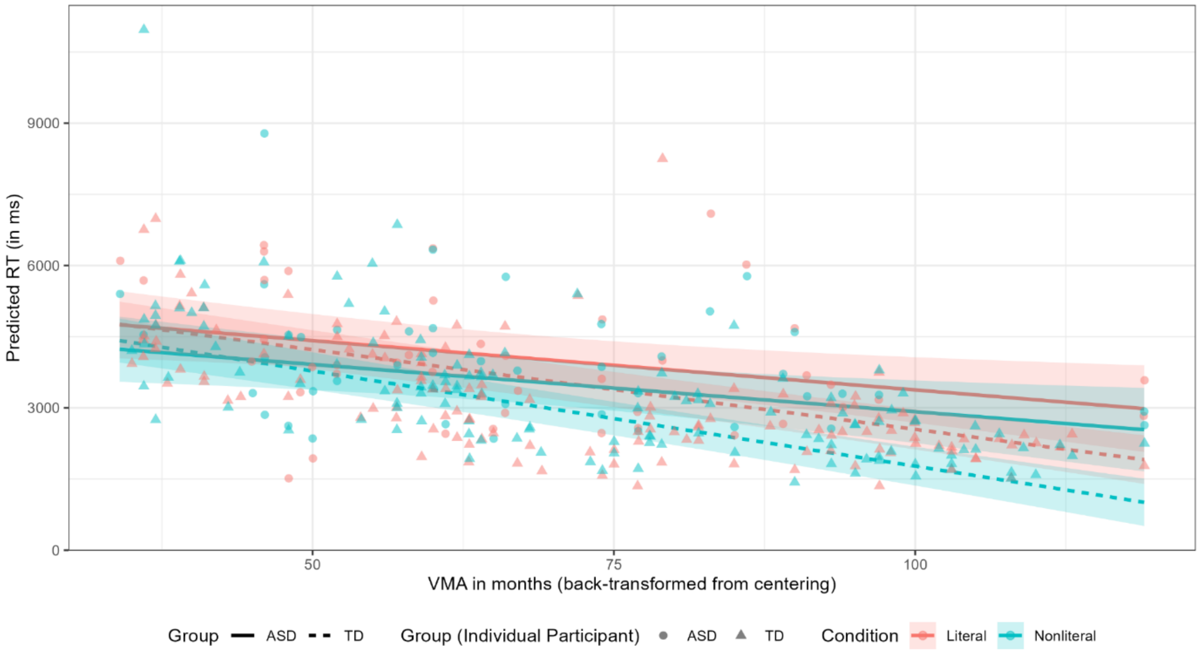
To formally test effects on RTs, we fitted a linear mixed-effects model using RT (in ms) as the dependent measure. Extremely fast or slow responses were excluded. We calculated the mean RT for each item and for each participant per group and we excluded trials that exceeded 3SD above those means. 194 trials (out of 5146 observations) were excluded (3%) according to these criteria. This choice of outlier removal did not affect the overall interpretation of results. RTs were analyzed only for correct responses. This led to removing 17% of critical trials.

The model included a three-way interaction between verbal mental age (centered), group, and condition, together with a second three-way interaction between group, condition, and category, and all lower-order effects. Random intercepts for participants and items were included. The results are summarized in Table 3. The overall model revealed that there is an interaction for response times between VMA and GROUP ($t = 2.10$, $p = .037$), in which we get shorter response times as VMA increases particularly in the TD group. The control group shows consistently quicker responses as compared to the ASD group which shows slightly slower responses. Another significant interaction was found between CONDITION and CATEGORY ($t = 5.88$, $p < .001$). Multiple comparisons of means using Tukey contrasts revealed the source of the interaction. While in the LVC category of items, literal conditions are significantly slower than nonliteral conditions (p < .001), in the MHI category the reverse holds, nonliteral conditions take more time to get responded than literal conditions (p < .001). We did not find any significant differences between autistic and typically developing children regarding the conditions and categories of items, nor within each of these groups. Autistic children showed slower responses in general. Figure 5 displays overall RTs for each category of item, LVC and MHI.

Table 3. ***Regression model output for reaction latencies.***

Term	β	95% CI	t-value	df	p-value
Intercept	3,586.50	[3,346.20, 3,826.81]	29.25	37.80	< .001
VMA (centered)	−28.61	[−36.29, −20.92]	−7.30	160.21	< .001
GROUP	604.84	[275.51, 934.17]	3.60	168.56	< .001
CONDITION	90.38	[−56.73, 237.50]	1.20	1683.85	.229
CATEGORY	−200.84	[−579.50, 177.83]	−1.04	15.35	.315
VMA × GROUP	16.48	[1.12, 31.84]	2.10	160.10	.037
VMA × CONDITION	2.84	[−3.63, 9.31]	0.86	1676.11	.390
GROUP × CONDITION	135.80	[−157.62, 429.23]	0.91	1680.88	.364
GROUP × CATEGORY	−190.58	[−479.77, 98.60]	−1.29	1663.64	.197
CONDITION × CATEGORY	889.22	[592.73, 1,185.71]	5.88	1700.66	< .001
VMA × GROUP × CONDITION	−7.37	[−20.29, 5.54]	−1.12	1674.92	.263
GROUP × CONDITION × CATEGORY	−474.13	[−1,064.75, 116.49]	−1.57	1694.54	.116

Although autistic participants took slightly longer to respond, no differences were found either between groups or within each group between conditions or interpretation trials. The response pattern was similar for both groups. From the visual inspection of the data, one can conclude that the tendency is towards literal responses to take more time than nonliteral responses overall. However, once the item category is taken into account the pattern continues to be that for LVC but reverses for MHI items. The lack of condition effect in both autistic and typically developing children may very well stem from large within group variation on RTs.



Note: The ribbon around the lines indicates 95% confidence intervals

Figure 4. ***Results of response times with verbal mental age as a continuous predictor and superimposed fitted slopes. Individual dots and triangles show participant averages. Error ribbons show 95-percent confidence intervals.***

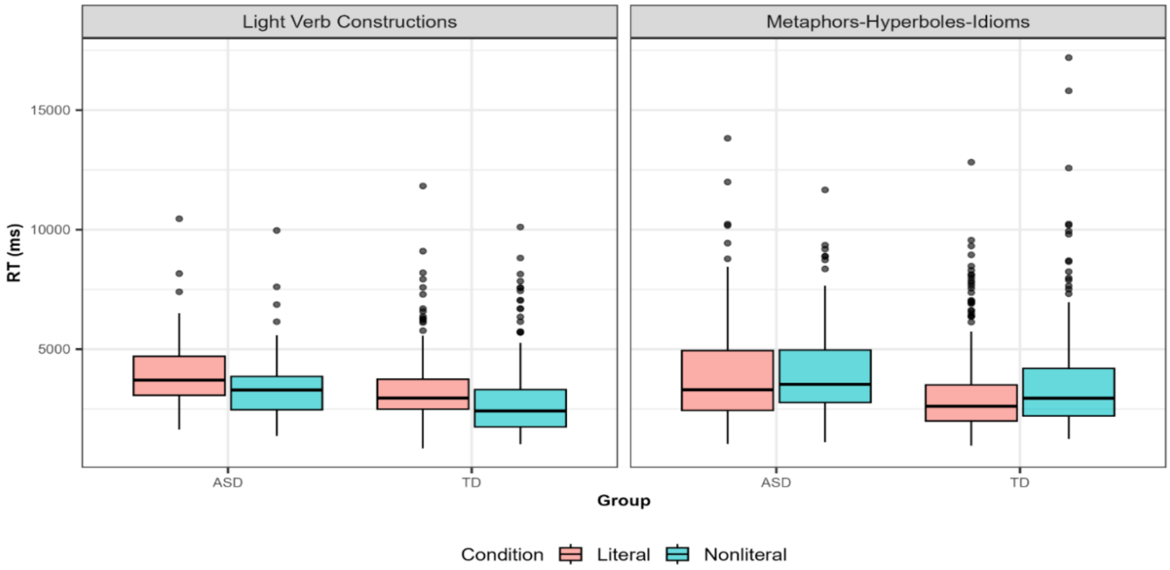


Figure 5. Turkey box plots to represent RT by group, condition and category.

We were interested in mean performance across literal and nonliteral trials broadly and within that we explored the effect of LVC and MHI specifically. To sum up these results, with increasing verbal mental age, performance changed for the best for both groups in both conditions (i.e., nonliteral and literal interpretation trials). There was an effect of the group, critically suggesting that performance on correct responses from autistic children differed from TD peers, but that only seemed to be the case in the MHI category of items, in particular, in the literal condition. No other differences were found between both groups. While the performance in different categories of items (LVC *versus* MHI) differed regarding both literal and nonliteral interpretation trials (i.e., conditions) for controls, the trend was much more subtle for autistic children for which there was a difference between interpretations for LVC, but not for MHI. The only difference we found was in the literal condition in the MHI category of items. Regarding response times, the picture that emerges is quite different. There were no differences between groups for each condition, nor within each group regarding each condition x category of item pairing. What seemed to be the case is that overall, autistic children tended to respond more slowly than their TD peers across verbal mental ages. The general trend where nonliteral responses are faster than literal responses in LVC and slower than literal responses in MHI does not hold consistently within each group. This inconsistency is likely due to significant individual variation among participants, but it could also be influenced by how participants understood or engaged with the task, as this might affect their response patterns or make certain items more challenging. Additionally, variability in familiarity with specific items or expressions could influence processing speed, potentially disrupting the trend. Factors like reduced attention or fatigue during parts of the task might also contribute to the inconsistencies in response times, although we attempted to visualize this, and it did not seem to be the case.

All the above analyses considered the 42 autistic participants as a group. Given the variable performance specifically at the youngest verbal mental ages (and the variable profiles that we included with minimally verbal children to highly verbal), we decided to explore individual differences in the autistic group descriptively, focusing on qualitative observations (see Figure 6), following Panzeri et al. (2022), Saban-Bezalel et al. (2019) and many others on individual variation of performance. Group mean performance might have been masking a wide range of individual variation on performance, limiting interpretability, so a closer inspection of the data seemed warranted.

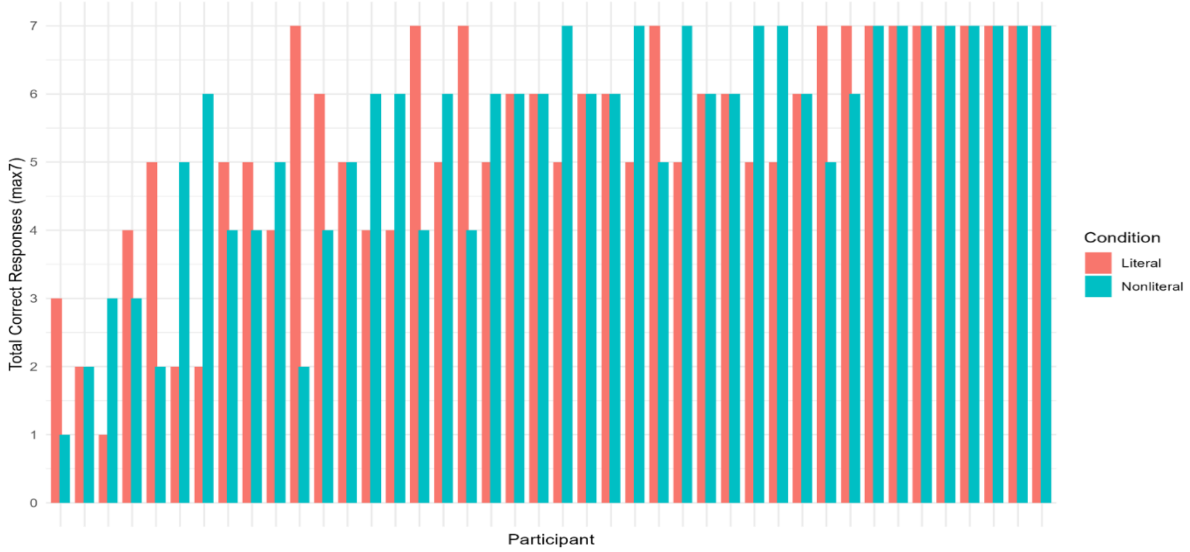


Figure 6. Individual autistic participants' scores (sum of correct responses) in literal and nonliteral interpretation trials.

Descriptive and Exploratory Post-hoc Analyses of the Autistic Group

Leaving reaction latencies aside, in this final analysis, we focused on performance on individual participants, descriptively. The idea was to simply describe profiles of autistic children regarding their response performance. We looked at their performance across conditions and types of items and described it in terms of z-scores (i.e., above or below the group's mean). We also wanted to explore whether their performance was related to child individual characteristics (i.e., CA, VMA, the difference between these two variables, and non-verbal IQ). In some cases, we could also add autism severity scores to these variables. By *autism severity,* we refer to the ADOS calibrated severity score, a metric derived from the transformation of ADOS-2 raw scores, as proposed by Gotham, Pickles and Lord (2009), with a range from 1 to 10. However, as we did not have ADOS scores for all children, we could not use autism severity as a variable for every participant. We mention autism severity scores only in the case of subgroups for which we had more than half of such scores. Performance was far more variable than that of TDs, and some fairly exploratory patterns were found.

There were four children who overall performed low across conditions and item categories, showing a very small proportion of correct responses (LVC *lit.* -0.9*z and nonlit. -0.7z; MHI lit. -1z and nonlit. -1z*). Their profile was heterogeneous: NVIQ ranging from 73 to 113; VMA from 3;10 years to 7;6 years; difference between CA and their VMA from 6 months to 2;9 years; and autism severity from 5 to 7.

By contrast, there were nine autistic children who had slightly higher scores, above the group's mean (LVC *lit.* +0.7*z* and *nonlit.* +0.5*z;* MHI *lit.* +0.4*z* and *nonlit.* +0.5*z*). They performed *at ceiling* irrespective of condition or item category. The high scorers group formed a heterogeneous group with regard to NVIQ (ranging from 82 to 117), VMA (4;2 years to 9;11 years), difference between CA and their VMA (for those who got higher VMA: 1 month to 1;4 years, for those who got lower VMA: 1 month to 3;11 years).

Interestingly, there were four children who performed better in the literal interpretation trials than on the nonliteral ones across item categories (*lit.* LVC *M* = 100, *SD* = 0 *vs nonlit.* LVC *M*= 42.85, *SD*= 53.45; and *lit.* MHI *M*=100, *SD* = 0 *vs nonlit.* MHI *M*=57.14, *SD*=51.35; lit. +0.6*z* and *nonlit.* -0.6*z*). They seemed to be the only ones exhibiting a strong literalism (including LVC expressions). These four children formed again a heterogeneous group with respect to NVIQ (ranging from 83 to 106), VMA (3 years to 5;4 years), difference between CA and their VMA (from 7 months to 2;6 years), and autism severity (5 to 7, though one severity score was missing).

Discussion

The aim of this paper is to add to the literature on nonliteral language development in autistic children. In particular, we aimed to test whether autistic children go through a phase of strong literalism, such that they would understand highly frequent nonliteral uses of language in a word-by-word, compositional, way. For that purpose, we used a picture selection task where children were required to choose between representations of either the literal or nonliteral meaning of an expression, along with two distractors. This paradigm was chosen primarily because it allowed us to compare the accuracy and ease of access to literal *versus* the accuracy and ease of access to nonliteral meanings in both groups, rather than to assess children's preference for one interpretation over the other. The expressions we used as stimuli were not just pieces of conventional nonliteral language, but their nonliteral uses were conventional *and* more frequent than their word-by-word literal counterparts. The expressions included frequent idioms, hyperboles and metaphors, as well as light verb constructions.

The primary aim of this study was to compare literal and nonliteral interpretation trials to identify potential differences between groups and also within each group. To assess the potential impact of VMA, VMA was included as a linear predictor for both autistic and typically developing participants. Additionally, item categories (LVC *versus* MHI) were included for exploratory analysis, based on the results obtained in Castroviejo et al. (2024). The study aimed to answer the following two research questions. The first one was:

RQ1: Are there any differences in the performance of TD and autistic children when interpreting nonliteral expressions across verbal mental age?

Our hypothesis was that autistic children's performance on nonliteral uses in general would be lower compared to their TD peers, even when matched on VMA. We were aware that some empirical evidence tends to support the opposite hypothesis (see Kalandadze, 2019), as well as some theories (Andrés-Roqueta & Katsos, 2017) that hold that comprehension of the nonliteral uses of the expressions in the current experiment would relate to VMA. However, based on some contrary evidence supporting literalism, and the fact that literalism is an interpretive tendency for many autistic people, especially in their infancy and youth, we hypothesized that we would find group differences, expressed either in difficulties in accessing nonliteral meanings, or in a relative ease in accessing the literal ones (or both at the same time). We did not have any specific predictions for autistic participant's performance in item categories.

With regard to RQ1, and contrary to what was predicted, results suggest that autistic children did not differ from VMA-matched TD children's performance regarding literal and nonliteral interpretation trials across development. The above-chance performance in the task suggests that the task itself was accessible and appropriate for assessing comprehension in both groups, indicating that autistic children are capable of processing both types of interpretations comparably to their TD peers when matched by VMA. In spite of those high scores, children from both groups exhibited an upwards linear performance as their vocabulary size developed.

In the same line, we observed that VMA (receptive vocabulary) was a predictor of success on both literal and nonliteral conditions in the autistic group. As said, since Norbury (2005), several authors have argued that language abilities (particularly lexical knowledge in Norbury's own case), play a key role in figurative language comprehension (e.g., Andrés-Roqueta & Katsos, 2020). Some authors have found that the hypothesis that figurative language comprehension relates to language level applies particularly well to *conventional* figurative language (Kasirer & Mashal, 2014; Olofson et al., 2014), which is the kind of nonliteral language that we were concerned with in our study: all the expressions in our list of stimuli were highly conventional in their nonliteral use. In this regard, as we did not compare conventional with novel nonliteral uses of language, we cannot add to the discussion concerning the role of linguistic development (or, rather, of receptive vocabulary levels) to these different kinds of nonliteral language. However, in line with Norbury (2005), Kasirer & Marshal (2014), and Olofson et al. (2014), we found a clear effect of VMA on overall task performance, suggesting that vocabulary breadth may facilitate the comprehension of both literal and nonliteral language in autistic children. As said, we had not predicted that nonliteral language would be so clearly predicted by VMA, because we thought that autistic children's difficulties with nonliteral language could be independent from their linguistic development. As we explain below, we found some evidence that this could be the case only for a small subgroup of children whose degree of literalism did not correspond to their vocabulary level.

We also explored differences between and within each group regarding item categories. We found that autistic participants only differed significantly from TD children in the MHI item category during literal interpretation trials, showing slightly lower performance. This was a surprising finding, which could suggest more difficulties in accessing literal interpretations when there is competition with nonliteral interpretations. On the other hand, we observed no significant differences between literal and nonliteral interpretation trials within the MHI item category for autistic participants, which would mean that it only makes sense to speak about some difficulties in autistic children when we compare autistic and non-autistic performance. However, we suspect that the observed difference is due to the heterogeneity of the autistic group: as we mention below, a majority of autistic children exhibited an interpretive flexibility comparable to their TD peers.

As already shown in the previous study on typical developmental trajectories (Castroviejo et al., 2024), TD controls exhibited higher accuracy scores in the nonliteral condition (as compared to the literal condition) for LVCs across developmental stages. The opposite pattern emerged for MHI items, in which case the literal condition got higher correct responses in comparison to the nonliteral condition across ages. Crucially, the difference between both types of items for TD controls lay in the literal interpretation trials. Autistic participants in our sample also followed that performance pattern except for the literal interpretation trials in the MHI item category. Underlyingly, their overall performance pattern seemed to be similar to the bimodal performance of TD children.

Concerning response times, autistic participants were overall slower to respond than their TD peers in both conditions as well as in the filler trials. There were apparent differences between groups across verbal mental ages. With increasing age, TD children’s responses became faster, possibly meaning that the task became easier for them. That seems not to be the case for autistic children. The task might have still been somewhat demanding or effortful, and attention and/or motor skills may have also played a role. No differences were found between literal and nonliteral interpretation trials between groups and within each group, nor for each item category. What we found is that once we collapsed groups, the emerging pattern converges with the results from the picture selection task.

Our second research question was the following:

RQ2: Do autistic children exhibit an easier access to literal, compositional meanings at different developmental stages?

Despite being implicitly addressed in the preceding research question, autistic children at group level were not more accurate in literal interpretation trials than in nonliteral interpretation trials. Nor did they perform better in literal interpretation trials at any point across VMA development. Most importantly, we did not find differences across interpretation trials in response times for autistic children. Here we found differences in picture selection between interpretation trials in the LVC item category, with more accurate performance in nonliteral interpretations and no differences between conditions in the MHI item category. However, these empirical findings might underlyingly suggest bimodal performance as in typical development. This would most likely be the case if there were not so much within-group variation. In other words, there might be a lingering tendency toward better performance in literal interpretation trials for MHI and in nonliteral ones for LVC.

One reason why we may not have found support for a general pattern of literalism may be the lack of ecological validity. This forced-choice paradigm, although widely used, falls short of accurately representing interpretation in naturalistic settings. Still, even with a relatively modest sample size and a minimally demanding task with visuals as supportive contexts, we found some subgrouping, in particular a few children who did better in literal interpretation trials, therefore exhibiting literalism.

Although 25 autistic children exhibited a behavior similar to that of TDs of similar vocabulary size, we observed that a small number of children (four) displayed a lower performance in all items and conditions, that another group of nine children performed *at ceiling* in all items and conditions (LVC 100% in both conditions; MHI *lit.* 100% and MHI *nonlit.* 98%), and that a group formed by another four participants adjusted to a strong literalist profile. Concerning the first group of four children, whose performance was lower than the rest in the literal and nonliteral conditions, our hypothesis is that their structural language abilities did not correspond to their vocabulary size: that is, despite being VMA-matched to TDs on receptive vocabulary, their sentence comprehension abilities were probably not matched to that of their TD peers.

The group that was more interesting for the purposes of this paper was the group of four children who were more accurate than their VMA-matched TD peers only in the literal condition in both categories (LVC ASD 100% vs TD 60% and MHI ASD 100% vs TD 92%), and who performed significantly less accurately in the nonliteral condition (ASD; LVC 43% and MHI 57%; TD; LVC 79% and MHI 77%). Children in this group were more accurate on literal interpretation trials than on nonliteral interpretation trials, that is, they seemed to have obvious difficulties understanding expressions in a non-compositional way. It is particularly noteworthy that they were 100% accurate on the literal interpretation of expressions like *hacer la cama* ('to make the bed') and below 50% on their nonliteral interpretation. As mentioned above, these four children, on the other hand, formed a heterogeneous group with respect to NVIQ (ranging from 83 to 106), VMA (3 years to 5;4 years), the difference between CA and their VMA (from 7 months to 2;6 years), and autism severity (5 to 7, though one severity score was missing). This suggests that literalism may be more acute in some autistic individuals than in others, and that it may relate to factors we did not explore. Perhaps literalism is characteristic of autism in the sense that only autistic individuals exhibit such a way of understanding nonliteral language, while at the same time not being widespread in the autistic population. In other words, it is not observed in all autistic individuals, and as such cannot be considered a defining feature, but it can be considered a cue validity feature (Rosch, 1978).

The paradigm that we employed had some limitations, so the current findings should be interpreted with caution. First, the task was simple and easy, as we already mentioned in several places and as accuracy results showed, and children were given the right interpretation as an option, which is not what occurs in real life contexts. In general, multiple-choice paradigms significantly reduce the uncertainty surrounding the comprehension of any utterance of an expression used nonliterally. Second, an obvious limitation of the current study is the relatively small size of our autistic sample which naturally impacts on the exploratory analysis of possible subgroups. The relatively low number of autistic participants and consequently low statistical power for detecting true effects was in large part due to the challenges of recruiting participants. Accordingly, findings for subgroups should be seen as an exploration of possible interpretive performance patterns coming from individual variation. Increasing the autistic sample size would allow for stabler developmental trajectories and possibly more homogeneous clusters. A further constraint of child-friendly paradigms is the reduced number of critical items used with the addition of an unbalanced number of item types in each item category.

Two limitations of our study relate to the notion of VMA. The first limitation is that our VMA measures are of receptive vocabulary, while the task consisted in understanding sentences. Especially in the case of autistic children, receptive vocabulary may not provide an adequate measure of linguistic development, and, in fact, we suspect that the poor performance in both conditions of a subgroup of autistic children may relate to the fact that their linguistic abilities did not correspond to the VMA measure. The second limitation is assuming that TD children's receptive vocabulary is aligned with their chronological age, without verifying this through direct testing. We did not test receptive vocabulary in the TD participants because they constituted the experimental group of a TD developmental study in which VMA was not a variable. This implied that we had to assume that chronological age would match verbal mental age in the TD group.

Lastly, in order to try to explain why certain subgroups performed the way they did, we realized that we had limited information about the participants in the study. As mentioned above, we saw nothing in common between the children who exhibited a stronger literalist behavior, for instance. In that respect, we missed having data about ToM abilities, executive function, local processing or rigid behaviors, all of which may relate to literalism. However, since we had ADOS-2 (Lord et al., 2012) scores for more than half of the participants, we were able to observe that subgroups were also heterogeneous with respect to the social-affective scores of the ADOS-2. These scores can be considered an indirect measure of ToM abilities, further highlighting the complexity of the subgroup differences.

Conclusion

The current study contributes to research studying nonliteral language in providing empirical evidence on how different expressions in Spanish are interpreted and processed by both autistic and typically developing pre-school and early elementary school-aged children across a wide verbal mental age range (three to nine). In particular, this study compared autistic children to their TD peers on their comprehension of expressions that had a dominant nonliteral meaning. We wanted to see whether autistic children could exhibit a strong form of literalism (word-by-word processing) with that kind of expressions.

Overall, our results revealed that autistic children performed similarly to typically developing peers. Their verbal mental age trajectories did not differ between the groups either, regardless of the literal or nonliteral interpretation trials. Finally, their correct response performance in both groups increased with verbal mental age. Contrary to expectations, difficulties with nonliteral uses of language were not spotted across the autistic spectrum. However, typically developing children performed differently in two types of nonliteral expressions: they were more accurate with respect to literal conditions for Metaphors, Hyperboles and Idioms and with respect to the nonliteral conditions for Light Verb Constructions. This performance pattern was not as clearly observed in the autistic group, as no differences were found between literal and nonliteral interpretation trials for Metaphors, Hyperboles and Idioms, although we hypothesize that the lack of a similar dichotomous pattern is masked by the variability in individuals' performances.

In trying to account for such variable performance, we observed that a small group of children were strongly literalists, performing better in literal interpretation trials even in examples involving light verb constructions. At the same time, there were some very good performers who could flexibly entertain both interpretations. This suggests that there is a lot of variability concerning literalism across the spectrum. However, given the characteristics of our design, we cannot exclude that the literalism observed in many autistic individuals may appear in more demanding communicative contexts. Future work should explore further this possible heterogeneity in pragmatic profiles within the spectrum and relate eventual subtypings to other variables or traits.

References

Allott, N., & Textor, M. (2022). Literal and metaphorical meaning: in search of a lost distinction. *Inquiry*, 1–28. https://doi.org/10.1080/0020174X.2022.2128867

Andrés-Roqueta, C., & Katsos, N. (2017). The Contribution of Grammar, Vocabulary and Theory of Mind in Pragmatic Language Competence in Children with Autistic Spectrum Disorders. *Frontiers in Psychology, 8*, 996. https://doi.org/10.3389/fpsyg.2017.00996

Asperger, H. (1991). 'Autistic psychopathy' in childhood. In U. Frith (Ed.), *Autism and Asperger Syndrome* (pp. 37–92). Cambridge University Press. https://doi.org/10.1017/CBO9780511526770.002

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). *The lme4 package* (Version 2) [R package]. https://cran.r-project.org/web/packages/lme4/

Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review, 112*(1), 193–216. https://doi.org/10.1037/0033-295X.112.1.193

Castroviejo, E., Ponciano, M., Hernández-Conde, J.V. and Vicente, A. (2024), Development of nonliteral interpretations in typically developing Spanish speaking children: light verb constructions and figurative expressions. *Studia Linguistica, 78*, 8-34. https://doi.org/10.1111/stul.12222

Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H., & Vulchanova, M. (2016). Can You Play with Fire and Not Hurt Yourself? A Comparative Study in Figurative Language Comprehension between Individuals with and without Autism Spectrum Disorder. *PLoS One, 11*(12), e0168571. https://doi.org/10.1371/journal.pone.0168571

Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H., & Vulchanova, M. (2017). Can you tell it by the prime? A study of metaphorical priming in high-functioning autism in comparison with matched controls. *International Journal of Language & Communication Disorders, 52*(6), 766-785. https://doi.org/10.1111/1460-6984.12314

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar Inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 40*(9), 1104–1117. https://doi.org/10.1007/s10803-010-0960-8

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective Pragmatic Impairment in Autism Spectrum Disorder: Indirect Requests Versus Irony. *Journal of Autism and Developmental Disorders, 48*(9), 2938–2952. https://doi.org/10.1007/s10803-018-3561-6

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *PPVT-III Peabody: Test de vocabulario en imágenes.* Tea Ediciones.

Fellbaum, C., Geyken, A., Herold, A., Koerner, F., & Neumann, G. (2006). Corpus-based Studies of German Idioms and Light Verbs. *International Journal of Lexicography, 19*(4), 349–360. https://doi.org/10.1093/ijl/ecl031

Fleischer, W. (1997). *Phraseologie der deutschen Gegenwartssprache*: De Gruyter. https://doi.org/10.1515/9783110947625

Gernsbacher, M. A., & Pripas-Kapit, S. R. (2012). Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. *Metaphor and Symbol, 27*(1), 93–105.

Geurts, B., Kissine, M., & Tiel, B. van. (2019). Pragmatic reasoning in autism. In K. Morsanyi & R. M. J. Byrne (Eds.), *Thinking, Reasoning, and Decision Making in Autism* (1st ed., pp. 113–134). Routledge. https://doi.org/10.4324/9781351060912-6

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

He, A. X., & Wittenberg, E. (2020). The acquisition of event nominals and light verb constructions. *Language and Linguistics Compass, 14*(2), e12363. https://doi.org/10.1111/lnc3.12363

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar Implicature in Absence of Epistemic Reasoning? The Case of Autism Spectrum Disorder. *Language Learning and Development, 14*(3), 224–240. https://doi.org/10.1080/15475441.2017.1343670

Kasirer, A., & Mashal, N. (2014). Verbal creativity in autism: Comprehension and generation of metaphoric language in high-functioning autism spectrum disorder and typical development. *Frontiers in Human Neuroscience, 8,* 615.

Kasirer, A., & Mashal, N. (2016). Comprehension and generation of metaphors by children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 32,* 53–63.

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

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism: The International Journal of Research and Practice, 22*(2), 99–117. https://doi.org/10.1177/1362361316668652

Kanner, L. (1944). Early infantile autism. *The Journal of Pediatrics, 25,* 211–217. https://doi.org/10.1016/S0022-3476(44)80156-1

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

Kissine, M., Cano Chervel, J., Carlier, S., De Brabanter, P., Ducenne, L., Pairon, M.-C., Deconinck, N., Delvenne, V., & Leybaert, J. (2015). Children with autism understand indirect speech acts: Evidence from a semi-structured act-out task. *PloS One, 10,* e0142191. https://doi.org/10.1371/journal.pone.0142191

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models* (Version 2.0-7) [R package]. https://cran.r-project.org/package=lmerTest

Lampri, S., Peristeri, E., Marinis, T., & Andreou, M. (2023). Figurative language processing in autism spectrum disorders: A review. *Autism Research,* 1–16. https://doi.org/10.1002/aur.3069

Lord C., Luyster R., Gotham K., & Guthrie W. (2012). *Autism Diagnostic Observation Schedule.* 2nd. Torrence, CA: Western Psychological Services. (ADOS-2) *Manual (Part II): Toddler Module.*

MacKay, G., & Shaw, A. (2004). A comparative study of figurative language in children with autistic spectrum disorders. *Child Language Teaching and Therapy, 20*(1), 13–32. https://doi.org/10.1191/0265659004ct261oa

Marocchini, E., Di Paola, S., Mazzaggio, G., & Domaneschi, F. (2022). Understanding indirect requests for information in high-functioning autism. *Cognitive Processing, 23*(1), 129–153. https://doi.org/10.1007/s10339-021-01056-z

Martín-González, I., Schroeder,K., Castroviejo, E., Lossius Falkum, I., & Vicente, A. (2025). Novel metaphor processing in young autistic children. *Language Development Research, 5*(2), 28-66. https://doi.org/10.34842/im-g

Mazzaggio, G., Foppolo, F., Job, R., & Surian, L. (2021). Ad-hoc and scalar implicatures in children with autism spectrum disorder. *Journal of Communication Disorders, 90*, 106089. https://doi.org/10.1016/j.jcomdis.2021.106089

McMahon C.M., Lerner M.D., & Britton N. (2013) Group-based social skills interventions for adolescents with higher-functioning autism spectrum disorder: a review and looking to the future, *Adolescent Health, Medicine and Therapeutics, 4,* 23-38, https://doi.org/10.2147/AHMT.S25402

Melogno, S., Pinto, M., & Di Filippo, G. (2017). Sensory and Physicopsychological metaphor comprehension in children with ASD: A preliminary study on the outcomes of a treatment. *Brain Sciences, 7*(12), 85. https://doi.org/10.3390/brainsci7070085

Morra, L. (2016). Raising awareness of how Asperger persons perceive their capacity to use Metaphors. *Medicina e Storia, 11*(2), 129–146.

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology,* I(3), 383–399. https://doi.org/10.1348/026151005X26732

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language, 70*(3), 491–538. https://doi.org/10.2307/416483

O'Shea, A., Cersosimo, R. & Engelhardt, P.E.(2024). Online Metaphor Comprehension in Adults with Autism Spectrum Disorders: An Eye Tracking Study. J *Journal of Autism and Developmental Disorders,* 1-15. https://doi.org/10.1007/s10803-024-06562-5

Olofson E. L., Casey D., Oluyedun O., van Herwegen J., Beccera A., Rundblad G. (2014). Youth with autism spectrum disorder comprehend lexicalized and novel primary conceptual metaphors. *Journal of Autism and Developmental Disorders, 44*(10), 2568–2583. http://doi.org/10.1007/s10803-014-2129-3

Ozonoff, S., & Miller, J. N. (1996). An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language, 52*(3), 411–434. https://doi.org/10.1006/brln.1996.0022

Panzeri, F., Mazzaggio, G., Giustolisi B, Silleresi, S., & Surian, L. (2022). The atypical pattern of irony comprehension in autistic children. *Applied Psycholinguistics, 43(*4), 757-784. https://doi.org/10.1017/S0142716422000091

Pastor-Cerezuela, G., Fernández-Andrés, M.-I., Tordera-Yllescas, J. C., & González-Sala, F. (2020). Metaphor comprehension in children with and without autism spectrum disorder. *Research in Autism Spectrum Disorders*, *76*, 101588. https://doi.org/10.1016/j.rasd.2020.101588

Paul, R., & Cohen, D. J. (1985). Comprehension of Indirect Requests in Adults with Autistic Disorders and Mental Retardation. *Journal of Speech, Language, and Hearing Research*, *28*(4), 475–479. https://doi.org/10.1044/jshr.2804.475

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.-P., & Geurts, B. (2009). Pragmatic Inferences in High-Functioning Adults with Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders, 39*(4), 607–618. https://doi.org/10.1007/s10803-008-0661-8

Psychology Software Tools. (2016). *E-Prime 3.0* (Version 3.0) [Computer software]. https://support.pstnet.com/

R Core Team (2022). *R: A language and environment for statistical computing.* Vienna, R Foundation for Statistical Computing. https://www.R-project.org/.

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter International Performance Scale - Third edition.* Stoelting Company.

Rosch, Eleanor (1978), Principles of categorization, in Eleanor Rosch & Barbara B. Lloyd (Eds.), *Cognition and Categorization*, Hillsdale, New Jersey: Lawrence Erlbaum, pp. 27–48.

Rundblad G., & Annaz D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism, 1,* 29-46. https://doi.org/10.1177/1362361309340667

Rutter, M., LeCouteur, A. & Lord, C. (2003). *Autism Diagnostic Interview - Revised*. Los Angeles, CA: Western Psychological Services. Adaptación española 2006.

Saban-Bezalel, R., Dolfin, D., Laor, N., & Mashal, N. (2019). Irony comprehension and mentalizing ability in children with and without Autism Spectrum Disorder. *Research in Autism Spectrum Disorders, 58,* 30–38. https://doi.org/10.1016/j.rasd.2018.11.006

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The understanding of scalar implicatures in children with Autism Spectrum Disorder: Dichotomized responses to violations of informativeness. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01266

Su, Y., & Su, L.-Y. (2015). Interpretation of logical words in mandarin-speaking children with autism spectrum disorders: Uncovering knowledge of semantics and pragmatics. *Journal of Autism and Developmental Disorders, 45*. https://doi.org/10.1007/s10803-014-2350-0

Taylor, J. R. (2006). Polysemy and the lexicon. In K. G., M. Achard, R. Dirven, & F. J. Ruiz de Mendoza Ibánez (Eds.), *Cognitive Linguistics: Current Applications and Future Perspectives* (pp. 51-80). Berlin: Mouton de Gruyter.

Van Herwegen, J., & Rundblad, G. (2018). A cross-sectional and longitudinal study of novel metaphor and metonymy comprehension in children, adolescents, and adults with autism spectrum disorder. *Frontiers in Psychology*, 9, 945. https://doi.org/10.3389/fpsyg.2018.00945

Van Tiel, B., & Kissine, M. (2018). Quantity-based reasoning in the broader autism phenotype: A web-based study. *Applied Psycholinguistics, 39*(6), 1373–1403. https://doi.org/10.1017/S014271641800036X

Vega-Moreno, R. E. (2005). Idioms, transparency and pragmatic inference. *UCL Working Papers in Linguistics, 17,* 389–425.

Vicente, A., Ábalos, Z., Martín-González, I., Ramos-Cabo, S., Castroviejo, E. (2024) Autism. In Vandenbroucke, M., Declercq, J., Brisard, F. & D'hond, S. (Eds.), *Handbook of Pragmatics* (pp. 1-40). John Benjamins. https://doi.org/10.1075/hop.27.aut4

Vicente, A., & Falkum, I. L. (2023). Accounting for the preference for literal meanings in autism spectrum conditions. *Mind & Language, 38,* 119–40. https://doi.org/10.1111/mila.12371

Vulchanova, M., Talcott, J. B., Vulchanov, V., Stankova, M., & Eshuis, H. (2012). Morphology in autism spectrum disorders: Local processing bias and language. *Cognitive Neuropsychology, 29*(7–8), 584–600. https://doi.org/10.1080/02643294.2012.762350

Vulchanova, M., & Vulchanov, V. (2022). Rethinking figurative language in autism: What evidence can we use for interventions? *Frontiers in Communication, 7.* https://doi.org/10.3389/fcomm. 2022.910850

Walenski, M., & Love, T. (2017). The real-time comprehension of idioms by typical children, children with specific language impairment and children with autism. *Journal of Speech Pathology & Therapy*, 3(1), 130. https://doi.org/10.4172/2472-5005.1000130

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230-259). Palgrave.

Wittenberg, E., (2016). With light verb constructions from syntax to concepts. PhD Dissertation. Universität Potsdam. *Potsdam Cognitive Science Series.*

Data, Code and Materials Availability Statement

List of materials, data and analysis script are available on the open science framework online data OSF repository at: https://osf.io/v3kq9/

Ethics Statement

Ethical approval was issued by University of the Basque Country's (UPV/EHU) Ethics Committee for research with human beings (CEISH), code M10_2019_205. Written informed consent was obtained from parents/caregivers of all participants.

Authorship and Contributorship Statement

MP: Conceptualization, Software, Investigation, Data curation, Formal analysis, Writing - Original Draft Preparation, Writing - Review & Editing. AV: Conceptualization, Methodology, Resources, Writing - Original Draft Preparation, Writing - Review & Editing, Supervision. JVHC: Formal analysis. EC:

Conceptualization, Methodology, Resources, Writing - Original Draft Preparation, Writing - Review & Editing, Supervision. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

The authors wish to thank all the children and their families for their participation in this study. Special thanks are also extended to: Early Attention Service and Autism Association from Araba; APNABI Association from Bizkaia; CEIP Gabriel Celaya HLHI Eskola Publikoa and CPI Antonio Trueba IPI Eskola Publikoa in Portugalete; C.E.I.P. Itsasoko Ama Ikastetxea H.L.H.I. in Santurtzi (Spain). We are also grateful to our colleagues at the Lindy Lab (UPV/EHU), as well as to the editors of LDR and the three anonymous reviewers who commented on the original version of this paper. They have significantly helped us improve our work.

Earlier versions of this research were presented at the 31st Colloquium on Generative Grammar (CGG31, 2022); the 3rd Joint Meeting of the Society for Philosophy and Psychology (SPP) and the European Society for Philosophy and Psychology (ESPP) (SPP/ESPP, 2022); and the 9th Experimental Pragmatics Conference (XPRAG, 2022). We are grateful to the audiences at these events for their insightful feedback.

This research has been partially supported by grant PID2021-122233OB-I00, funded by MICIU/AEI/10.13039/501100011033 and by "ERDF/EU", and by the IT1537-22 Research Group (Basque Government). This work has also been supported by the Basque Government's Department of Education through its Predoctoral Training Programme for Non-Doctoral Research Staff (Programa Predoctoral de Formación de Personal Investigador No Doctor, 2020–2024) (M.P.).

Appendix

Table A1. ***Sentences used as experimental items. The nonliteral translations of the items are also shown.***

Item	Literal interpretation	Nonliteral translation / interpretation
Benito hace pesas	Benito makes (builds) weights.	Benito lifts weights.
El técnico corta el agua	The technician cuts the water.	The technician turns off the water.
Ibai se queda helado al oír la noticia	Ibai gets frozen when hearing the news.	Ibai is shocked by the news.
Idoia le da la mano a Martín	Idoia gives Martín her hand.	Idoia shakes Martín's hand.
Ione coge el autobús	Ione catches (lifts) the bus.	Ione takes the bus.
Juan hace la cama	Juan makes (builds) the bed.	Juan makes the bed.
Juan le pone mala a Elena	Juan makes Elena sick.	Juan drives Elena mad.
Pedro está hecho polvo	Pedro is made of dust.	Pedro is exhausted.
Sandra duerme con los angelitos	Sandra sleeps with the angels (literally).	Sandra sleeps with the angels / peacefully.
Sergio se muere de aburrimiento	Sergio is (literally) dying of boredom.	Sergio is dying of boredom / is extremely bored.
Silvia está en las nubes	Silvia is (physically) in the clouds.	Silvia has her head in the clouds / is distracted.
Tania se parte de risa	Tania splits herself with laughter.	Tania bursts out laughing / laughs uncontrollably.
Unax es una tortuga	Unax is a turtle.	Unax is a snail / is very slow.
Xabi pone la mesa	Xabi places the table (somewhere).	Xabi lays the table.

Table A2. Model results with verbal mental age as a predictor in autistic children

Term	β	95% CI	z-value	p-value
Intercept	1.73	[1.23, 2.24]	6.71	< .001
VMA (centered)	0.04	[0.02, 0.06]	3.74	< .001
CONDITION	−0.18	[−0.68, 0.32]	−0.69	.489
VMA × CONDITION	−0.02	[−0.05, 0.00]	−1.81	.071

Table A3. Model results with chronological age as a predictor in autistic children

Term	β	95% CI	z-value	p-value
Intercept	1.59	[1.07, 2.10]	6.05	< .001
AGE (centered)	0.02	[0.00, 0.04]	1.54	.123
CONDITION	−0.01	[−0.44, 0.42]	−0.03	.975
AGE × CONDITION	−0.02	[−0.04, 0.01]	−1.52	.129

License

Decoupling literalist behavior from children's early metaphor comprehension abilities

Mary Beth Neff and Ingrid Lossius Falkum
University of Oslo, Norway

Abstract: Children's literalist responses to metaphor comprehension tasks are often taken to indicate deficient metaphor comprehension. We aimed to decouple this assumed (literalist) performance–(metaphor) competence link and investigate whether children's observed literalism is best explained by an early difficulty with metaphor. We assessed 3- to 7-year-olds' metaphor comprehension abilities using different novel functional, attributional, and psychological metaphors in a between-subjects design. We found that when not provided with literal options, children could derive metaphorical interpretations successfully. This was further supported by longer reaction times for metaphorical over literal interpretations. However, when literal options were available, even adults predominantly chose them over metaphorical interpretations. These findings challenge the view that children's literalism stems solely from difficulty with metaphor and urge researchers to more clearly distinguish studies assessing sensitivity to metaphorical meaning from those investigating the ability to prioritize a metaphorical interpretation over a literal one.

Keywords: metaphor; children; literalism; experimental; pragmatics

Corresponding author(s): Mary Beth Neff, Department of Philosophy, Classics, History of Art and Ideas, Blindernveien 31 Georg Morgenstiernes hus 0313 Oslo, Norway. Email: marybeth-neff@gmail.com

ORCID ID(s): Mary Beth Neff: https://orcid.org/0000-0002-9549-5936; Ingrid Lossius Falkum: https://orcid.org/0000-0002-1203-8036

Citation: Neff, M.B., & Falkum, I. L. (2025). Decoupling literalist behavior from children's early metaphor comprehension abilities. *Language Development Research, 5*(2), 130–171.
http://doi.org/10.34842/ldr2025-819

Introduction

If you were to stub your toe and say, “Ah - my toe is on fire!”, there are at least two possible interpretations to take—that your toe is literally on fire or that your toe is in a lot of pain. Given the context, adults would likely infer that you were in a lot of pain. Children, however, may take you to mean that your toe was actually on fire. We explored why, between these alternatives, children often favor literal interpretations. Specifically, we investigated whether children’s preference for literal interpretations is best explained by early difficulties with understanding novel metaphors, as often suggested in previous research.

Recent evidence has suggested that children as young as 3 years old can make sense of novel metaphorical statements (5et al., 2024; Pouscoulous & Tomasello, 2020). However, despite this early ability, children often favor literal interpretations, even when the context does not support them (e.g., Martín-González et al., 2025) and sometimes up until preadolescence (e.g., Van Herwegen et al., 2013). This observed literalist preference among young children has often been referred to as their “literal bias,” a concept that has been a long-standing issue in research on metaphor comprehension development (see Vosniadou, 1987 and Winner, 1988/1997 for reviews). More generally, children’s fragile performance on pragmatic tasks involving nonliteral uses of language (e.g., metaphor, irony, implicature) is puzzling in light of the growing body of evidence showing children’s early pragmatic sophistication in other domains, including prelinguistic communication, word learning, and referential communication (see Matthews, 2014 and Zufferey, 2015 for reviews).

In early studies, researchers presumed children’s literalist tendencies reflected a developmental period in which children are incapable of accessing metaphoric meaning until they reach preadolescence (so called metaphor deficit or literal stage accounts; see Vosniadou, 1987 and Winner, 1988/1997 for reviews). Researchers with this view claimed children moved from an initial creative and flexible stage—observed most notably in studies of children’s early metaphor productions (e.g., Billow, 1981; Winner et al., 1980)—toward a strictly literal stage that persists throughout their preschool years (Gardner et al., 1975; Levorato & Cacciari, 2002). This idea of a fixed stage has since been criticized, as children have shown some capacity for figurative language within this supposed literal period (e.g., Di Paola et al., 2020; Gentner, 1977; Gottfried, 1997; Pouscoulous & Tomasello, 2020; Özçalişkan, 2005). However, while literal stage accounts have fallen out of favor, it is still commonly assumed that literalist tendencies reflect a difficulty with figurative language comprehension (e.g., Cacciari & Padovani, 2012; Long et al., 2021; Vosniadou et al., 1984; Winner, 1988/1997; Winner et al., 1976).

In light of other cases where children's competencies had previously been underdemonstrated in experimental settings (e.g., scalar implicature; Horowitz et al., 2018; Skordos & Papafragou, 2016), more recent research has begun to explore whether children's past difficulty with metaphor could be explained by task difficulty instead of a deficit in metaphor comprehension (Colston, 2020; Di Paola et al., 2020; Pouscoulous, 2011; Vosniadou et al., 1984). For example, previous tasks often relied on a sophisticated verbal reasoning ability where children had to explain what the experimenters really meant in context to demonstrate metaphor comprehension (for reviews, see Pouscoulous, 2011 and Vosniadou, 1987). The ability to comprehend is distinct from the ability to succinctly verbalize what is comprehended, and even adults struggle with this discrepancy in experimental settings (Faitaki & Murphy, 2019). Now, instead of using complicated verbal reasoning tasks, recent metaphor comprehension tasks typically use forced-choice paradigms that require children to disambiguate between literal and metaphorical interpretations (e.g., Di Paola et al., 2020; Long et al., 2021; Van Herwegen et al., 2013). For example, in a study with 5-year-olds, researchers asked children to pair one of three images with the statement "Lucy is a parrot" from either a literal depiction of a parrot, a metaphorical depiction of a girl resembling a parrot, or a distractor girl (Long et al., 2021). The use of these forced-choice style tasks often corresponds with findings showing an earlier onset of metaphor comprehension (e.g., Almohammadi et al., 2025; Pouscoulous & Tomasello, 2020). However, a preference for literal interpretations still appears in many studies on young children's metaphor comprehension (Vicente & Falkum, 2021; Winner, 1988/1997). For example, in Long et al. (2021) younger children chose literally ~80% of the time (see also Martín-González et al., 2025 for similar results among older children).

Because these newer forced-choice tasks are assumed to better scaffold children's developing abilities, any error children make on them tends to be taken as even stronger evidence against robust metaphor comprehension. In fact, presenting children with both literal and metaphorical interpretations in these forced-choice scenarios stems from the assumption that proving children's ability to comprehend metaphorical meaning requires demonstrating their capacity to ignore literal alternatives (Winner, 1988/1997). In these cases, using literal competitors is meant to act as a more rigorous test of comprehension because it shows that participants can ignore a highly competing alternative. However, we argue that while pitting literal competitors against metaphorical ones is a common practice, it makes the task about why children prioritize one interpretation over another rather than a test of sensitivity to metaphorical meaning alone.

However, not only has this (literalist)performance–(metaphor)competence relationship never been tested but inferring children's metaphor comprehension from their literal responses assumes literalist responding is inherently incorrect. These interpretations are reasonable if we are to assume that the literal interpretation is just a distractor, or that the metaphorical interpretations are the only correct choices. However, because both literal and metaphorical interpretations are made explicitly available in these tasks (usually via visual evidence), the 'correct' answer is in fact, ambiguous. This ambiguity makes it difficult to determine whether children's literalism is due to an inability to reason metaphorically or whether children perhaps privilege literal interpretations when they are available.

Long and colleagues (2021) also commented on the ambiguity of these literal versus metaphor disambiguation tasks (see also Gardner & Winner, 1978). In their comparison study with 13-year-olds (an age at which previous research has asserted children should be sensitive to metaphorical meaning, Vosniadou, 1987; Willinger et al., 2017; Winner, 1988/1997), they found that older children performed at chance and were not strictly metaphor-biased. They concluded that older children's sensitivity to both literal and metaphorical meanings likely made them sensitive to the ambiguity of the task as well—compared to the 5-year-olds whose literally biased performance suggested sensitivity to the literal meaning only. It could be that younger children's more apparent literalist preferences reflect a lack of sensitivity to metaphorical meaning; however, the fact that there *was* ambiguity makes this difficult to determine.

Despite the literature moving away from literal stage (or metaphor deficit) wording, children's errors—most of which are literal responses—are still presumed to reflect a difficulty in understanding metaphors. Regardless of whether this link between literalist performance and difficulty with metaphor exists, if the aim of metaphor comprehension tasks is to see if children can derive metaphorical interpretations, then including a highly competing literal interpretation complicates that goal. Decoupling these literal and metaphorical choices may therefore offer a more sensitive measure of early metaphor comprehension.

In a recent study, Pouscoulous and Tomasello (2020) replaced literal alternatives with distractors and indeed found that 3-year-olds could comprehend novel, perceptually based metaphors. For example, children were given two toy cars—one with a large sack on its roof (metaphorical) and one with a similar sack inside (distractor)—and asked to "Pick the car with the backpack." Children chose metaphorical depictions over distractors, demonstrating that even 3-year-olds can access metaphorical meaning in a minimal linguistic context.

Given the participants' comparatively young age in this study, their success in the task may also support a possible distinction between children's literalist tendencies and their ability to understand metaphor. However, because Pouscoulous and Tomasello (2020) did not provide literal options, it is unclear whether children's success should be attributed to the type of metaphor tested or the absence of literal interpretation options. It has been suggested that visual or attributional metaphors, like those used in Pouscoulous and Tomasello, are more accessible and easier to comprehend (Gentner, 1977; Winner et al., 1976), so it could have been that children's success was limited to those specific items. Additionally, because children appear to understand these more straightforward attributional metaphors, they may have been less likely to select literal alternatives had such options been provided. Since the task did not include literal options, it remains unclear what underpins children's success. Consequently, it cannot yet be determined whether literal responses reflect a deficient metaphor comprehension.

In the current study, we sought to unpack this coupling between literalist responses and metaphor comprehension by replicating and extending Pouscoulous and Tomasello (2020) to include more abstract metaphor types and literal competitors. We tested 3- to 7-year-olds to see if they could understand novel metaphors in the absence of literal options (Experiment 1) and whether the presence of literal options affected performance (Experiment 2). In Experiment 1, children chose which image, between a metaphorical and a distractor, matches a metaphoric statement. In Experiment 2, we tested different children on the same four sets of metaphoric statements and replaced the distractor images with literal depictions. We also ran a third experiment in which we gave children the literal images from Experiment 2 and the distractor images from Experiment 1 to get a sense of the different age groups' understanding of the test items and to compare possible processing differences across the experiments.

If children could make sense of metaphoric statements in the first experiment but performed at chance or responded literally in the second, that would suggest that metaphorical meaning may be accessible to children, but the presence of literal competitors could mask their early reasoning abilities in ambiguous settings. This distinction between performance in Experiments 1 and 2 would also support previous work demonstrating that children can grasp metaphors early on and challenge the idea that literalism results from protracted metaphor comprehension development.

In addition to our original question, we explored how children progress through the tasks using their reaction time. We also looked for any developmental patterns or differences in children's understanding of the different types of metaphors.

Previous research suggests a linear development of metaphor type (Asch & Nerlove, 1960; Gentner, 1977; Winner et al., 1976)—where children understand functional metaphors before psychological metaphors but both later than attributional metaphors[1]—so it could be that children's overall task performance improves with age. However, additional research has found a U-shaped curve in some nonliteral language development, a possibility that might extend to metaphor comprehension as well (Gardner et al., 1975; see also Köder & Falkum, 2020 for an example in children's metonymy acquisition). We extended the items to include these other metaphor types to provide a more robust test of metaphor comprehension. However, any predictions of metaphor type, age, and their interactions were purely exploratory, as neither these metaphor types nor their interactions with age have been tested systematically under this decoupled lens.

Our reaction time predictions were equally speculative, as few reaction time measures existed in metaphor tasks using developmental populations. However, using a reaction-time-as-processing-effort approach (as in Di Paola et al., 2020), we assumed the following. If children in Experiment 2 responded more slowly than those in Experiment 3, it would indicate that they were still sensitive to the competing options in Experiment 2, even if their responses were equally literal. Conversely, if their response times in Experiment 2 were comparable to those in Experiment 3, it would suggest that the children did not perceive or engage with the ambiguity of the experimental context. If children chose metaphorical interpretations above chance in Experiment 1 and took longer to respond compared to children who chose literal interpretations in Experiment 3, this would align with findings suggesting that metaphorical interpretations are more costly than literal interpretations (as in Noveck et al., 2001).

Experiment 1

To investigate children's early metaphor comprehension abilities without the presence of a competing literal interpretation, we first replicated the study by Pouscoulous and Tomasello (2020) and extended it to include novel concrete (i.e., attributional) and abstract (i.e., functional and psychological) metaphors and children from 3 to 7 years old. As in the original study, we expected children to pick metaphorical depictions above chance by eliminating the conflicting literal information. However, if children could not understand the metaphors, we expected them to perform at chance because there should have been no way for them to reconcile the differences between metaphor and distractor images if they were not capable of overcoming literal meaning.

---

[1] Terminology for these metaphors varies (e.g., functional and psychological often are referred to as "abstract" metaphors and attributional as "concrete" or "perceptual"). For the purpose of this study, functional metaphors involve metaphorical relations derived from an object's function, whereas psychological metaphors involve relations to internal states (see Table 1 for a list of items).

In Pouscoulous and Tomasello (2020), 3-year-olds performed nearly at ceiling on attributional metaphor understanding. We expected to replicate this performance on trials adapting their original stimuli. However, as research on metaphor comprehension rarely tests multiple types of metaphors and age groups, we also explored the interaction between metaphor type and age to isolate further whether comprehension difficulties could be specific to different types of metaphors at specific developmental periods. All experimental protocols adhered to The National Committee for Research Ethics in the Social Sciences and the Humanities guidelines and have been approved by the Norwegian Agency for Shared Services in Education and Research (Reference Number: 596365). The study was also preregistered (https://osf.io/vauw2), and all materials, data, and analysis scripts are available on our OSF project page (https://osf.io/jkq9w/). In the next section, we describe ways our study differs from the original and the rationale.

Method

Model Selection and Sampling Plan

Prior to running the study, we ran full and null model simulations to assess the probability of successful model convergence (n = 120; for 1000 simulations) and confirmed model feasibility. For the complex model, including all correlational parameters, the probability of successful model convergence was 0.965; for the simple model excluding those parameters, the probability was 1. For each age group, planned sample size estimates were 20–26 for experimental conditions and 10–15 for the Experiment 3. Minimum estimates were informed by our model feasibility simulation (with total n = 120 for experimental conditions) and maximums by the reported samples in Pouscoulous and Tomasello (2020), referencing both final ($N_{\text{experiment1-2}} = 24$; $N_{\text{experiment3}} = 12$) and excluded ($N = 5$) samples. Supplementary analysis information for this project appears on our OSF page (https://osf.io/jkq9w/).

We chose to run our studies between subjects because we were interested in whether children could generally access metaphorical meaning and if they preferred literal meaning when it was available. Additionally, because we propose that providing literal alternatives may be problematic for testing metaphor understanding, we are less interested in exploring children's individual literal biases in the context of these paradigms. However, to mitigate concerns regarding potential individual variances across samples, we tested experiments across each testing session. We also assessed children's inhibitory control abilities using the DayNight task (Gerstadt et al., 1994) and collected language and demographic information from guardians, adapting the protocol from The ManyBabies Consortium (2020) to create a profile for each group of participants. Profiles were largely similar across samples and appear in the Supplementary Information section of this paper.

Participants and Design

We tested 82 3- to 7-year-old Norwegian-speaking, typically developing children (28 3-year-olds, 27 5-year-olds, and 27 7-year-olds). Of those tested, three children withdrew (1 per age group), and three were excluded (all 3-year-olds: one for a technical error, one for sibling interference, and one for practice failure). Children were given the same 20 test items in a fully randomized order, with metaphor and picture location on the right or left counterbalanced across participants.

Materials

In the original study, Pouscoulous and Tomasello (2020) always constructed the metaphors from perceptual similarities relating to the human body. Previous research (e.g., Gentner, 1977) has shown that children tend to perform better with body-related metaphors, supposedly because visual comparisons and spatial relations to the human body (e.g., "head = top") are more intuitive and easier to grasp. To avoid having the items all share a specific relation, we broadened the metaphoric devices to include functional and psychological metaphors as well as created new attributional metaphors without this body relation.

Additionally, following concerns regarding contrastive inference confounds discussed in Pouscoulous and Tomasello (2020, pp. 164–166), we adapted two of the original items. For "The dog with the brown shoes," we changed the distractor image to a dog with brown ears instead of the original dog with a brown bow so that both referents contained multiple brown features. For "The bottle with the big belly," we omitted the word "big" to avoid children choosing pictures from size (i.e., adjective) cues alone.

We made our metaphors by first compiling metaphor stimuli used in previous research testing similar age groups. We then either refined ones taken from that list or devised additional novel metaphors, taking inspiration from literature on children's understanding of object functions and mental state language (e.g., Bloom, 2001; Callanan et al., 2007; Deák, 2006; Estes et al., 1989; Harris et al., 2005). We also confirmed that all vocabulary used to construct the test referents were generally acquired by typically developing, Norwegian-speaking 3-year-olds. We did this by referencing the Norwegian vocabulary database, Ordforrådet (Lind et al., 2013; Simonsen et al., 2014), and the Stanford Wordbank (Frank et al., 2016). Additionally, we ran a control experiment in which we tested children on literal versus distractor images to ensure that children of similar ages could understand the test utterances without metaphorical constructions. See Experiment 3 later in this paper for this methodology and results.

Once we had a list of workable statements, we adapted them to the format used by Pouscoulous and Tomasello (2020) such that metaphors described a property of the subject of the utterance and were embedded in a referential statement (e.g., *Vis meg tårnet med hatten*/Show me the tower with the hat). After this process, we chose the final 20 metaphors based on how easily they could be depicted. Ten of these 20 were concrete (five from Pouscoulous and Tomasello and five new attributional items), and 10 were abstract (five novel functional and five novel psychological items).

For the images, we replicated the visual depictions from the Pouscoulous and Tomasello (2020) study near exactly (they are the cartoon equivalents of the physical toys used in the original study). All other images followed similar protocols to these items, where each image had to contain the referent and match on overall visual salience. We modeled the literal images on the distractor images and constructed the distractor images by mirroring the visual characteristics of the metaphorical images in a nonmetaphorical way. For example, in the phrase "The tree with the arms," the metaphorical tree had two branches that extended from the middle of the tree to look like arms. In contrast, the distractor tree had two roots that extended equally from the bottom of the tree, so they did not in any way relate to the metaphoric referent "arms." Table 1 presents the final 20 metaphoric statements and their English translations. Visual depictions appear on OSF (Løvstakken & Neff, 2024) though, see Figure 1 for an example.

Lastly, we ran a preference assessment test to account for potential preferences for one image over the other across sets. Similar to Pouscoulous and Tomasello (2020), we presented 3- to 7-year-old children either with metaphor and distractor items (n = 35), metaphor and literal items (n = 36), or literal and distractor items (n = 14), and we asked them to *Vis meg en* (Show me one). Children showed no preferences for either metaphor or literal over distractor images (p > .05). However, individual binomial tests confirmed significant preferences for a subset of images in the metaphor versus literal comparison group (of which, most were preferences for the metaphorical items over the literal ones). We made some adjustments following this finding, and preferences were removed (n = 20, p > .05). Raw data for these assessments, alongside original images, appears on OSF (https://osf.io/jkq9w/).

Table 1. ***Test sentences and their English translations***

Metaphor Type	Statement in Norwegian	Statement in English
Replicated attributional	Tårnet med hatten	The tower with the hat [a]
	Bilen med ryggsekken	The car with the backpack [a]
	Gulroten med håret	The carrot with the hair [a]
	Hunden med den brune sko	The dog with the brown shoes [a, b]
	Flasken med magen	The bottle with the belly [a, b]
New attributional	Treet med armene	The tree with the arms
	Hodet med spagettien	The head with the spaghetti
	Himmelen med kjeksen	The sky with the cookie
	Kakaoen med putene	The hot cocoa with the pillows
	Treet med de stekte eggene	The tree with the fried eggs
Functional	Frosken med paraplyen	The frog with the umbrella
	Reven med lommelykten	The fox with the flashlight
	Apen med hammeren	The monkey with the hammer
	Ekornet med koppen	The squirrel with the cup
	Larven på flyet	The bug on the plane
Psychological	Planten som er trist	The plant that is sad
	Ballene som er glade	The balls that are happy
	Tegningen som er sint	The drawing that is angry
	Baggen som sover	The bag that is sleeping
	Gutten som brenner	The boy that is on fire

[a] Items are an original subset from Pouscoulous and Tomasello (2020).
[b] Items changed to circumvent contrastive inferences.

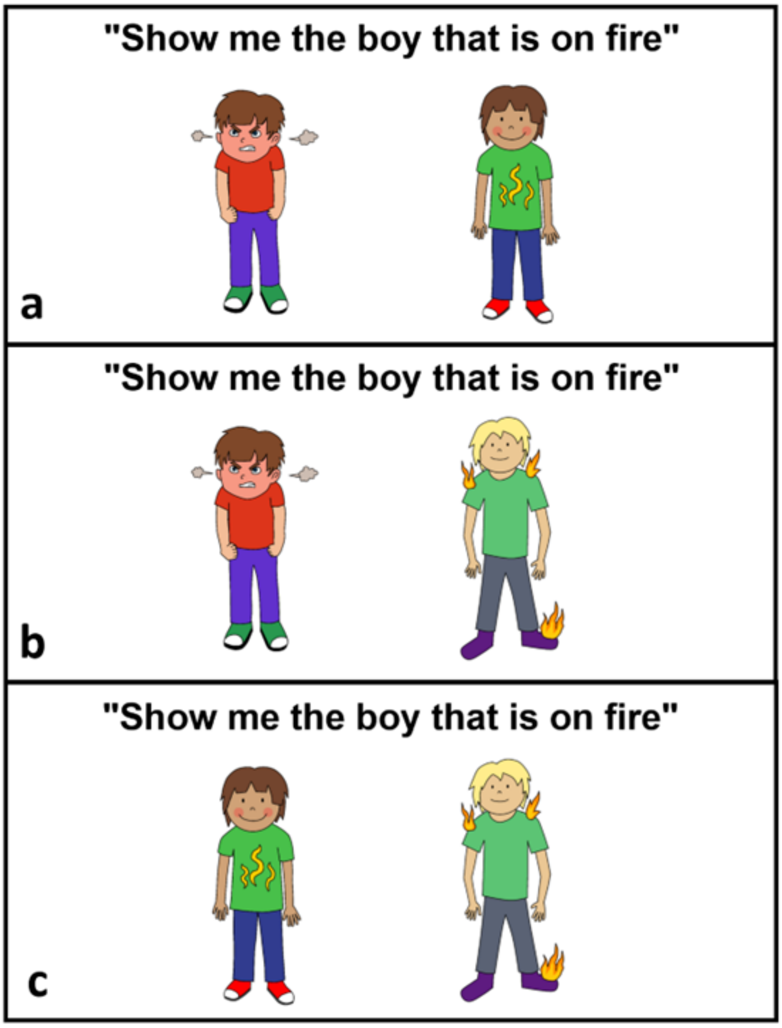


Figure 1. ***In child point of view: Schematic of test trials for Experiments 1 (a), 2 (b), and 3 (c).***

Note. All statements were aurally presented in Norwegian, and no text bubbles were used.

Procedure

Testing was done on tablets using the program PsychoPy/Pavlovia (Bridges et al., 2020; Peirce et al., 2019). The testing occurred in museums, kindergarten classrooms, and schools around Oslo. Testing always took place in a separate room to avoid other children and distractions. We only invited children whose guardians returned the signed consent form to participate. In addition to the guardians' informed consent, we made participating children aware of the voluntary nature of the study and all children had to assent to participate. Children also received a small gift (e.g., an eraser or a diploma) for their participation.

To begin the task, an experimenter introduced the children to two characters on a tablet and explained that they would be playing a game where the characters would ask questions about different pictures on the screen. In the original Pouscoulous and Tomasello (2020) study, children interacted with two experimenters and real objects. We made these changes to familiarize the children with one character and play the game with another. We chose characters because it reduced the need to control for social information and allowed us to remove the other character from the environment when they were not speaking. This removal also reduced the amount of communicative style switching in the task (i.e., speaking literally and figuratively). Children interacted with both characters; however, character identity was fixed in the practice and test sessions. The experimenter was a trained research assistant and a native Norwegian speaker.

Practice Trials. To ensure children understood the game, the experimenter told them they would play a practice game with one of the characters first. After introducing the child to the two characters on the screen, one character left, and the other remained to take the child through four practice trials. To be included in the final experiment, children had to have answered three out of four practice trials correctly, as was the case in Pouscoulous and Tomasello (2020).

The practice trials were identical to the test trials; except they did not include metaphorical statements. For example, the practice character asked the child to “Pick the apple that is red” while presenting an image of a red apple and a green apple. We took two of the practice trials directly from Pouscoulous and Tomasello (2020), and we altered the remaining two to familiarize children with more pragmatic choices. In piloting, we noticed that using only literal practice trials trained children to expect a specific referential match (i.e., to look for a “correct” and literal response). In our experiments, the literalness varied, making the correctness less defined, so we created two ‘pragmatic’ practice trials to discourage this type of response heuristic. Pragmatic trials included referents that were not entirely satisfactory or prototypical, but one was always more appropriate than the other. For example, we asked children to pick “the line that is straight” and presented them with one nearly but not perfectly straight line and one squiggly line (Ronderos et al., 2022). Like in the original Pouscoulous and Tomasello trials, children got feedback on their responses.

Test Trials. After practice, the experimenter told the children they had finished practicing and would now begin to play the actual game with the other character from the introductions. In the test, the character from the familiarization left, and the other character from the introduction returned and told the child that they would also like to play. As in practice, two images (now a metaphor and a distractor image) for each statement appeared on the screen, and the character asked the child to select one of the images that aligned with the instruction: "Show me the" plus a metaphoric statement. For example, the character said, "Vis meg tårnet med hatten" (Show me the tower with the hat) while an image of a tower with a red balcony (i.e., distractor image) and an image of another tower with a large red roof (i.e., metaphoric depiction) appeared. After the child selected an image, the subsequent trial immediately began. All test statements were metaphorical, and no feedback was provided.

The experiment session was audio recorded, and children's picture selections and reaction times were recorded directly via the tablet. Distractor image location, which we noted as left or right according to the child's point of view, was fixed in practice (i.e., ABBA) and counterbalanced in the test, such that half the metaphoric images appeared on the children's right and half on the left. Statement order was also fixed in practice but fully randomized in the test. See Figure 1 for a schematic of the procedure.

Results

Model Selection

Children's picture selections were analyzed using a generalized linear mixed model in R (lme4 package, Bates et al., 2015; R Core Team, 2021) using a binomial distribution with metaphoric responses coded as 1 and the distractor coded as 0. The full model included age group (z transformed), metaphor type, and their interactions as fixed factors; items and subject identity as random effects (Clark, 1973); and random slopes of age and metaphor type. The null model was identical but with metaphor type and its interaction with age removed. We made inferences regarding effects on picture selections via full-null model comparisons using an analysis of variance ($p \leq 0.05$) via the lmerTest R package (Kuznetsova et al., 2017).

In our preregistration (https://osf.io/vauw2) we outlined a confirmatory model that only included the effect of experiment type on picture selection, as well as an additional exploratory model that looked at potential interaction effects of age and metaphor type. We made these choices because, for our comparison across experiments, the model isolating experiment was more theoretically coherent. However, in isolating the effects of this first study, we decided to progress with the exploratory model as the maximal structure that included effects of metaphor type, age, and their interactions more appropriately fit our experimental design. See Barr et al. (2013) and Schielzeth and Forstmeier (2009) for further discussion regarding maximal slope structures. There were two additional changes from preregistration, namely that location ID and trial ID were removed from random effects structures as testing location was balanced and randomization was set to full random, so both were collapsed within subject identity.

Model Results

The full model (AIC = 871.99) did not provide a significantly better fit to the data compared to the null model: AIC = 873.01, $\chi^2(2) = 2.98$, $p = .225$, suggesting that age alone had a significant effect on the proportion of metaphorical picture selections but not metaphor type or their interactions. From visual inspection of the data (see Figure 2), we surmised that all children selected metaphorical pictures at above-chance levels (chance = .50) but that 3-year-old children chose slightly fewer metaphorical pictures compared to 5- and 7-year-olds. However, for a more detailed overview of metaphor type and individual item performance, see the Supplementary Information section for a graph depicting all four metaphor types alongside tables of mean picture selections for each item by age group.

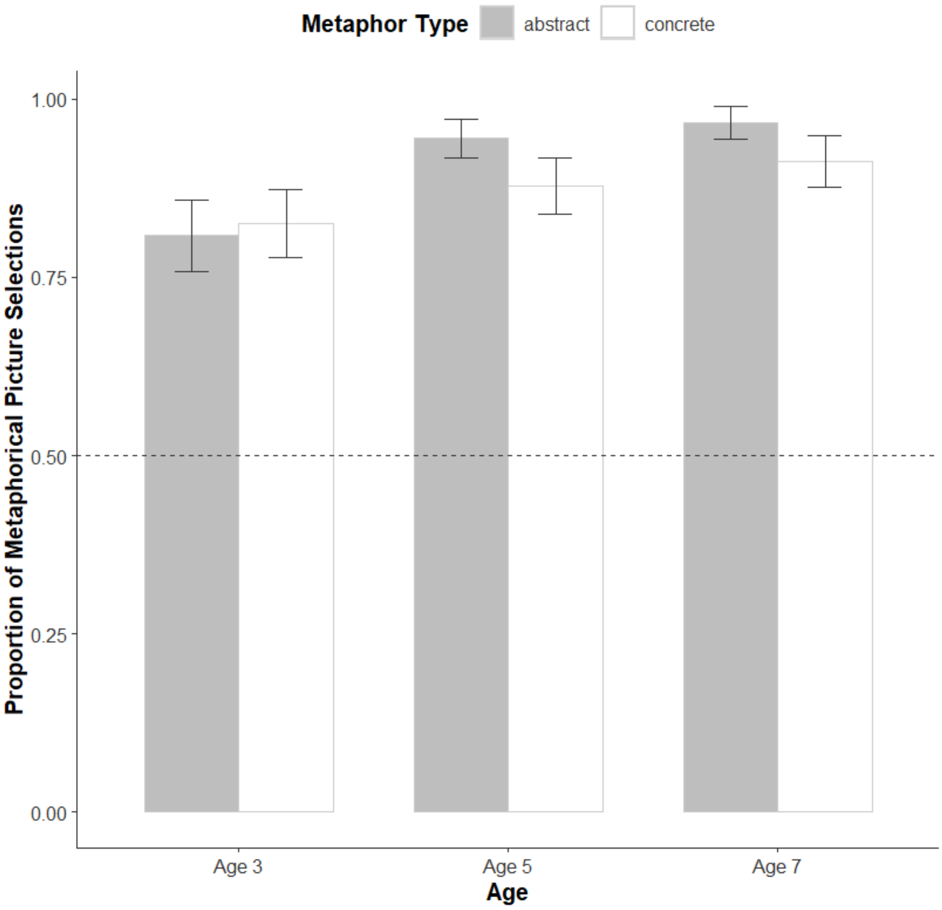


Figure 2. ***Proportion of metaphor selections in Experiment 1 by age and metaphor type.***

Experiment 2

Experiment 2 followed the exact procedure as Experiment 1—that is, we kept the metaphorical statement and depictions constant. However, in the test, we replaced the distractor images with literal depictions. We predicted reduced metaphoric selections in this second experiment because the literal interpretation was now available. For the predicted comparisons between experiments, if children's literalism reflected an early metaphor comprehension deficit, we would expect literal responding in Experiment 2, chance responding in Experiment 1, and for this overall pattern to be more apparent in our younger age groups. However, if we encountered literalist responding in Experiment 2 but above-chance metaphor selection in Experiment 1, this would contradict a direct mapping between literalism and protracted metaphor comprehension.

Method

Participants and Design

We tested 82 3- to 7-year-old Norwegian-speaking, typically developing children (29 3-year-olds, 27 5-year-olds, and 26 7-year-olds). Of those tested, one 3-year-old child withdrew, and three were excluded (two 3-year-olds and one 5-year-old for practice failures). Aside from replacing the distractor images with literal ones, we gave children the exact same 20 test items as Experiment 1 in a fully randomized order with metaphor picture location (i.e., right vs. left) counterbalanced across participants.

Results

Our model followed the same parameters as the first experiment, but to investigate the effect of experiment type on picture selection, we additionally included fixed effects of experiment type (including its interaction with metaphor type and age), as well as its random slope on metaphor type. The null model did not include the effect of experiment type.

The full model (AIC = 1,468.5) provided a significantly better fit to the data compared to the null model: AIC = 1,560.5, $\chi^2(4) = 100.04$, $p < 0.001$, meaning that including the effect of experiment type explained children's picture selections better than age alone. Visual inspection of the data (see Figure 3) showed that not only did children select fewer metaphorical images in Experiment 2, but they also selected them at below-chance levels, thus indicating a preference for the literal alternatives.

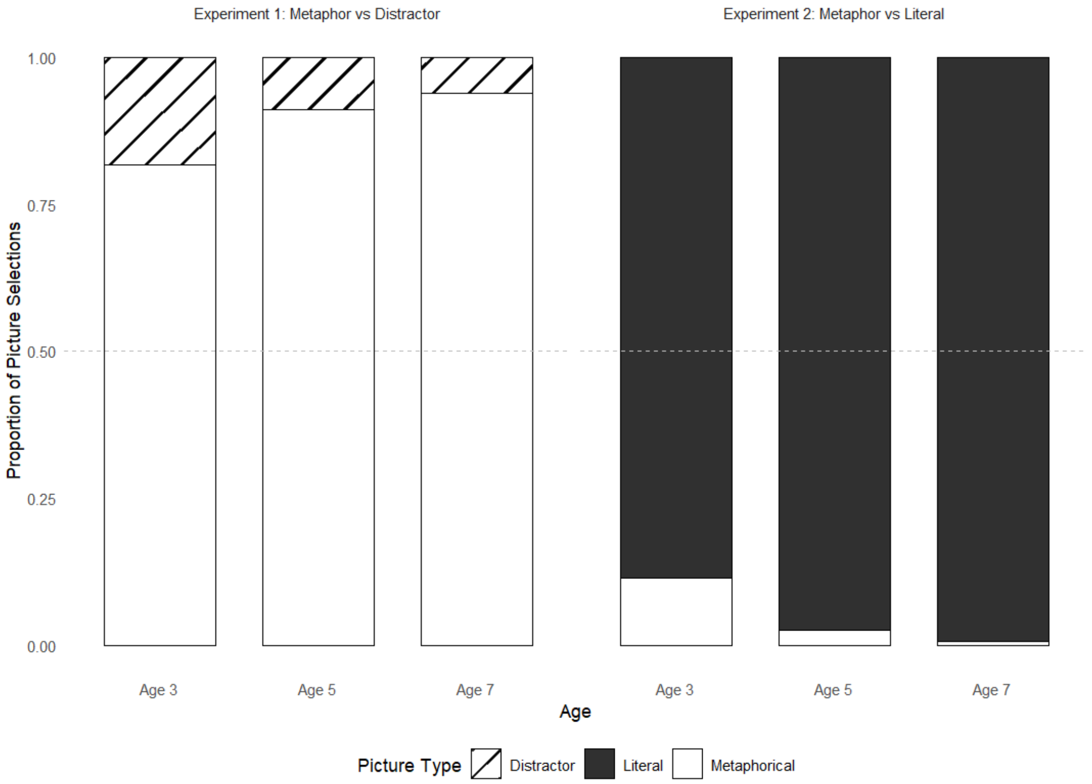


Figure 3. ***Proportion of picture selections in Experiments 1 and 2 with dotted line representing chance at 0.50.***

Experiment 3

To test whether children in these age ranges generally understood the test vocabulary, we ran a control experiment using the same procedure as Experiments 1 and 2. Although we constructed the items from words generally acquired by 30 months (Frank et al., 2016; Lind et al., 2013; Simonsen et al., 2014), we wanted a measure of children's relative understanding of our test vocabulary. We presented children in this study with the same utterances as in Experiments 1 and 2, but we did not give them metaphoric options. Instead, we presented the children with the distractor images used in Experiment 1 and the literal images used in Experiment 2. For example, for “the tree with the arms,” the experimenter showed children a tree with cartoon arms (i.e., literal) and a different tree with its roots showing (i.e., distractor; see also Figure 1 for an example schematic). We expected children of all ages to choose literal items at above-chance levels, with older children responding at ceiling.

Additionally, in line with the exploratory reaction time data for Experiments 1 and 2, we planned to compare children's response times in Experiment 2 with these in Experiment 3 to determine whether children remained sensitive to the metaphorical interpretation in Experiment 2 despite their literalist responding. If this were the case, we expected to find faster responses in Experiment 3 compared to Experiment 2. Additionally, if children were slower in Experiment 1 compared to Experiment 3, it would suggest that children were pragmatically constructing metaphorical interpretations at an added cost in Experiment 1 compared to the literal interpretations they could access more directly in Experiment 3. Because our reaction time hypotheses were partly motivated by adult comparison research, we present these data in the Adult Experiment section to include all data in a single model and avoid multiple testing (see also Figure 5 for these data).

Method

Participants and Design

We tested 50 3- to 7-year-old Norwegian-speaking, typically developing children (17 3-year-olds, 16 5-year-olds, and 17 7-year-olds). Of those tested, three participants withdrew (one per age group), and we excluded two (one 3-year-old for practice failure and one 7-year-old for language requirements). Procedures followed the other experiments except that we set picture identities in the test to literal versus distractor images.

Results

Model selection followed previous experiments; however, we removed fixed effects of metaphor type and experiment. The confirmatory model, including fixed effects of age, explained the variance in the data better in the full model (AIC = 137.0) than in the null model. The null model just included a random intercept: AIC = 148.3, $\chi^2(3) = 17.296$, $p < 0.001$. Visual inspection of the data showed all participants selected literal pictures at above-chance levels; however, 5- and 7-year-olds selected target (i.e., literal) pictures slightly more often than 3-year-olds, but this difference was not significant ($p = 0.138$). See Figure 4 for all picture selection results across ages.

Adult Experiment

Children's metaphor comprehension is often assessed in terms of their ability to overcome a literal interpretation. We assert that pitting literal and metaphorical referents against one another adds additional challenges to these tasks and ultimately makes the goal of these tasks more ambiguous. Therefore, we sought to disentangle these response patterns to determine whether literalist response patterns are best explained by an early difficulty with metaphor.

We found in Experiment 1 that when there were no competing literal alternatives, children were able to select metaphorical interpretations at well above-chance performance. Additionally, despite the general capacity for metaphorical reasoning at these age ranges, we found that children chose literally when given the option in Experiment 2. The above-chance performance on metaphor comprehension in Experiment 1 and the literally biased responses in Experiment 2 speak to a distinction between the two phenomena (i.e., between literalism and early metaphor comprehension abilities). The fact that the differences in successful metaphorical selections were relative to the presence or absence of a literal competitor, rather than an effect of age or metaphor type, also supports our previously asserted suspicions regarding the potential impact literal items have on performance. However, given the general skepticism surrounding young children's capacity for figurative language, we set out to replicate these findings in an adult sample.

Additionally, something that has motivated this research from the beginning is that even from our own perspective, we could not discern why literal competitors would be inherently incorrect in these settings where they had been made explicitly available. Upon hearing the statement "My toe is on fire," one may not immediately expect the literal interpretation, but why would you not endorse that interpretation once provided with visual evidence that it is indeed afforded in that communicative context (see also Winner et al., 1980 for related discussion). Long et al. (2021) argued that the reduced literalist performance among 13-year-olds indicated a sensitivity to both metaphorical and literal meanings and therefore the ambiguity of the test interaction, whereas the literalist responding among 5-year-olds showed a lacking sensitivity. However, given the absence of context in these settings, it seems no less rational for even adults to choose literal interpretations when this is likely to be the least effortful way to resolve the ambiguity, and also one that enables reference assignment (Wilson & Sperber, 2006).

Given the implications of an early metaphor comprehension deficit derived from children's literalist responding in these settings, we wanted to see how adults perform in these tasks where literal and metaphorical meanings are pitted against one another. Assuming an adult's sensitivity to metaphorical meaning is less fragile than that of a 3-year-old, if even adults choose literal meanings when given the same choice scenarios, then this would reinforce our argument for making a distinction between literalist performance and developing competence with metaphor.

Method

Participants and Design

We tested 67 Norwegian-speaking adults (26 in the following Experiment 1, 26 following Experiment 2, and 15 in Experiment 3). The adult tasks were nearly identical to the children's tasks with minor adaptations (e.g., shortening the intertrial intervals and changing the inhibitory control task to the Flanker task; Eriksen, 1995). We made these modifications so the adults could participate autonomously. All of the experimental materials appear on the OSF website (https://osf.io/8s2fm/). We did not collect demographic profiling information for adults, but their eligibility to participate was contingent on their self-reporting of Norwegian as a main language.

As we did with the children, we ran experiments in parallel, randomly assigning adults to one of the three experimental conditions (i.e., Experiments 1, 2, and 3) with the target picture's side counterbalanced and test item presentation fully randomized across participants. We recruited participants from museums, libraries, and universities in greater Oslo. All participants signed a written consent before participating, were made aware of the voluntary nature of the task and had the opportunity to opt-in to a prize draw for participating.

Results

As we were comparing adult data across experiments, we ran a full-null model comparison in which the full model included age, experiment, and their interaction terms and the null model only included age. Both random effects structures accounted for random slopes of age and experiment on each item and random intercept of subject ID and metaphor type. However, because we included all experiments in the model, we removed metaphor type as a fixed variable as there was no manipulation of metaphor in Experiment 3.

We found that the full model specifying the effect of experiment type (AIC = 1,496.1) better explained the rate of target picture selections (with metaphor coded as the target in Experiments 1 and 2 and literal coded as the target in Experiment 3) than the null model specifying age alone (AIC = 1,625.5, $\chi^2[4] = 137.42$, $p < 0.001$). This finding suggests that 3-year-olds tended to select fewer target pictures compared to older participants and that Experiment 2 negatively affected target picture selection rates for all subjects. Like children, adults similarly chose metaphorical pictures fewer times in Experiment 2 compared to those in Experiment 1. Adults also chose these metaphorical pictures at above-chance levels in Experiment 1 but below-chance levels in Experiment 2, with near-ceiling and floor rates, respectively ($M = 0.97$, $SD = 0.17$; $M = 0.02$, $SD = 0.16$).

However, there were potential multicollinearity issues with the full model as age had a high variance inflation factor value (VIF = 3.97). Though see Fox and Monette (1992), Harrison et al. (2018), and Quinn and Keough (2002) for discussions regarding issues with multicollinearity calculations and thresholds (especially with models specifying interaction effects). We did not consider the correlation an issue for our data (following Allison, 2012), but to see if it affected possible inferences, we reran the models where we dropped the fixed effect of age in the full model because the experiment was always our registered effect of interest. We included a random intercept model for the null and found comparable results (with full model AIC = 1,635.6 and null AIC = 1719.4, $p < 0.001$).

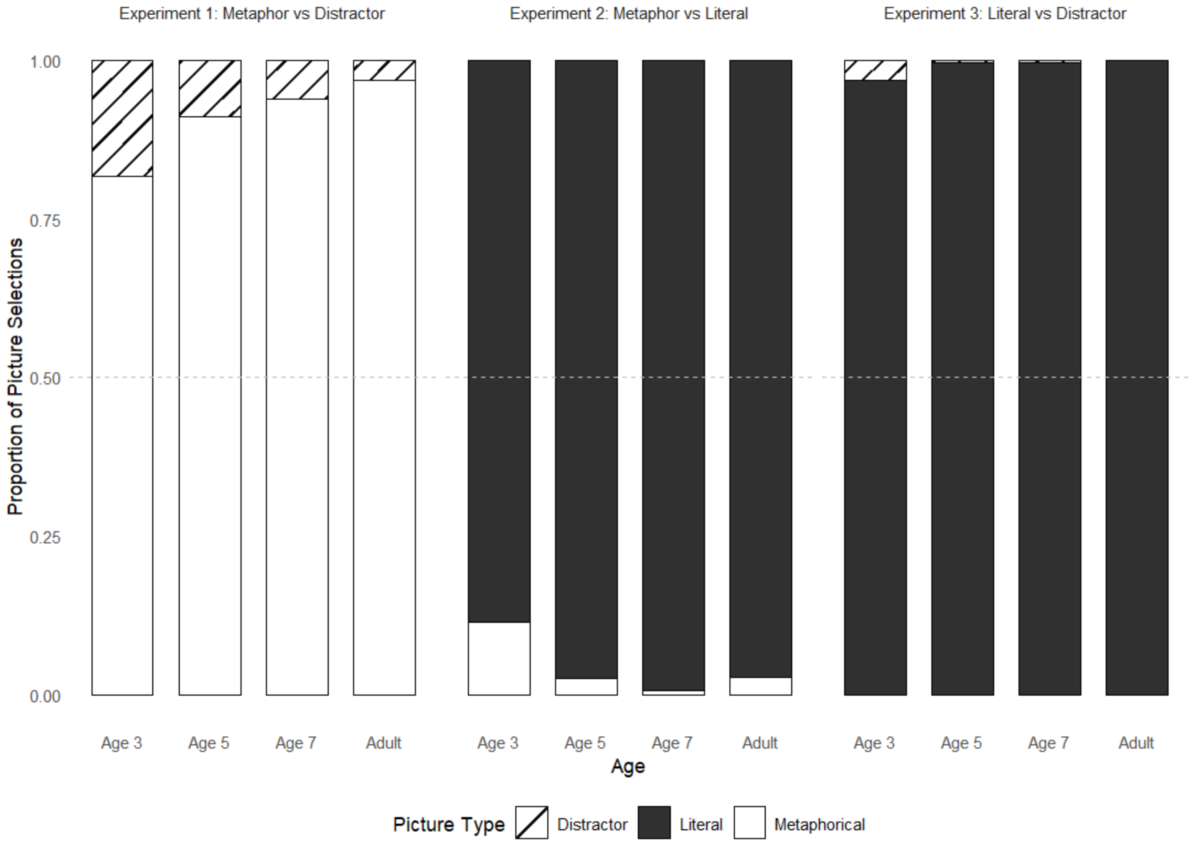


Figure 4. ***Proportion of picture selections by experiment and age with dotted line representing chance at 0.50.***

Exploratory Analysis of Reaction Time

Our primary aim in this paper was to disentangle literal and metaphorical response patterns. In doing so, we have also argued that directly contrasting literal and metaphorical items can obscure children's sensitivity to metaphorical meaning by introducing communicative ambiguity. Prior research has suggested that a shift from literal to metaphorical responses may reflect a developing sensitivity to ambiguity (Long et al., 2021). However, because most studies test metaphor comprehension by requiring children to override literal alternatives, it was difficult to determine at what point literal responses reflect a lack of sensitivity to metaphorical meaning versus interpretation difficulties due to communicative ambiguity. To look into these different factors, we conducted exploratory tests of reaction time to see whether there were any processing differences across literal choices.

We proposed that slower responses in Experiment 2 compared to Experiment 3 could reflect sensitivity to competing interpretations, even when responses were equally literal. In contrast, similar response times would suggest limited engagement with the ambiguity in Experiment 2. Additionally, if children in Experiment 1 selected metaphorical interpretations above chance and took longer than those giving literal responses in Experiment 3, this would support the view that deriving novel metaphorical meanings is more demanding (e.g., Noveck et al., 2001). However, while these analyses were planned from the outset, the specific predictions were speculative as our method departs from conventional approaches in the literature. Accordingly, we interpret any findings here with caution.

Results

We recorded reaction times from the offset of the test utterance to the picture selection and analyzed them using a linear mixed model mirroring the parameters explained in the picture selection analysis in the adult experiment. After fitting the model, we checked for normality by examining the Q-Q plot of residuals (Field, 2005), which we confirmed violated assumptions in a follow-up Shapiro test ($p < 0.001$). We then performed the Box-Cox procedure (Box & Cox, 1964) to find the appropriate transformation for our data ($\lambda = 0.02$) and reran all models using the transformed data which met model assumptions (Lo & Andrews, 2015).

The full model (AIC = 8,495.3) provided a better fit for the data compared to the null model (AIC = 8,756.1), indicating that participants' reaction times varied significantly based on the type of experiment and its interaction with age ($\chi^2[24] = 277.13$, $p < 0.001$). Generally, reaction times decreased with age. Specifically, 3-year-olds had the longest reaction times across experiments with Experiment 1 (M = 4.42 [3.97–4.87]), Experiment 2 (M = 2.64 [2.35–2.94]), and Experiment 3 (M = 1.85 [1.64–2.07]). However, the interaction between age and experiment had mixed effects. Upon visual inspection (see Figure 5), children took longer to respond in Experiment 1 compared to Experiments 2 and 3. In contrast, adults showed similar reaction times in Experiments 1 and 2, both of which were slower compared to Experiment 3. Only 7-year-olds responded with similar reaction times in Experiments 2 and 3, although 5-year-olds appeared to show a similar trend. However, 3- and 5-year-olds and adults seemingly took longer to make selections in Experiment 2 when compared to Experiment 3. We revisit these exploratory results in the General Discussion.

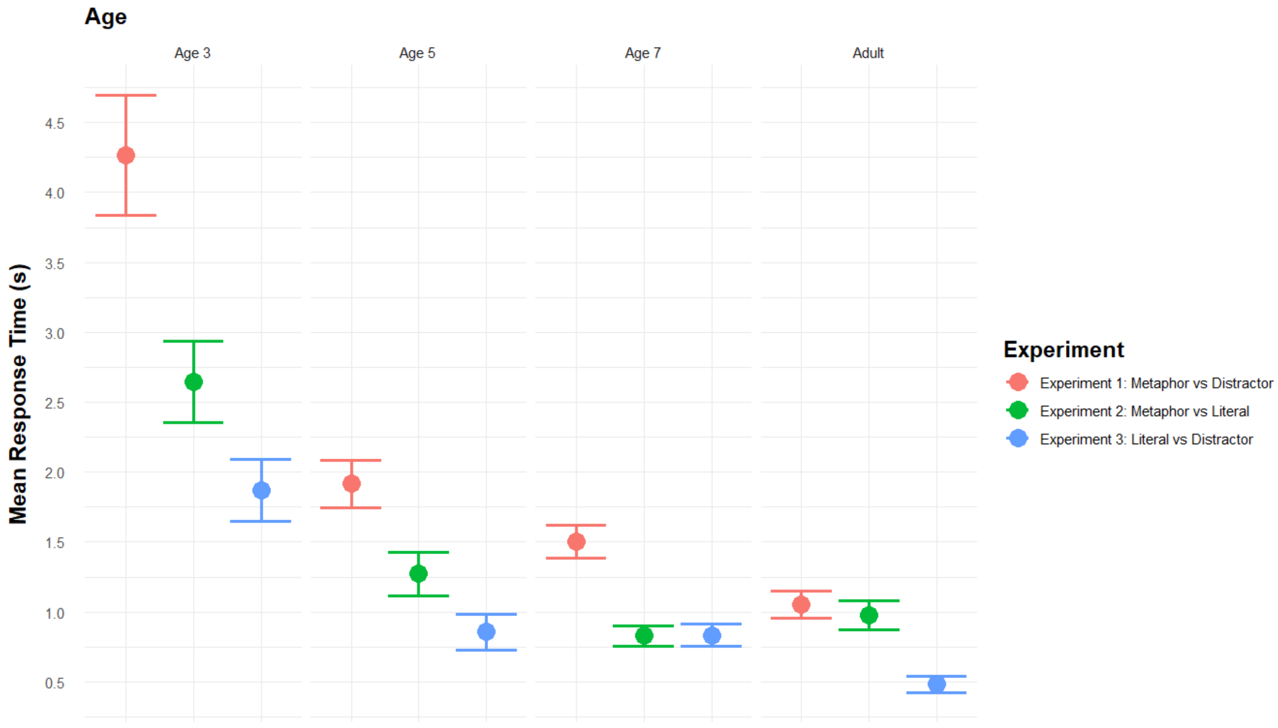


Figure 5. ***Mean response times in seconds observed across different age groups and experiment types with 95% confidence intervals.***

General Discussion

Children's literal interpretations were often assumed to depend on whether they could understand metaphors; however, this assumed link between literalism and early metaphor comprehension had never been tested directly. We investigated whether children's observed literalism in metaphor comprehension tasks is best explained by a developing sensitivity to metaphorical meaning using a between-subjects design. We found no support for claims that young children simply cannot understand metaphors, which is also in line with more recent findings (Almohammadi et al., 2025; Pouscoulous & Tomasello, 2020). Children from 3 years old were adept at our metaphor comprehension task when we removed the picture representing the literal interpretation from their response options. When we included literal options, younger children were not more literally biased than older children—which would follow from a metaphor deficit view, where children would become less literalist and more sensitive to figurative meanings with age. And even adults chose literal options exclusively in this context, where the target utterance was ambiguous between a metaphorical and a literal interpretation. Contrary to previous assumptions, young children's tendency toward literal interpretations does not appear to be explained by an underdeveloped sensitivity to figurative meaning.

Research on children's metaphor comprehension has often interpreted literal responses as evidence of limited sensitivity to metaphorical meaning. If this were the case, 3-year-olds—given their generally weaker performance on metaphor tasks (including in our Experiment 1)—should have shown the strongest literal bias in Experiment 2. Instead, they were no more likely than older children or adults to choose literal interpretations and, if anything, appeared slightly less biased. Adults' own tendency to prefer literal interpretations when both options were available further underscores that literalism cannot be straightforwardly equated with immature metaphor understanding. These findings challenge the assumption that literalist behavior declines linearly with age and call for a reevaluation of past research that equates literal responses with deficient metaphor competence.

One reinterpretation may be that literalist responses reflect children's preference for literal meanings, regardless of their understanding of metaphors, which consequently affects their performance on metaphor comprehension tasks. In tentative support of this claim, the presence of literal competitors, much like age, often predicts children's poor metaphor comprehension task performance (e.g., Van Herwegen et al., 2013; Vosniadou, 1989; Winner, 1988/1997). When given the opportunity in previous research, children chose literal interpretations (Vosniadou, 1987; Winner, 1988/1997), similar to what we found in our second experiment.

It was previously assumed that children chose literal options because they were unable to consider metaphorical alternatives. However, consistent with our findings, earlier studies that did not include literal options showed that children's responses were not random (Pouscoulous & Tomasello, 2020; Vosniadou, 1989; see Winner 1988/1997 for a review), nor did an ability with metaphor reduce literalist selections when a literal option was present (e.g., Long et al., 2021; Vosniadou, 1989). Instead, studies that did not include literal competitors often showed early sensitivity to metaphorical meaning (Pouscoulous & Tomasello, 2020; Vosniadou, 1989), and those that included literal competitors tended to find less metaphor-biased responding, even among older children (e.g., Long et al., 2021; Van Herwegen et al., 2013). Therefore, a misunderstanding of the tasks does not necessarily mark children's protracted metaphor comprehension development: children do not behave wholly randomly or appear confused; rather, they seem to demonstrate preferences for literal interpretations.

Relatedly, there are other factors that have been previously suggested to improve children's performance in metaphor comprehension tasks, which would also arguably help them overcome a literal preference. For example, Deamer (2013) found that children with better inhibitory control performed better on metaphor comprehension tasks. Traditionally, these findings have been argued to suggest that inhibitory control is necessary for children's metaphor comprehension because it allows them to handle conflict between literal and metaphorical meaning (Deamer, 2013; see also Pouscoulous, 2011 for a similar discussion). However, making this argument implies that literal meaning is inherently active, which children must overcome to appreciate metaphorical meaning—a processing account that has been debated (Gibbs & Colston, 2012). Additionally, in light of previous presumptions that young children could *not* access metaphorical meaning, it remains unclear how inhibitory control would aid children's sensitivity to metaphorical meaning because its relative absence would suggest the children would encounter no such conflict. Therefore, in keeping with a decoupled account, our data instead adds to these findings on inhibitory control by supporting the idea that literal meaning is salient (Deamer, 2013). As such, including these literal competitors introduces a nonarbitrary conflict in these tasks of metaphor comprehension and adds additional demands beyond testing sensitivity alone.

Although our findings suggest that literalist selections appear independent of metaphor comprehension, we cannot fully speak to the reasons behind children's early literalist tendencies. The goal of this paper was to disentangle metaphorical and literal responses from one another and not necessarily specify what underpins each individually in light of this decoupling. However, because children could successfully derive metaphorical interpretations in Experiment 1, and even adults chose literal options more often when available in Experiment 2, we assert that protracted metaphor comprehension is an unsatisfactory explanation.

While using literal alternatives remains a widely accepted method for testing metaphor comprehension, using literalism as evidence of a deficient development risks mischaracterizing both literal tendencies and underestimating children's metaphor comprehension abilities. Consequently, we caution against perpetuating the assumed mapping of literalist responding and (reduced) sensitivity to metaphorical meaning and suggest a need for more research into the underpinnings of literalist preferences independent from tests of figurative language comprehension. This is particularly relevant for research using literal competitors, as differences in our metaphorical picture selection and reaction time data across experiments showed that literal competitors add processing demands that researchers should consider when evaluating and making assumptions about early sensitivity to nonliteral meaning.

Methodological Considerations and Limitations

In our study, we replicated and adapted the choice paradigm used by Pouscoulous and Tomasello (2020). Though there have been discussions of potential confounds with this paradigm, such as contrastive inference abilities (e.g., Davies et al., 2021) and visual-associative processes (Pouscoulous & Tomasello 2020, pp. 164–166), the paradigm remains straightforward, has been successfully used in different populations (e.g., Almohammadi et al., 2025; Buehler et al., 2018), and, importantly, did not include literal competitors. However, to address these concerns, we adapted items that we felt could have been resolved via contrastive inferencing and still replicated and extended the original findings. We also broadened the metaphor types used to include more abstract metaphors and collected reaction time data to test the robustness of the task—particularly concerning its test of metaphor comprehension instead of visual associations.

In the original study, children could possibly have resolved the referential assignment via visual associations. For example, the word "hat" in the phrase "The tower with the hat" could have triggered the expectation that there would be something on top. Therefore, like Pouscoulous and Tomasello (2020), children may have selected the metaphorical option in our first experiment because it was the only option with a salient property on top. Although this is possible, it remains unclear how this resolution process would qualify as inherently nonfigurative or purely associative. One could argue that this process mirrors the one described for lexical broadening (Wilson & Carston, 2007), where the concept encoded by "the hat" is broadened to also denote other objects that occupy the same spatial relation of being on top to their bearers. Nonetheless, we also included more abstract metaphors to avoid these potential visual heuristics and found that children performed above chance on all metaphor types. Although this heuristic might be available in more attributional trials, it is unlikely to uniformly apply to our other metaphor types, where this meaning is more abstracted (e.g., as in "the drawing that is angry").

Also, because associative accounts are usually highlighted as a leaner contrast to a richer cognitive account, it remains to be explained why participants took generally longer to progress through the first experiment. Although contested as an account of how communication works, one could argue that the initial search for the literal meaning adds extra time (Grice, 1975; though cf. Sperber & Wilson, 2008). However, the fact that children took longer to respond literally in Experiment 2 compared to Experiment 3 contradicts this account because it suggests that children are not simply searching for literal meaning in the first instance. We also controlled for potential low-level features of the images that could contribute to how children responded. Our preference assessment data showed that children chose randomly between image comparisons (i.e., metaphorical vs. distractor, metaphorical vs. literal, and literal vs. distractor) when instructed with the neutral phrase "Show me one." Only when we made the metaphorical or literal meaning available in the main experiments did they reliably attend to one image over the other. Thus, regardless of broader concerns about what forced-choice tasks reveal about metaphor comprehension, our reaction time data and results across different metaphor types and test prompts suggest that children reliably selected specific referents based on the communicative context, rather than relying solely on visual shortcuts (though see Shanks, 2010, for a discussion on how associative and cognitive accounts can complement each other).

Lastly, our task did not use any disambiguating discourse context in line with previous research (though cf. Keil, 1986 and Gardner et al., 1975 for arguments against testing metaphor sans discourse). Because we did not have any disambiguating context, we cannot speak to the literal "bias" explicitly, as even though participants exclusively chose literally, they did not do so despite more appropriate alternatives (see Tversky & Kahneman, 1974 for discussions operationalizing biased responding). We do not see this as an issue with the current design, as one of our core motivations was to highlight potential issues with traditional forced-choice paradigms. Therefore, we intentionally chose these ambiguous scenarios to cohere with previous research. Additionally, although our findings show that literal competitors add nonarbitrary impact to tasks of metaphor comprehension, we do not claim to have resolved the literal bias full stop. However, considering that children could derive metaphorical interpretations without this prior discourse suggests that the facilitative effect of context observed in previous studies may help children resolve the *choice* between literal and metaphorical interpretations, rather than explain their general ability to derive metaphorical meaning. We therefore suggest that considering participants' responses to literal competitors, even if provided with disambiguating context, that literally biased responding should not be taken simply as evidence of a deficit, or even fragility, with metaphor. Instead, equal attention should be given to why literal meaning might be particularly salient or prioritized by children.

Developmental Trajectories

We replicated Pouscoulous and Tomasello's (2020) finding that 3-year-olds can understand attributional metaphors and extended it to more abstract metaphors such as functional and psychological ones. Previous research found that younger children performed worse on abstract metaphors compared to older children (Asch & Nerlove, 1960; Gentner, 1977; Winner et al., 1976), so when extending our paradigm, we included older children to be able to detect any such developmental differences. Though broadly, children of all ages and adults performed similarly, we did find some evidence to suggest that 3-year-olds less successfully derived metaphorical interpretations compared to 5- and 7-year-olds. However, we also found that 3-year-olds were less successful in the control condition (Experiment 3), so it could be that word knowledge instead of the ability to derive metaphorical interpretations explains these differences (see also Keil, 1986; Norbury, 2005; Vicente & Martín-González, 2021). Whether these small differences in our data resulted from simple word knowledge differences or differences in metaphor processing would require further research. However, for the purpose of our study, the fact that children performed above chance regardless of these potential differences—and that younger children were not more literally biased in the second experiment—highlights children's general ability with novel, nominal metaphors and suggests that literalist responses should not be taken as direct evidence of deficient metaphor comprehension.

Children of all ages and adults similarly passed our metaphorical selection task in Experiment 1 and chose literally in Experiments 2 and 3. However, even though surface accuracy levels appear comparable, the reaction time data indicate different processing strategies. Across all ages, participants responded more slowly in the metaphor versus distractor condition (i.e., Experiment 1) than in the literal versus distractor condition (i.e., Experiment 3). This finding aligns with research suggesting that deriving metaphorical meaning is more effortful (e.g., Noveck et al., 2001). We also found that reaction times overall decreased with age, which could be due to factors other than children's developing competence with metaphor comprehension. For instance, it might be that access and retrieval of conventional senses are generally slower in children than in adults.

To explore whether participants were still sensitive to the metaphorical meaning when literal options were available, we conducted a third experiment to see if children's literal selections were processed faster in this unambiguous context than in Experiment 2. We found that 3- and 5-year-olds, as well as adults, took longer to respond literally in Experiment 2 compared to Experiment 3. The fact that participants were highly literal in both studies but took significantly longer when the context was more ambiguous demonstrates their lingering sensitivity to the competing metaphorical alternatives despite their literalist responses.

However, we do not claim that this sensitivity can be extrapolated to reflect participants' full comprehension of the situation's ambiguity, as we might expect that to result in longer reaction times overall (as seemingly if one were sensitive to the ambiguity of the choice, then making that choice should take longer compared to situations without any ambiguity). Instead of suggesting that younger children's choices are literal because they cannot access metaphorical meaning, some of the developmental differences seen across the literature may result from developing sensitivities to referential ambiguity (see Long et al., 2021 for a similar discussion with their data with 13-year-olds). It could also be that this sensitivity interacts with the ability to overcome the less effortful choice (i.e., literal interpretation). Because we did not provide disambiguating discourse context we cannot speak to a clear distinction between these alternatives, especially in light of adults' literal responses in the ambiguous context. However, we still argue that decoupling these alternatives from children's metaphor comprehension provides a better explanation for the data. Even considering the curious reaction time shift in the 7-year-olds, where they appeared to choose literally in Experiment 2 just as quickly as in Experiment 3: It seems more likely that these older children were more influenced by the presence of a literal alternative than a loss of sensitivity to metaphorical meaning (see Lee et al., 2022 for similar findings with surface-meaning biases in 8-year-olds and Köder & Falkum, 2020 for accounts on related U-shaped developments in nonliteral language contexts).

Conclusion

Our study challenges the common assumption that literal responses represent fragile metaphor comprehension. Contrary to deficit-based expectations, young children successfully derived metaphorical interpretations, and we found no age-related decline in literalist responses within ambiguous contexts where both metaphorical and literal options were available. We cannot argue whether these findings generalize to all metaphor comprehension because we did not probe children's understanding of metaphor forms. However, given that even adults chose mostly literally when provided literal options, it is clear that literal options introduce a meaningful conflict that would likely persist across tasks. Consequently, previous research, which often included literal competitors, may have underestimated children's early metaphor comprehension abilities.

Ultimately, our findings call for a decoupling of literalist responses from metaphor comprehension ability and for researchers to more clearly distinguish between studies assessing sensitivity to metaphorical meaning from those investigating when metaphorical interpretations "ought" to be prioritized over literal ones. Further, we urge researchers who use literal options to interpret any literalist responding cautiously rather than treating it as evidence of figurative language difficulties, especially in light of increasing evidence of children's early sensitivity to metaphorical meaning and harmful implications resulting from deficit-worded accounts.

References

Allison, P. (2012, September 10). When can you safely ignore multicollinearity? Statistical Horizons. https://statisticalhorizons.com/multicollinearity/

Almohammadi, A., Gaskins, D. K., & Rundblad, G. (2025). Metaphor comprehension in the acquisition of Arabic. *Journal of Child Language, 52*(2), 334-354. https://doi.org/10.1017/S0305000923000740

Asch, S. E., & Nerlove, H. (1960). The development of double function terms in children. In B. Kaplan & S. Wapner (Eds.), *Perspectives in Psychological Theory: Essays in Honor of Heinz Werner* (pp. 47-60). International Universities Press.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4,* 328. https://doi.org/10.3389/fpsyg.2013.00328

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

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

Bloom, P. (2001). Precis of how children learn the meanings of words. *Behavioral and Brain Sciences, 24*(6), 1095-1103. https://doi.org/10.1017/S0140525X01000139

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B, 26*(2), 211-243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing megastudy: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, e9414. https://doi.org/10.7717/peerj.9414

Bühler, D., Perovic, A., & Pouscoulous, N. (2018). Comprehension of novel metaphor in young children with Developmental Language Disorder. *Autism & Developmental Language Impairments, 3,* 2396941518817229. https://doi.org/10.1177/2396941518817229

Cacciari, C., & Padovani, R. (2012). The development of figurative language. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 505–522). Cambridge University Press.

Callanan, M. A., Siegel, D. R., & Luce, M. R. (2007). Conventionality in family conversations about everyday objects. *New Directions for Child and Adolescent Development, 2007*(115), 83-97. https://doi.org/10.1002/cd.184

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 135-142. https://doi.org/10.1016/S0022-5371(73)80014-3

Colston, H. L. (2020). Figurative language development/acquisition research: Status and ways forward. *Journal of Pragmatics, 156,* 176-190. https://doi.org/10.1016/j.pragma.2019.07.002

Davies, C., Lingwood, J., Ivanova, B., & Arunachalam, S. (2021). Three-year-olds' comprehension of contrastive and descriptive adjectives: Evidence for contrastive inference. *Cognition, 212,* 104707. https://doi.org/10.1016/j.cognition.2021.104707

Deák, G. O. (2006). Representing object functions: The cognitive basis of tool-use by children. In *Proceedings of the Fifth International Conference on Development and Learning, Bloomington, Indiana, USA.* http://quote.ucsd.edu/cogdevlab/files/2013/02/Deak_function_ICDL06.pdf

Deamer, F. M. (2013). *An investigation into the processes and mechanisms underlying the comprehension of metaphor and hyperbole* [Doctoral dissertation, University College London]. UCL Discover. https://discovery.ucl.ac.uk/id/eprint/1420534/

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

Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition, 2*(2-3), 101-118.

https://doi.org/10.1080/13506289508401726

Estes, D., Wellman, H. M., & Woolley, J. D. (1989). Children's understanding of mental phenomena. *Advances in Child Development and Behavior, 22*, 41-87. https://doi.org/10.1016/S0065-2407(08)60412-7

Faitaki, F., & Murphy, V. (2019). Oral language elicitation tasks in applied linguistics research. *The Routledge Handbook of Research Methods and Applied Linguistics (1st ed.).* https://doi.org/10.4324/9780367824471-31

Field, A. (2005). *Discovering statistics using SPSS (2nd ed.)*. Sage Publications, Inc.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association, 87*(417), 178-183. https://doi.org/10.1080/01621459.1992.10475190

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project.* MIT Press. https://doi.org/10.7551/mitpress/11577.001.0001

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, 4*(1), 123-141. https://doi.org/10.1086/447976

Gentner, D. (1977). If a tree had a knee, where would it be? Children's performance on simple spatial metaphors. *Papers and Reports on Child Language Development, 13,* 157-164.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition, 53*(2), 129-153. https://doi.org/10.1016/0010-0277(94)90068-X

Gibbs, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning.* Cambridge University Press. https://doi.org/10.1017/CBO9781139168779

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

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (pp. 41-58). Brill.

Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science, 14*(2), 69-73. https://doi.org/10.1111/j.0963-7214.2005.00337.x

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ 6*:e4794. https://doi.org/10.7717/peerj.4794

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development, 89*(5), e572-e593. https://doi.org/10.1111/cdev.13014

Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development, 1*(1), 73-96. https://doi.org/10.1016/S0885-2014(86)80024-7

Köder, F., & Falkum, I. (2020). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics, 156*, 1-13. https://doi.org/10.1016/j.pragma.2019.07.007

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

Lee, R., Chambers, C. G., Huettig, F., & Ganea, P. A. (2022). Children's and adults' use of fictional discourse and semantic knowledge for prediction in language processing. *PLOS One, 17*(4), e0267297. https://doi.org/10.1371/journal.pone.0267297

Levorato, M. C., & Cacciari, C. (2002). The creation of new figurative expressions: Psycholinguistic evidence in Italian children, adolescents and adults. *Journal of Child Language, 29*(1), 127-150. https://doi.org/10.1017/S0305000901004950

Lind, M., Simonsen, H. G., Hansen, P., Holm, E., & Mevik, B.-H. (2013). "Ordforrådet"-En leksikalsk database over et utvalg norske ord ("Ordforrådet"-A lexical database of a selection of Norwegian words). *Norsk Tidsskrift for Logopedi, 59*(1), 18-26.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Long, M., Shukla, V., & Rubio-Fernandez, P. (2021). The development of simile comprehension: From similarity to scalar implicature. *Child Development, 92*(2), e1020-e1035. https://doi.org/10.1111/cdev.13507

Løvstakken, E., & Neff, M. B. (2024, May 1). Novel metaphor stimuli for children. OFSHome. https://doi.org/10.17605/OSF.IO/VJN42

The ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science, 3*(1), 24-52. https://doi.org/10.1177/2515245919900809

Matthews, D. (2014). *Pragmatic development in first language acquisition (Vol. 10).* John Benjamins Publishing Company.

Martín-González, I., Ronderos, C. R., Castroviejo, E., Schroeder, K., Lossius-Falkum, I., & Vicente, A. (2025). That kid is a grasshopper! Metaphor development from 3 to 9 years of age. *Journal of Child Language, 52*(4), 945-970. https://doi.org/10.1017/S0305000924000187

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology, 23*(3), 383-399. https://doi.org/10.1348/026151005X26732

Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol, 16*(1-2), 109-121. https://doi.org/10.1080/10926488.2001.9678889

Özçalişkan, Ş. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language, 32*(2), 291-318. https://doi.org/10.1017/S0305000905006884

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Re-search Methods, 51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics, 25*(1), 51-79. https://doi.org/10.1075/bjl.25.04pou

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.pragma.2019.05.021

Quinn, G. P., & Keough, M. J. (2002). *Experimental designs and data analysis for biologists.* Cambridge University Press. https://doi.org/10.1017/CBO9780511806384

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ronderos, C. R., Falkum, I. L., Noveck, I. A., & Mathisen, E. (2022, June 3). Development of imprecise usages of absolute adjectives. *OSF Registries.* https://doi.org/10.17605/OSF.IO/KUGPF

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416-420.

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology, 61*, 273-301. https://doi.org/10.1146/annurev.psych.093008.100519

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language, 34*(1), 3-23. https://doi.org/10.1177/0142723713510997

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition, 153*, 6-18. https://doi.org/10.1016/j.cognition.2016.04.006

Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. In R. W. Gibbs Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84-105).

Cambridge University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Van Herwegen, J., Dimitriou, D., & Rundblad, G. (2013). Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities, 34*(4), 1300-1311. https://doi.org/10.1016/j.ridd.2013.01.017

Vicente, A., & Falkum, I. L. (2021). Accounting for the preference for literal meanings in ASC. *Mind & Language, 38*(1), 119-140. https://doi.org/10.1111/mila.12371

Vicente, A., & Martín-González, I. (2021). El sesgo literalista en las condiciones del espectro autista: revisión de las teorías existentes [The literalist bias in the autistic spectrum conditions: Review of existing account]. *Studies in Psychology, 42*(4), 298-333. https://doi.org/10.1080/02109395.2021.1909248

Vosniadou, S. (1987). Children and metaphors. *Child Development, 58*(3), 870-885. https://doi.org/10.2307/1130223

Vosniadou, S. (1989). Context and the development of metaphor comprehension. *Metaphor and Symbol, 4*(3), 159-171. https://doi.org/10.1207/s15327868ms0403_4

Vosniadou, S., Ortony, A., Reynolds, R. E., & Wilson, P. T. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development, 55*(4), 1588-1606. https://doi.org/10.2307/1130028

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In Noel Burton-Roberts (Ed.), *Pragmatics* (pp. 230-259). Palgrave-Macmillan.

Wilson, D., & Sperber, D. (2006). Relevance theory. In K. Brown (Ed.), *The handbook of pragmatics* (pp. 607-632). Blackwell Publishing.

Winner, E. (1997). *The point of words: Children's understanding of metaphor and irony.* Harvard University Press. (Original work published 1988)

Winner, E., Engel, M., & Gardner, H. (1980). Misunderstanding metaphor: What's the problem? *Journal of Experimental Child Psychology, 30*(1), 22-32.

https://doi.org/10.1016/0022-0965(80)90072-7

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

Zufferey, S. (2015). *Acquiring pragmatics: Social and cognitive perspectives.* Routledge.

Data, Code and Materials Availability Statement

All data, materials, and analysis scripts are available on the Open Science Framework at https://osf.io/jkq9w/, except for children's demographic profiles. These are instead provided as aggregated summaries in the Supplementary Information section of the manuscript. As these data were considered non-anonymized personal data, they could not be shared in raw form.

Ethics Statement

All experimental protocols adhered to The National Committee for Research Ethics in the Social Sciences and the Humanities guidelines and have been approved by the Norwegian Agency for Shared Services in Education and Research (Reference Number: 596365). All participants gave informed written consent before taking part in the study.

Authorship and Contributorship Statement

Mary Beth Neff: conceptualization, methodology, software, validation, resources, writing – original draft, writing – review & editing, visualization, project administration, data curation, investigation, and formal analysis. **Ingrid Lossius Falkum**: supervision, methodology, resources, writing – review & editing, funding acquisition.

Funding

This research was supported by the European Union's Horizon 2020 research and innovation program, under grant agreement 853211 (an ERC Starting Grant 2019 awarded to Ingrid Lossius Falkum).

Acknowledgements

We would like to thank Roger Mundry for advice on statistical simulations, as well as Nausicaa Pouscoulous for her help with replicating the original study. We would also like to thank all of the participating families and the DEVCOM research assistants who made this research possible. In particular, we thank Eline Løvstakken for her help in creating the metaphorical stimuli, Emma Krane Mathisen and Henriette Johansen for voicing the characters, Tyler Green and Peter Philipp for trouble-shooting PsychoPy, and Håkon Bjørn Due for testing all those participants!

Supplementary Information

Table S1. ***Children's average language development demographics per experiment and age group***

age	exp	*N*	% female	% exposure to lang1	% lang1 = norsk	% w/ language dev't issues	% w/ dev't issues	% care-giver A w/ master's edu	% care-giver B w/ master's edu
3	3	16	19	95	100	19	0	50	44
3	2	16	38	97	94	0	0	50	25
3	1	18	50	84	92	11	0	50	56
5	3	17	29	84	91	6	6	31	44
5	2	17	24	83	94	12	0	29	53
5	1	13	54	82	92	15	0	33	33
7	3	14	21	90	96	14	7	15	31
7	2	24	50	83	94	13	0	50	32
7	1	10	70	96	85	40	10	14	0

Note. All children were born in Norway. Sample sizes will vary from the original experimental groups as the form was optional for caregivers and could include excluded participants (for practice failures). We did not collect these data from adult participants.

Table S2. ***Average inhibitory control task scores per experiment / age group***

Age Group	Experiment	*N*	*Mean*	*SD*
Adult	Experiment 1	26	0.81	0.39
	Experiment 2	26	0.88	0.32
	Experiment 3	16	0.83	0.38
7	Experiment 1	25	0.97	0.06
	Experiment 2	26	0.98	0.05
	Experiment 3	15	0.96	0.07
5	Experiment 1	27	0.74	0.31
	Experiment 2	26	0.74	0.32
	Experiment 3	16	0.83	0.15
3	Experiment 1	28	0.76	0.23
	Experiment 2	28	0.67	0.33
	Experiment 3	17	0.54	0.32

Note. Sample sizes will vary from the original experimental groups as this task was always run last, some participants withdrew before taking the task, and data is from all possible subjects (including replacements and those excluded for practice failures). Task = DayNight for children and = Flanker for adults.

Table S3. ***Proportion of metaphorical picture selections in Experiment 1 by item and metaphor types for 3-year-olds***

Age	Metaphor Type	Metaphor Item Statement in English	Mean Correct (proportion)
3	Replicated Attributional	The tower with the hat	1.00
		The car with the backpack	0.73
		The carrot with the hair	1.00
		The dog with the brown shoes	0.92
		The bottle with the belly	0.85
	New Attributional	The tree with the arms	1.00
		The head with the spaghetti	0.89
		The sky with the cookie	0.54
		The hot cocoa with the pillows	0.85
		The tree with the fried eggs	0.42
	Functional	The frog with the umbrella	0.92
		The fox with the flashlight	0.88
		The monkey with the hammer	0.69
		The squirrel with the cup	0.77
		The bug on the plane	0.96
	Psychological	The plant that is sad	0.73
		The balls that are happy	0.65
		The drawing that is angry	0.81
		The bag that is sleeping	0.69
		The boy that is on fire	0.81

Table S4. ***Proportion of metaphorical picture selections in Experiment 1 by item and metaphor types for 5-year-olds***

Age	Metaphor Type	Metaphor Item Statement in English	Mean Correct (proportion)
5	Replicated Attributional	The tower with the hat	1.00
		The car with the backpack	0.93
		The carrot with the hair	1.00
		The dog with the brown shoes	1.00
		The bottle with the belly	1.00
	New Attributional	The tree with the arms	0.89
		The head with the spaghetti	1.00
		The sky with the cookie	0.89
		The hot cocoa with the pillows	1.00
		The tree with the fried eggs	0.07
	Functional	The frog with the umbrella	1.00
		The fox with the flashlight	0.85
		The monkey with the hammer	0.89
		The squirrel with the cup	0.93
		The bug on the plane	0.96
	Psychological	The plant that is sad	0.96
		The balls that are happy	0.93
		The drawing that is angry	0.96
		The bag that is sleeping	1.00
		The boy that is on fire	0.96

Table S5. ***Proportion of metaphorical picture selections in Experiment 1 by item and metaphor types for 5-year-olds***

Age	Metaphor Type	Metaphor Item Statement in English	Mean Correct (proportion)
7	Replicated Attributional	The tower with the hat	1.00
		The car with the backpack	0.92
		The carrot with the hair	1.00
		The dog with the brown shoes	1.00
		The bottle with the belly	1.00
	New Attributional	The tree with the arms	1.00
		The head with the spaghetti	1.00
		The sky with the cookie	0.92
		The hot cocoa with the pillows	1.00
		The tree with the fried eggs	0.28
	Functional	The frog with the umbrella	1.00
		The fox with the flashlight	1.00
		The monkey with the hammer	0.92
		The squirrel with the cup	1.00
		The bug on the plane	0.96
	Psychological	The plant that is sad	1.00
		The balls that are happy	0.88
		The drawing that is angry	1.00
		The bag that is sleeping	1.00
		The boy that is on fire	0.92

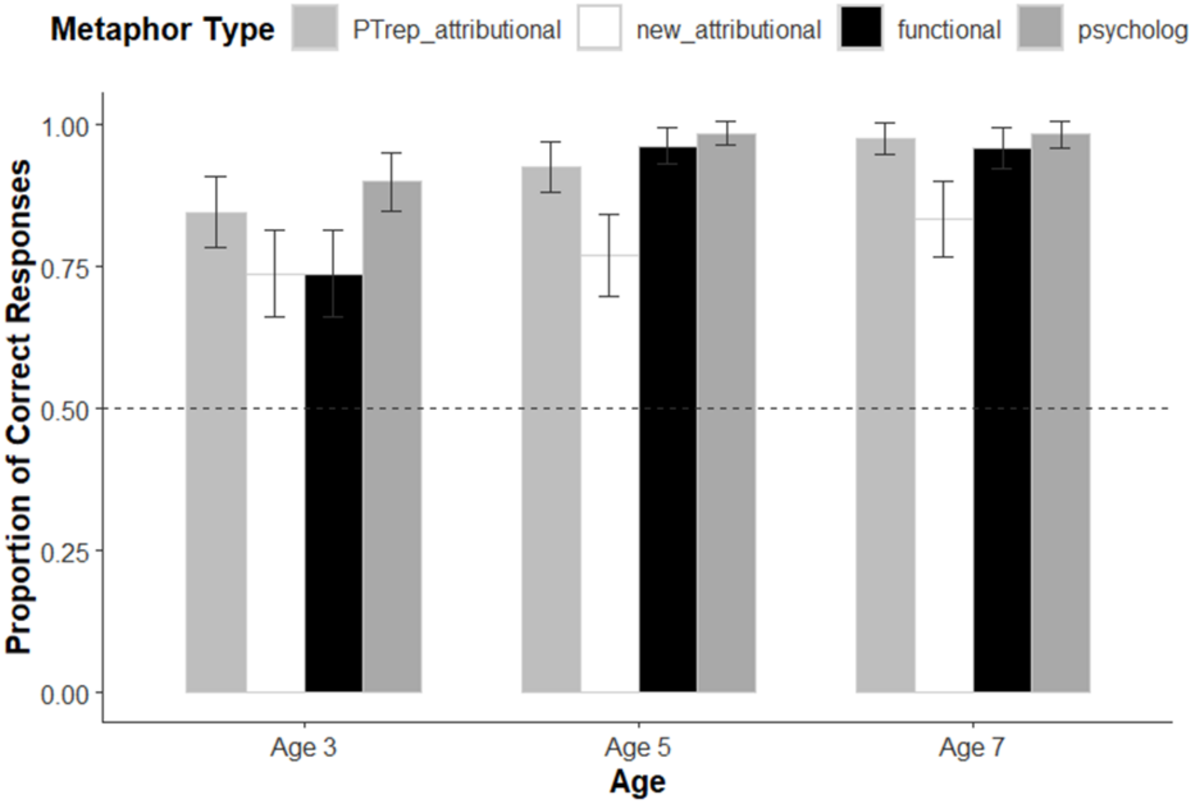


Figure S1. ***Proportion of metaphorical picture selections in Experiment 1 across all metaphor types and ages.***

License

Metaphor comprehension in preschool children: Individual differences and metaphor-related factors affecting metaphor comprehension

Pelin Küçükerdoğan
Deniz Tahiroglu
Boğaziçi University, Turkey

Abstract: Metaphor comprehension in children has been a topic of interest for decades. Contrary to the previous belief that metaphor comprehension emerges at the formal operational stage, recent findings suggest that preschool children can also comprehend metaphors when the metalinguistic demands are controlled. However, these studies have primarily focused on children's physical metaphor comprehension, with only a few investigating psychological metaphor comprehension and the factors that influence this process. This study addresses this gap by presenting stories containing psychological metaphors to 62 Turkish-speaking preschool children. Participants were asked to identify the emotion of the character in the story (revealed in the metaphoric phrase only) by selecting the correct picture. Effects of individual differences in cognitive abilities (cognitive flexibility skills, pretend play, and language complexity) and metaphor-related factors (context and familiarity) were also investigated. Overall, children were more likely to choose the picture depicting the correct emotion compared to the distractors and incorrect pictures. A positive relationship was found between cognitive flexibility and metaphor comprehension; however, no correlation was observed between metaphor comprehension and either pretend play or language complexity. Contrary to our expectations, context and familiarity with the metaphors did not significantly affect metaphor comprehension. However, a marginal interaction between cognitive flexibility and metaphor familiarity suggested that cognitive flexibility might be especially important for understanding less familiar metaphors. Generally, the current study replicated and extended the findings, suggesting that preschool children have some understanding of psychological metaphors and emphasized the importance of simultaneously investigating the effects of cognitive and metaphor-related factors.

Keywords: metaphor comprehension; psychological metaphors; cognitive flexibility; pretend play; language complexity

Corresponding author: Pelin Küçükerdoğan, Department of Psychology, Boğaziçi University. Email: pelinkckrdgn@gmail.com

ORCID ID(s): Pelin Küçükerdoğan: https://orcid.org/0009-0008-3984-5939; Deniz Tahiroglu: https://orcid.org/0000-0003-3871-8811

Citation: Küçükerdoğan, P. & Tahiroglu, D. (2025). Metaphor comprehension in preschool children: Individual differences and metaphor-related factors affecting metaphor comprehension. *Language Development Research*, *5*(2), 172–203. http://doi.org/10.34842/ldr2025-807

Introduction

The development of figurative language comprehension, particularly metaphor comprehension, is crucial for linguistic and cognitive growth in children. Metaphors are prevalent in daily communication, storybooks, and cartoons (e.g., Marriott, 2002; McCrindle & Odendaal, 1994; Taggart et al., 2019). They are utilized to express complex ideas, emotions, and abstract concepts (Vosniadou, 1987). Therefore, understanding how children acquire and comprehend metaphors can provide valuable insights into their cognitive development and linguistic capabilities.

Although children’s metaphor comprehension has received substantial attention, a dispute persists regarding how well they understand and learn from metaphorical expressions. The literature presents conflicting findings concerning the extent to which children possess a solid understanding of metaphorical phrases at an early age. While previous investigations by Piaget (1926), Asch and Nerlove (1960) and Winner et al. (1976) suggest that children under the age of 10 struggle to understand metaphors, more recent studies by Deamer (2013), Pouscoulous and Tomasello (2020) and Özçalışkan (2005) indicate that younger children do possess some degree of metaphor comprehension when tested with age-appropriate tasks and with physical metaphors (e.g., “tower with a hat” implying the roof over the tower). These recent studies provide evidence for early metaphor comprehension and highlight the role of age-appropriate assessment methods. Extending this line of inquiry, we also consider it important to investigate psychological metaphor comprehension specifically (e.g., “shining like the sun” implying happiness), as well as cognitive and metaphor-related factors that may influence the comprehension process in preschool children. Psychological metaphor comprehension is a critical area of inquiry, as it holds potential implications for both abstract language development and practical applications in the clinical field. For instance, facilitating the acquisition of psychological metaphor comprehension may enhance children’s ability to recognize and express emotions, thereby contributing to more effective therapeutic interventions and improved emotional well-being.

Therefore, in the current study, we attempt to assess preschool children's psychological metaphor comprehension abilities in relation to both individual differences in cognitive abilities (e.g., executive function skills) and metaphor-related factors (e.g., context and familiarity). First of all, we present the theoretical concept of metaphors and discuss existing theories in regards to children's understanding of metaphors. Further on, we discuss the potential impact of cognitive abilities and metaphor-related factors on the metaphor comprehension process, with a focus on psychological metaphor comprehension.

Theoretical Definition of Metaphors

According to the Conceptual Metaphor Theory (Lakoff & Johnson, 1980), metaphors are composed of a vehicle concept (base) and a topic concept (target). In a metaphorical expression, the role of the vehicle is utilized to reshape our understanding of the topic in an abstract fashion. Namely, the expression "lion-hearted kid" implies that the topic 'kid' bears a resemblance to a 'lion', which serves as the vehicle in this scenario, in terms of the bravery and power lions possess.

The nature of the relationship between the vehicle and the topic has received various interpretations from different accounts. For instance, Gentner (1983) proposes the Structure Mapping Theory, in which she claims that the metaphor comprehension process involves a structural mapping between the vehicle and topic based on their shared features. In other words, this account suggests a comparison between two concepts, focusing on their similarities. Hence, according to the comparison account, “the kid” is similar to “a lion” because they are both brave.

In another line of research, Glucksberg and Keysar (1990) claim that in a metaphorical phrase, the topic is assigned to the category of the vehicle by abstraction. Thus, the key idea of the class inclusion (categorization) account is that metaphors are not merely comparisons, but rather statements of category inclusion, where the topic is considered a member of the superordinate category represented by the vehicle. According to this account, “the kid” is in the same category as “a lion,” because both are characterized by bravery. In yet another account, the Career of the Metaphor Theory, Bowdle and Gentner (2005) suggest that whether the relationship between vehicle and topic is based on comparison or categorization depends on the conventionality of the metaphor (also see Blank, 1988; Giora, 1997; Turner & Katz, 1997 for similar approaches). To be more specific, they suggest that novel metaphors are more likely to be perceived as comparisons, while conventional metaphors (ones that have become widely recognized over time) are more likely to be perceived as categorizations.

Although these accounts differ on the exact nature of the relationship between the vehicle and the topic, they all suggest a level of abstraction based on shared attributes of the two concepts. Given these characteristics of metaphors, researchers propose that metaphors play a fundamental role in shaping how we understand and conceptualize the world (Lakoff & Johnson, 1980; Vosniadou, 1987). Thus, they serve a meaning-forming function, particularly in relation to novel or abstract concepts. It is argued that metaphors shape how we think, allowing us to grasp complex and abstract ideas by relating them to more familiar, concrete experiences. For example, we often understand time in terms of physical entities or actions ('time is money' or 'time flies'), which helps make the intangible concept of time more comprehensible through everyday experiences. Additionally, metaphors may be especially valuable for fostering emotion comprehension (Fainsilber & Ortony, 1987; Fetterman et al., 2016; Kovecses, 1988, 2000). For instance, in a study emphasizing how metaphors can be effective in emotion regulation, Fetterman et al. (2016) found that people experience less negative affect when instructed to articulate their negative memories using metaphorical language, compared to recalling them without any additional prompt, implying an emotion-regulating function of metaphorical language. These findings indicate that metaphors may play a significant role in the context of emotions.

Research examining the role of metaphors in this particular way has primarily focused on adults. However, understanding whether this function of meaning-forming in the context of emotions also holds for children is important, since children have fewer tools than adults to understand and express their emotions (Chronaki et al., 2015). Therefore, identifying additional mechanisms that might enhance children's emotional understanding could be helpful. Nevertheless, before investigating whether children can utilize metaphors to gain a better understanding of their own and others' emotions in further research, it is necessary, as a first step, to address whether young children can understand metaphors that conceptualize psychological states.

Children's Metaphor Comprehension

Earlier accounts of children's metaphor understanding propose that children under the age of 10-12 struggle to understand the abstract connection between the vehicle and topic in a metaphorical expression (Cometa & Eson, 1978; Piaget, 1926; Smith, 1976). Consistent with this perspective, Asch and Nerlove (1960) provided evidence that children must have reached the formal operational stage to effectively articulate the abstract connection between the topic and vehicle in a metaphorical expression.

However, these initial investigations were criticized for relying solely on children's capacity to explicitly verbalize the meaning of metaphors and treating performance scores as definitive proof of metaphor comprehension (e.g., Deamer, 2013; Özçalışkan, 2005; Pouscoulous, 2014, 2023; Pouscoulous & Tomasello, 2020; also see Gardner, 1974; Gentner, 1977; Pearson, 1990; Waggoner et al., 1985). Verbally explaining metaphors involves meta-linguistic abilities and the capacity to engage in discussion about language itself, which differs from understanding metaphorical phrases (Vosniadou, 1987). In other words, one can understand the meaning of a metaphorical phrase but fail when trying to articulate the reason, due to the lack of meta-linguistic skills. With this criticism in mind, recent studies on metaphor comprehension have employed alternative methods that are not limited to verbal responses (e.g., Deamer, 2013; Pouscoulous & Tomasello, 2020; Waggoner & Palermo, 1989; Zhu & Gopnik, 2023). These studies recognize that poor performance on earlier metaphor comprehension tasks may be attributed to younger children's limited meta-linguistic abilities rather than insufficient comprehension. For example, Pouscoulous and Tomasello (2020) conducted an experiment in which the experimenter described the toys using metaphorical phrases, such as "the tower with a hat" (referring to the roof) and asked 3-year-olds to choose the correct toy from two possibilities. The children in the test group (receiving metaphorical instruction) performed above chance levels and demonstrated a preference for the metaphorically correct item compared to the control group, which was not given a definition and was simply asked to select one of the toys.

The majority of recent research focuses on physical metaphor comprehension, and there are only a few studies addressing psychological metaphor comprehension (e.g., Lecce et al., 2019; Nippold et al., 1984). In such a study by Waggoner and Palermo (1989), participants of various ages, including 5-, 7-, 9-year-olds, and college students, listened to ambiguous stories in which the emotional state of the protagonist was revealed only at the end through a metaphorical expression (e.g., 'Betty was a bouncing bubble'). This phrase served as the final sentence of the story, followed by a question offering two verbal options for the protagonist's emotion (e.g., 'Was she happy or sad?'). The findings provided evidence that even 5-year-olds could choose the accurate emotion for the protagonist above chance levels and that metaphor comprehension increased with age.

While this study significantly contributes to our understanding of the psychological metaphor comprehension abilities of children across different age groups, it presents a limitation that warrants further investigation. In that study, the children were asked to choose between options where one showed a positive emotion and the other showed a negative emotion; thus, the design did not allow for distinguishing whether they understood the general valence of the metaphorical phrase or grasped the specific emotion the metaphor conveyed, which is important for understanding the nuanced process of children’s emotional understanding. In the current research, we aim to address this gap. In addition to assessing preschool children's comprehension of psychological metaphors, we also aim to examine whether their cognitive abilities and metaphor-related factors are correlated with their psychological metaphor comprehension.

Cognitive Factors

Metaphor comprehension is believed to be a cognitive and linguistic skill (Özçalışkan, 2005; Vosniadou, 1987). However, existing research lacks a comprehensive discussion of how individual variations in cognitive and linguistic abilities impact metaphorical understanding. This is exemplified by studies that involve different age groups, the results of which consistently show that metaphorical understanding ability increases with age (ages 5 to 12), implying that cognitive abilities enabling metaphorical understanding unfold over time (Asch & Nerlove, 1960; Waggoner & Palermo, 1989; Winner et al., 1976). Nevertheless, it has not been systematically investigated which cognitive abilities facilitate the process. In this study, we intend to investigate how executive function (EF) skills, the tendency to engage in pretend play, and complex language use might relate to metaphor comprehension in preschoolers.

Executive function (EF) skills are higher-order cognitive capacities with three commonly acknowledged subcomponents: inhibition, cognitive flexibility, and working memory (see Miyake et al., 2000; Spiegel et al., 2021). We propose that out of these three, cognitive flexibility, which involves the ability to adapt and shift perspective based on new information, particularly plays a significant role in metaphor comprehension. According to fundamental metaphor theories (Gentner, 1983; Glucksberg & Keysar, 1990; Lakoff & Johnson, 1980), understanding a metaphor involves inhibiting the literal meaning of the topic and using structural mapping to transfer shared features from the vehicle to the topic. This process also requires a mental shift from literal to figurative language, which inherently demands a degree of cognitive flexibility.

There are a few studies that have already explored the relationship between EF skills and metaphor comprehension. For example, Deamer (2013) found a positive relationship between inhibitory control and physical metaphor comprehension in preschool children. Meanwhile, Carriedo et al (2016) did not find a prevailing impact of EF skills on physical metaphor comprehension in a sample of older children (aged 10-15), yet they did observe that EF skills had a positive influence on understanding more challenging metaphors. These studies provide valuable insights; however, the present research is novel in that it focuses on preschool children and their comprehension of psychological metaphors. This focus is crucial for several reasons. First, the relationship between EF skills and metaphor comprehension may differ from that observed in older children due to the gradual development of EF abilities (Miyake et al., 2000). Furthermore, in the present study, we chose to narrow our focus to study cognitive flexibility specifically, rather than inhibitory control, because the ability to shift between different perspectives or representations may be important for interpreting abstract, psychological metaphorical expressions. While inhibitory control may assist in suppressing literal interpretations, cognitive flexibility can be particularly relevant for psychological metaphors, which often require the child to go beyond surface-level features and integrate multiple mental representations. This is supported by findings suggesting that more abstract metaphors pose greater cognitive demands (Waggoner & Palermo, 1989). Therefore, we anticipate a positive relationship between cognitive flexibility and the comprehension of psychological metaphors in this age group.

Another cognitive factor we examine in the current study is children's tendency to engage in pretend play. During pretend play, children may treat an object as if it were something else (e.g., while pretending to talk on the phone, a child holds a banana to their ear, thinking of the banana as a phone), act as if they were someone else, or engage in an activity as if they were doing something else (e.g., feeding a baby as a caregiver) (Garvey, 1990; Lillard, 1993). In these cases, similar to metaphorical comprehension, children block the object's prominent function (banana) and reconceptualize it using another vehicle (phone). Due to the similarity between pretend play and metaphor production, several earlier studies even considered instances of pretend play as metaphor production (Billow, 1981; Winner, 1979; Winner et al., 1979; Winner et al., 1980). For example, making a toy animal eat grass, represented by the experimenter's hair, was interpreted as metaphorical (Billow, 1981). This treatment is problematic since to form a link between a vehicle and a topic in a metaphor, there needs to be a shared attribute, but this is not necessary for pretend play (Vosniadou, 1987). Yet, although it is not a requirement, perceptual similarities between the vehicle and topic are often observed in pretend play as well (Garvey, 1990). Therefore, it is clear that both concepts have similarities in terms of imposing familiar schemas on novel concepts, which provides a basis for anticipating a positive relationship between them.

The final cognitive factor we plan to investigate is the complexity of children's language use. Metaphor comprehension is considered a linguistic capacity (Özçalışkan, 2005; Vosniadou, 1987); however, there is limited research in this area. A few studies have found that metaphor understanding is positively associated with vocabulary abilities (Pouscoulous & Perovic, 2023) and verbal reasoning (Carriedo et al., 2016). We propose that another language-related domain that may be linked to metaphor comprehension is language complexity, which refers to the sophistication and structural variety of the language children use in their everyday interactions The complexity of language used in everyday contexts may serve as an indicator of children's capacity for metaphor comprehension, as both rely on advanced linguistic processing and abstract reasoning skills. With this in mind, analyzing the structural complexity of children's language can offer valuable insights into their ability to understand psychological metaphors.

To sum up, comprehension of psychological metaphors, cognitive flexibility, and the tendency to engage in pretend play may share similar underlying mechanisms. Therefore, we anticipate a positive association between all of these factors. Furthermore, given that metaphorical language is a more complicated form of speech (Lakoff & Johnson, 1980), we anticipate that children who are more proficient in using complex language will have higher psychological metaphor comprehension skills.

Metaphor-Related Factors

Looking at the metaphor comprehension process from a different angle, we can see that while children's cognitive and linguistic abilities contribute to its development, metaphor-related mechanisms also play a role. For example, in several investigations, researchers discovered that the same children understood certain metaphors but not others (Keil, 1986; Vosniadou et al., 1984), implying that the nature of metaphors influences the comprehension process as well.

Several observations support the common intuition that more complex metaphors are more difficult to understand (Johnson & Pascual-Leone, 1989; Vosniadou et al., 1984; Waggoner & Palermo, 1989; Winner et al., 1976). However, there is no consensus on the operational definitions of "complex" and "simple" metaphors. As a result, different researchers focused on varying aspects of metaphorical phrases that influence overall complexity and, consequently, the comprehensibility of metaphors. Namely, Waggoner and Palermo (1989) utilized the abstract-concrete distinction to compare complex and simple metaphors, whereas Vosniadou et al. (1984) employed the simile-predicative distinction. Additionally, Jones and Estes (2006) introduced aptness as a factor influencing the comprehensibility of a metaphor. We acknowledge the significant effects of these factors; thus, we aimed to control them by using only the predicative form of metaphors, restricting the selection to abstract (psychological) metaphors, and ensuring similar levels of aptness through a pilot study. With these controls in place, we aimed to examine different factors (context and familiarity) that have not been previously addressed but might be particularly important for children.

We believe that the context in which a metaphor is presented can influence its complexity and, consequently, its comprehensibility. As mentioned earlier, in life, we are exposed to abstract language, such as metaphors, within a specific context, which provides various cues to aid interpretation. One illustrative example is the use of facial expressions and intonations during speech, which may aid in understanding metaphorical phrases representing a person's emotions. These cues are universally present and can be noticed by everyone (Nelson et al., 1979). Nonetheless, understanding some cues might require higher-order reasoning; therefore, it may be harder to interpret a metaphor presented with such cues. For example, inferring what someone might feel in a given scenario or during an action can demand higher-order reasoning. While it might be easier to understand the metaphor based on facial expressions, interpreting an action or a scenario in order to understand a psychological metaphor may be less effective for individuals who have not yet developed higher-order reasoning skills. For instance, seeing someone's vase broken and assuming that they will be disappointed upon noticing the broken vase is more difficult than simply observing a person's sad facial expression. This is because it involves two steps: first, having the emotional understanding capacity necessary to recognize that someone would feel sad in such a scenario, and second, having the capacity to understand the metaphors that describe this feeling in the given scenario. Consequently, we expect that when a metaphor is presented with more accessible and easier cues (e.g., facial expression) that aid interpretation, children's comprehension will be higher compared to metaphors presented with less obvious cues (e.g., embedded in a scenario).

The familiarity of metaphors can be another factor that influences metaphor complexity and, as a result, metaphor comprehension. While we acknowledge that conventionality is a more commonly used measure and is often considered a key factor influencing a metaphor's complexity since it affects how a metaphor is conceptualized (Bowdle & Gentner, 2005), we argue that familiarity can be a more relevant factor when working with children. Assessing the conventionality of metaphors in children can be challenging, as they have varying levels of experience with metaphors. However, we propose that children's familiarity with a given metaphorical phrase may influence their comprehension process in a similar way to how conventionality influences adults' comprehension. Thus, the connection between the topic and the vehicle may be more readily apparent to children only when they are acquainted with the metaphor.

In summary, we expect that children's comprehension performance will differ depending on the level of complexity, influenced by contextual cues and familiarity. Specifically, we predict that easier contexts and more familiar metaphors will facilitate better metaphor comprehension. We also aim to investigate the possible interactions between the levels of metaphor-related factors and executive function, as indicated by the findings of Carriedo et al. (2016). In particular, we expect cognitive flexibility to be more influential when the context is hard and the metaphors are unfamiliar.

The Current Study

To examine how cognitive abilities and metaphor-related factors interact in psychological metaphor understanding, we utilized a methodology similar to the one employed by Waggoner and Palermo (1989). We created ambiguous short stories, each ending with a metaphorical statement. Children were then instructed to select the picture that best represented the ending of the narrative based on the metaphorical term. Through this approach, we intended to eliminate meta-linguistic requirements and offer children a setting in which they could encounter metaphors, mirroring the way they would encounter metaphors in real life rather than simply being presented with the metaphorical phrase.

Previous research (Waggoner & Palermo, 1989) has shown that preschool children (4–5-year-olds) have some level of metaphor comprehension. Accordingly, we chose this age group to explore these effects in a developmentally relevant context, based on the assumption that it would provide an opportunity to investigate the influence of the aforementioned cognitive and metaphor-related factors on metaphor comprehension.

In conclusion, this study is expected to provide insights into whether preschool children can comprehend psychological metaphors and how both cognitive abilities and characteristics of these expressions influence their understanding. We anticipated positive relationships between cognitive abilities (cognitive flexibility, pretend play, and language complexity) and psychological metaphor comprehension. Furthermore, we hypothesized that metaphors presented in an easier context would be easier to understand than metaphors presented in a harder context. Likewise, we expected better comprehension levels when metaphors are more familiar compared to their less familiar counterparts. Finally, we intended to explore interactions between cognitive and metaphor-related factors with the expectation of cognitive flexibility to facilitate psychological metaphor comprehension with metaphors from harder contexts and with less familiarity.

Method

Participants

The study was conducted with 62 Turkish-speaking preschool children (29 boys, 33 girls) aged from 45 months (3 years, 9 months) to 79 months (6 years, 7 months) (M_{age} = 61 months, *SD* = 7 months). An additional participant was tested but excluded from the analyses due to the parents' report of a developmental disorder. The targeted number of participants (*N* = 60) was decided a priori according to the reported sample sizes for the *t*-tests of Pouscoulous and Tomasello (2020) and Waggoner and Palermo (1989), because previous similar studies did not report effect size.

The children in the main study were recruited from municipality preschools and private preschools. At the end of testing, all of the children in the study received an age-appropriate storybook as a gift and a personalized certificate of “Participation in Science.”

Materials and Procedure

After written consent forms were received from the parents, the children were examined in a room at their preschools by the first author. Verbal assent was also obtained from the child before starting the tasks. The researcher conducted a series of tasks in the following order: language complexity, metaphor understanding, cognitive flexibility, and pretend play. As this is a standard practice in individual differences research, the tasks were given in a fixed order (e.g., Carlson & Moses, 2001; Pomareda et al., 2024). Language complexity and metaphor comprehension tasks were administered via computer. The duration of the testing procedure for each child was approximately 20-25 minutes. A video recording was done throughout the entire testing time, except for the three participants, whose parents refused to consent to visual filming. In those cases, only an audio recording was taken. Following the testing with the children, the parents were requested to complete an online questionnaire that contained demographic questions and familiarity ratings of metaphors using the Qualtrics platform.

Language Complexity

In order to evaluate the complexity of language, the researcher presented a brief (1 minute and 35 seconds long) video clip from a cartoon without sound. After watching the clip, the children were asked to describe what was happening in the cartoon. The depicted cartoon showcased Sylvester the Cat chasing Tweety, the bird in the cartoon, while some chickens on the farm help Tweety evade and hide from Sylvester (see supplementary material for the video clip). In the current study, the language complexity of the children's narratives regarding the cartoon was assessed using a coding system created by Berman and Slobin (1994).

In this system, a clause was defined as a phrase consisting of one or more predicates that describe an action, state, or event. If a clause contained only one predicate, it was considered a simple clause, and if it contained two or more predicates, it was considered a complex clause. For example, "Mary went home" was coded as a simple clause, while phrases such as "Mary went home because her mother called her", "Mary's mother yelled 'Come home'", and "While running back home, Mary fell " were coded as complex clauses. The total number of meaningful words (repetitions included), unique words, clauses, predicates, simple clauses, and complex clauses were coded. To determine the child's ability to use complex language, the number of complex clauses was divided by the total number of clauses uttered by the child.

Two research assistants from the Boğaziçi University's Family and Child Studies Laboratory were trained with pilot language samples, and inter-rater reliability was taken from the codings of six language samples from the main study. A high degree of inter-rater reliability was found between the raters on complex clause percentage measurements: ICC was .98 (95% CI [.86, .99]). Results also showed that ICCs between two raters' scores were high for the total numbers of meaningful words, number of unique words, clauses, predicates, simple clauses, and complex clauses (r's respectively .99, .99, .98, .99, .98, .99 all *p*'s < .05). Two coders then solved their disagreements in the six initially coded language samples and each worked on 28 of the remaining narratives.

Metaphor Comprehension

Short stories of 3-4 sentences were developed for the study. The six stories and twelve metaphors used in the main study were refined through pilot studies in order to counterbalance emotional valence between story sets and obtain similar levels of aptness between metaphors (see supplementary material for stories used in the main study). All stories had two possible endings, which differed by the character's emotions at the end. One ending reflected a positive emotion, whereas the other displayed a negative emotion. The endings included metaphorical terms that children needed to comprehend to understand the emotion depicted in the story. To ensure that the children see just one of the two possible endings for each story, two story sets (A and B) were created. This design helped us assess whether children comprehended the meaning of the metaphorical term rather than guessing what would happen according to the story's build-up.

Whether the context was hard or easy to understand was also controlled between the two story sets. In the easy context condition, children identified the character's feelings by choosing from pictures of facial expressions. In the hard context condition, the children inferred the protagonist’s emotion from pictures showing an action or situation (see Figure 1 for examples). Each story set contained six metaphors, three were presented in easy contexts and three were presented in hard contexts (see supplementary material).

As described in Figure 1, the children saw three pictures after each story: a correct picture, an incorrect picture, and a distractor. The distractor, which was different across the story sets, was based on Ekman’s (1992) work and depicted an emotion that had the same valence as the correct answer, yet it did not tap into the exact feeling the metaphor describes. More specifically, Ekman (1992) proposed that emotions like happiness and surprise are associated with pleasant outcomes, while sadness, anger, disgust, and fear are linked to unpleasant stimuli. Therefore, we used happiness or surprise as distractors for metaphors depicting positive emotions, and sadness, anger, and fear as distractors for metaphors depicting negative emotions. For instance, in the hard context story in Figure 1, if the story ended with a positive metaphor, then the distractor picture showed a surprised girl who did not find her friend at the doorstep but a mailman carrying a box, and if the story ended with a negative metaphor, then the distractor image depicted a frightened girl who was confronted by a lightning instead of being sad that her friend had not shown up. With this design, we intended to distinguish between whether children can understand the exact emotion the metaphor is describing or just understand the general valence of the emotion (positive or negative).

Procedure		Easy Context	Hard Context
Story Presentation		One day, Ayşe was playing games with her brother Ali. Suddenly, Ayşe noticed a spider in their room and got very scared. She asked Ali to remove the spider. In this situation, *Ali was a lion* .	One weekend, Lale invited her friend Deren to her house to play together. However, it snowed heavily on the day of the meeting and the roads were closed. While Lale was waiting for Deren with hope, she thought she heard a sound at the door. When she opened the door *her hopes faded away* .
Question		Which one of the following pictures could be Ali?	Which one of the following pictures could be Lale?
Picture Selection	Correct		
	Incorrect		
	Distractor		

Figure 1. ***Procedure and example materials from the metaphor comprehension task, story set A***

All participants saw two familiarization trial stories and six test stories during the testing, in the same order. The familiarization trials consisted of similes whereas the test trials consisted of metaphors. The experimenter read all the stories to the participants while a neutral filler picture that showed either the characters or the situation was on the screen. Then she asked, “Which of the following pictures could be the ‘name of the character?’” and showed the three pictures to the children. The children were asked to point to the picture they thought was the correct one. Initially, the correct responses were coded as 2, the distractor responses were coded as 1, and the incorrect responses were coded as 0. Using this coding system, we aimed to control for distractors and evaluate whether children's performance exceeded the chance level by selecting correct pictures over distractors and incorrect options. For the actual metaphor comprehension analyses, correct responses were coded as 1 and both distractor and incorrect responses were coded as 0.

The children’s familiarity with the metaphors used in the study was calculated based on the parents’ reports. After the data collection, we asked the parents to rate their children’s familiarity with the 12 metaphorical phrases that were used in the study, using a Likert scale from 1 “not familiar at all” to 5 “very familiar”. Based on these ratings, metaphors were categorized as either more or less familiar.

Cognitive Flexibility

The Dimensional Change Card Sorting (DCCS, Zelazo, 2006) task that was adapted from the study of Frye et al. (1995) to measure set-shifting abilities and flexibility in older preschoolers was used in this study. The experimenter presented the color game, the shape game, and the border game, respectively, in which children had to sort the cards according to the rules of each game and change their way of sorting once the rules changed. The child's scores were calculated separately for each rule change, and each correct placement was given 1 point. The total score was calculated by adding up the scores for each section with higher scores indicating higher cognitive flexibility skills (the maximum possible score was 24, see supplementary material for detailed procedures).

Pretend Play

Affect in Play Scale - Preschool Version (APS-P, Kaugars & Russ, 2009) was administered to assess the play behaviors of 4-6-year-old children during a 5-minute structured free play session. Children were given 5 minutes to freely play and narrate a story with the standardized toys provided (e.g. softball, plush bears, a car, etc., see supplementary materials for a detailed procedure of the task).

The play sessions were video recorded to be coded later. In the original scale, the coding was carried out in three main categories: Subjective coding (imagination-pretense, organization, elaboration-complexity, and interest-involvement-comfort), affect, and the play type (pretend, functional, and no play). For the current study, only the last category was coded. The type of play the child engaged in, whether pretend play (e.g. using a toy car as a telephone), functional play (e.g. rolling a ball), or no play, was coded for each 20-second period of the total playtime. The ratio of the periods in which the child engaged in pretend play to the total number of times the child continued playing was analyzed as a measure of the child's tendency to engage in pretend play.

Two research assistants from the Boğaziçi University’s Family and Child Studies Laboratory coded the play sessions according to the above-mentioned criteria. Coders were first trained using 5 free play sessions from the main study, and then the inter-rater reliability ratings were obtained using another 12 free play sessions. Results showed high inter-rater reliability for the detection of pretend play in the free play session: ICC was .97 with 95% CI [.79, .98] (F (11,11) = 29.14, $p < .001$). After resolving the disagreements on the 17 sessions they initially coded, the assistants coded the remaining play sessions separately ($N = 21$ each).

Demographic Form

The form given to the parents consisted of basic demographic questions and metaphor familiarity questions (see details in supplementary material). The form was shared with the parents after the session with the children to avoid possible clues parents might give to their children, when looking at the familiarity questions of metaphors.

Results

Preliminary Analyses

Before conducting the main analyses to address the research questions of this study, preliminary analyses were carried out to examine the descriptive characteristics of the dataset and to assess whether it met the assumptions required for the planned statistical analyses. All analyses were done with R (R Core Team, 2024) and IBM SPSS Statistics 29.

The children were randomly assigned to see one of the story set conditions (A = 30 or B = 32). Metaphor comprehension performance, familiarity ratings with metaphors used in this study and the age distribution of participants did not differ between the story sets A and B (Table 1). Therefore, the difference between story sets was not controlled in further analyses. Additionally, gender distribution (A: 16 girls, 14 boys; B: 17 girls, 15 boys; $X^2(1, N = 62) = .00$, $p = .99$) was found to be similar across story sets, as well. Then, how gender might be linked with metaphor comprehension was analyzed, and the boys and girls were found to perform similarly (Girls $M = 63.89$, $SD = 19.17$; Boys $M = 62.36$, $SD = 20.97$, $t(60) = .301$, $p = .76$); thus, gender would not be used in the upcoming analyses either.

Table 1. ***Mean metaphor comprehension, familiarity and age in two story sets***

	Set A Means *(SD)*	Set B Means *M (SD)*	*t*	*df*	*p*
Metaphor comprehension (%)	60.56 (22.09)	65.63 (17.55)	–1.00	60	.32
Familiarity	3.04 (.70)	2.97 (.68)	.375	57	.71
Age (months)	60.67 (6.93)	62.09 (7.66)	–.768	60	.45

Then, analyses were conducted to formulate a collapsed familiarity score per story and classify stories as more or less familiar, accordingly. To do so, familiarity ratings of the individual stories that were received from the parents ($N = 59$) were examined. A within-subject ANOVA showed that although familiarity did not differ between story sets, familiarity levels of metaphors differed significantly across stories (F (5,54) = 4.288, $p = .02$, $\eta p^2 = .28$). Familiarity ratings of the 1st story ($M = 3.54$, $SD = 1.10$) was higher than the 2nd ($M = 2.78$, $SD = 1.19$; 95% CI [.08, 1.44], $p = .02$) 3rd ($M = 2.86$, $SD = 1.17$; 95% CI [.10, 1.25], $p = .01$), 4th ($M = 2.78$, $SD = 1.23$; 95% CI [.18, 1.34], $p = .002$) and 6th ($M = 2.95$, $SD = 1.33$; 95% CI [.06, 1.13], $p = .02$) story but did not significantly differ from the 5th story ($M = 3.12$, $SD = 1.20$; 95% CI [–.23, 1.08], $p = .78$). The familiarity ratings of the 2nd, 3rd, 4th, 5th, and 6th stories did not significantly differ from each other. After considering the pairwise comparisons, mean scores of the 1st and 5th stories were calculated and used as more familiar metaphors while mean scores of the 2nd, 3rd, 4th, and 6th stories were used as less familiar metaphors. An additional within-subject t-test was run to make sure the familiarity ratings of more familiar (the 1st and 5th) metaphors are significantly higher than less familiar (2nd, 3rd, 4th, 5th, and 6th) metaphors. Results showed that the group classified as more familiar ($M = 3.33$, $SD = .81$) indeed had higher familiarity ratings compared to the group classified as less familiar ($M = 2.84$, $SD = .81$, $t(58) = 4.11$, $p < .001$).

After that, normality and outlier analyses were conducted with the main variables of interest. Normality and outlier assumptions were met for all variables except for the language complexity. Data revealed that children generally used less complex language; thus, the variable showed a floor effect (skewness = –1.88, $SD = .33$, kurtosis = 4.80, $SD = .64$), and it had two outliers with 80% ($Z = 3.07$) and 100% ($Z = 4.07$) complex language usage. However, since this usage was believed to be a valid representation of the children's language use and the sample size for the variable was large enough ($N = 56$) to assume robustness, these data points were kept in the dataset. Finally, assumptions like homogeneity of variance, collinearity, etc. were checked during the analysis and reported if there were any violations.

Main Analyses

Descriptives

Descriptives and correlations of age with the performances in behavioral tasks can be seen in Tables 2 and 3. Results revealed a significant relationship between metaphor comprehension and age, $r(62) = .27$, $p = .04$. Correlation analyses also showed a significant positive relationship between metaphor comprehension and cognitive flexibility, DCCS, $r(62) = .31$, $p = .02$. In contrast to our hypotheses, the correlations between metaphor comprehension and other cognitive variables (pretend play and language complexity) were not significant.

Table 2. ***Descriptives of age and performance in the behavioral tasks***

	N	Minimum	Maximum	Mean	SD
Age (months)	62	45	79	61.40	7.30
Metaphor Comprehension (%)	62	8.33	100	63.17	19.88
Language Complexity (%)	56	0	100	18.36	20.13
Cognitive Flexibility	62	6	23	15.13	5.45
Pretend Play (%)	60	0	100	62.74	30.78

Note. The summed-up scores of DCCS were used for the Cognitive Flexibility measure.

Table 3. ***Correlations for age and performance in the behavioral tasks***

	1	2	3	4	5
1. Age (months)	–				
2. Metaphor Comprehension (%)	.27*	–			
3. Language Complexity (%)	.12	.09	–		
4. Cognitive Flexibility	.18	.31*	–.14	–	
5. Pretend Play (%)	-.01	-.04	.22	.11	–

Note. The summed-up scores of DCCS were used for the Cognitive Flexibility measure. $*p < .05$.

Metaphor Stories

To understand whether the children understood the metaphors presented in stories, a within-subject ANOVA test was conducted and the test revealed that the response types (correct, distractor, incorrect) of the children differed significantly ($F(2, 60) = 39.566$, $p < .001$, $\eta p^2 = .57$). Bonferroni corrected pairwise comparisons further showed that children had significantly more correct responses ($M = 3.24$, $SD = .17$) than distractors ($M = 1.66$, $SD = .15$, 95% CI [.835, 2.326], $p < .001$) and incorrect responses ($M = 1.10$, $SD = .11$, 95% CI [1.544, 2.747], $p < .001$). Distractor responses were observed more than incorrect responses as well (95% CI [.056, 1.073], $p = .03$). These findings indicate that children not only understood the valence of the psychological metaphorical phrase but also understood the specific emotion the metaphor conveyed.

Subsequently, Chi-Square analyses were conducted for each story to examine individual patterns of stories. Tests yielded significant results (all *p*'s < .001) for all stories except for the 6th story ($\chi^2 = 4.290$, $p = .12$). In the 6th story, children's correct responses did not differ from incorrect responses. The correct response percentage was significantly higher than the chance level for the 1st, 3rd, 4th, and 5th stories (respectively 68%, 69%, 58%, and 62%). However, for the second story, the incorrect response percentage was significantly higher than the expected frequency by 54%. The response distribution of the stories can be seen in Figure 2. Overall, the metaphor comprehension task revealed that participants exhibited a substantial level of metaphor comprehension. Nevertheless, some items were more difficult to interpret than expected, suggesting variability in item difficulty across the stimulus set.

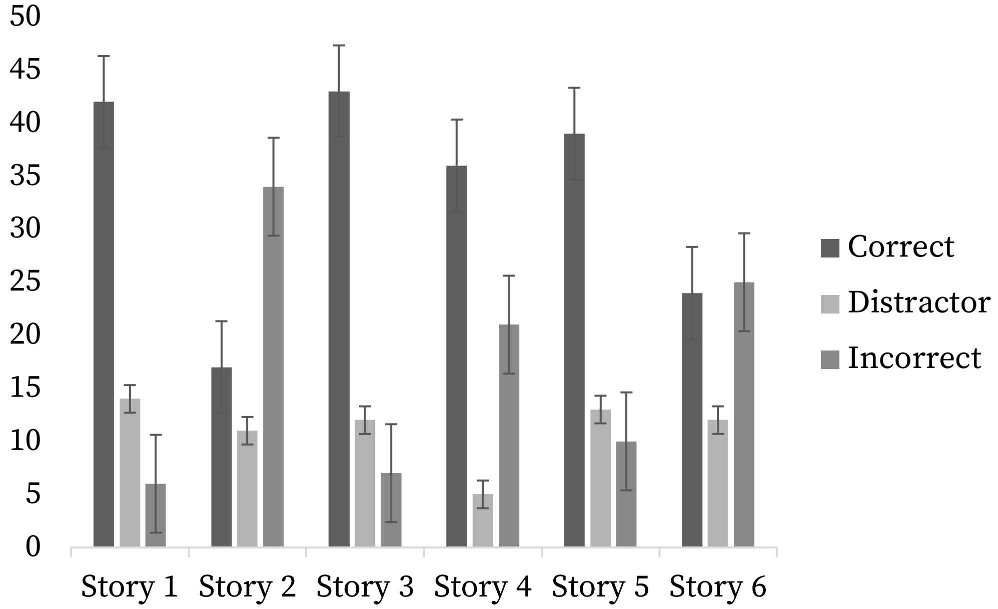


Figure 2. ***Percentage of the types of responses by stories***

Cognitive Factors

Initial correlation analyses indicated a connection between metaphor understanding and cognitive flexibility (Table 3), but no association was found between metaphor comprehension and individual differences in pretend play or language complexity. When looking into non-significant cognitive variables, we observed limited variability in both language complexity and pretend play tasks. The lack of variability in utterances in the language complexity task indicated that participants were reluctant to narrate the story (*N* = 6 missing cases with no verbal response, 25th percentile = 29 words, 50th percentile = 44,50 words, 75th percentile = 79 words). Similarly, while analyzing the patterns in free-play sessions, we saw that children were more inclined to engage in pretend play. The percentage of engaging in pretend play was 40 in the 25th percentile, 66.67 in the 50th percentile, and 93.33 in the 75th percentile, suggesting a ceiling effect. Therefore, the analyses for both factors were repeated with participants above the 50th percentile for clarification; however, results remained unchanged. Consequently, we presented the analyses based on the entire sample.

To further examine whether cognitive factors predicted metaphor comprehension, a logistic mixed-effects model was formed. We used the *glmer*() function from the lme4 package (Bates et al., 2015) in R (R Core Team, 2024) to fit the logistic mixed-effects model. The model used binary coded answers (1: chose the correct picture, 0: chose the distractor or incorrect picture) and included cognitive flexibility scores from the DCCS task, pretend play engagement percentage and complex language use percentage as fixed effects, with random intercepts for both children and story number to account for individual differences and repeated measures across the six stories each child read.

The results indicated that cognitive flexibility was a significant predictor of metaphor comprehension (β = .06, *SE* = .03, *p* = .01). Specifically, for every one-point increase in cognitive flexibility, the odds of comprehension increased by 6% (OR = 1.064, 95% CI [1.01, 1.12]). On the other hand, neither pretend play (β = –.00, *SE* = .00, *p* = .56) nor the use of complex language (β = .01, *SE* = .01, *p* = .45) significantly predicted metaphor comprehension. The variance attributed to the children was 0.12 (*SD* = 0.34), and the variance attributed to the story was 0.47 (*SD* = 0.69). These results indicated substantial variability in the children's baseline metaphor comprehension across both individuals and the stories they encountered. The model provided a reasonable fit to the data, with an AIC of 430.5 and a BIC of 453.2. The inclusion of random intercepts for both the children and the story number improved the model fit compared to a simpler model with only a random effect for children (ΔAIC = –19.2).

Metaphor-Related Factors

Another logistic mixed-effects model was conducted to examine the effects of context (easy vs. hard) and familiarity (less vs. more familiar) on metaphor comprehension, coded as a binary outcome. The model included random intercepts for the children and the stories to account for variability across participants and stories. Results indicated that neither context ($\beta = -.07$, $SE = .49$, $p = .88$) nor familiarity ($\beta = .76$, $SE = .52$, $p = .15$) had a significant effect on predicting metaphor comprehension. In the current model, the variance attributed to the children was 0.25 ($SD = 0.50$), and the variance attributed to the story was 0.29 ($SD = 053$). The model fit statistics were AIC of 498.7 and BIC of 518.3, which are significantly better compared to a model with only the random effects of children ($\Delta AIC = -11.5$).

Exploratory Analyses

Several analyses were conducted to examine whether different levels of cognitive flexibility interact with the levels of metaphor complexity. Before getting into the analyses, for easier interpretation, a median cut categorical cognitive flexibility variable is computed in which scores less than or equal to 18 were labeled as low cognitive flexibility performance ($N = 49$) while scores higher than 18 were labeled as high cognitive flexibility ($N = 13$) performance. First, to see if the effect of cognitive flexibility on metaphors presented in a harder context was more emphasized than metaphors in an easier context, a mixed design ANOVA was performed 2 (low-high cognitive flexibility) x 2 (easy-hard context) using metaphor comprehension as the DV. None of the main effects nor the interaction was significant.

Later, the interaction between cognitive flexibility and familiarity ratings on metaphor comprehension was explored. A mixed design ANOVA 2 (low-high cognitive flexibility) x 2 (more -less familiar metaphors) was conducted to address the hypothesis. The main effect of familiarity was found to be significant with the performance on more familiar metaphors ($M = 76.20$, $SD = 24.97$) being better than less familiar metaphors ($M = 56.65$, $SD = 24.77$; $F(1,60) = 8.611$, $p = .005$, $\eta p^2 = .13$). However, the main effect of cognitive flexibility was not significant ($F(1,60) = .799$, $p = .38$, $\eta p^2 = .01$). Finally, the interaction between cognitive flexibility and familiarity levels was marginally significant ($F(1,60) = 3.751$, $p = .058$, $\eta p^2 = .06$). Bonferroni corrected pairwise comparisons showed that, in the low cognitive flexibility group, the performance on more familiar metaphors ($M = 77.04$, $SD = 26.44$) was significantly higher than on the less familiar metaphors ($M = 53.57$, $SD = 24.34$; 95% CI [14.64, 32.30], $p < .001$) whereas the difference between more ($M = 73.08$, $SD = 18.99$) and less familiar metaphors ($M = 68.27$, $SD = 23.72$) did not significantly differ in the high cognitive flexibility group (95% CI [–12.33, 21.94], $p = .58$).

Discussion

The objective of the present study was to examine the relationship between cognitive and metaphor-related variables in preschoolers' psychological metaphor comprehension. The analyses provided partial support for the study's hypotheses. In contrast to previous theories (Asch & Nerlove, 1960; Piaget, 1926) and consistent with recent research involving preschool-aged children (e.g., Deamer, 2013; Pouscoulous & Tomasello, 2020; Rubio-Fernandez & Grassmann, 2016; Özçalışkan, 2005; Zhu & Gopnik, 2023), the present findings provide further evidence in favor of the hypothesis that preschool-aged children are capable of comprehending psychological metaphors when tested with tasks that have lower linguistic demands.

While recent research has demonstrated that children as young as three years old can understand physical metaphors and several studies indicate that elementary school children can understand psychological metaphors beyond chance levels (e.g., Lecce et al., 2019; Nippold et al., 1984), the present study is crucial in establishing that children as young as four or five years old can comprehend psychological metaphors. Moreover, we observed that children not only grasp the valence of the emotion (positive or negative) but also precisely identify the particular emotion that the metaphors communicate. Consistent with previous findings, our results show that metaphor comprehension increases with age (e.g., Di Paola et al., 2020; Rubio-Fernandez & Grassmann, 2016) and does not differ by gender (Pouscoulous & Perovic, 2023).

The data further confirmed the second hypothesis of the research, which shows the cognitive flexibility subcomponent of EF skills as one of the predictors of metaphor comprehension. To the best of our knowledge, only a limited number of studies have investigated the relationship between executive function skills and children's metaphor comprehension in preschool children (e.g., Deamer, 2013). As mentioned before, Deamer's (2013) work similarly found a relationship between EF skills and metaphor comprehension in 3-5-year-old children. However, unlike the present study, it focused on physical metaphors and examined inhibitory control as a subcomponent of EF. Despite the differences in methodology of the studies, the positive relationship observed in each suggests that executive function skills may be generally associated with metaphor comprehension. In another study, Carriedo et al. (2016) assessed the relationship between EF skills and metaphor comprehension in adolescents. EF skills, in Carriedo et al.’s (2016) study, were evaluated by employing multiple task batteries that assess inhibition, shifting, and updating. Contrary to the findings of the current study and Deamer’s (2013) study, no consistent effect of EF skills on metaphor comprehension was observed when utilizing distinct tasks in a sample of 11 and 15-year-olds. However, the discrepancy observed between the findings reported by Carriedo et al. (2016) and the present study may be because EF skills may have a more substantial impact on metaphor comprehension among younger children. To elaborate, since EF skills are still developing in children, they might have a stronger and more noticeable effect on their ability to comprehend metaphors. Therefore, the presently discussed findings are important for understanding the role of EF skills in a younger demographic, as it is likely to differ from the mechanisms identified in older children and adults.

While Carriedo et al. (2016) did not find a main effect of EF skills, their results indicated that “EF skills play a supplementary role when metaphor comprehension is highly demanding” (p.14). More precisely, they suggest that the contribution of EF skills to metaphor comprehension increases in the presence of special processing difficulties (e.g., limited semantic knowledge) or when metaphors are more difficult to comprehend (e.g., novel metaphors, absence of a context). The interaction analyses in the present data offer some evidence in favor of this hypothesis. Consistent with expectations, cognitive flexibility, in this study, demonstrated a (marginally) more pronounced influence on comprehending unfamiliar metaphors when compared with familiar metaphors. In general, the similarities and differences identified between the present study and prior investigations indicate that although EF skills are significantly related to understanding of metaphors, their effects might differ depending on the complexity level of the expression. Therefore, when analyzing the factors that contribute to the development of metaphor comprehension, it is vital to examine the interaction between individual differences and linguistic components of the metaphor itself, such as familiarity, contextual support, and abstractness.

The subsequent hypothesis, which posited a positive correlation between children's tendency to engage in pretend play and their comprehension of metaphors, was not supported by the data. Nevertheless, the current study provides valuable insight by distinguishing pretend play and metaphor comprehension as separate constructs rather than overlapping cognitive processes. This distinction is particularly important because previous research (e.g., Billow, 1981) often equated instances of pretense with metaphor production, whereas our findings suggest that these abilities may function independently. Alternatively, the lack of a relationship between these variables may be attributed to task characteristics. In the current study, children’s pretend play tendency was assessed using a test developed to assess a child's inclination for pretend play, whereas the metaphor task measured their comprehension ability. While it is thought that metaphor comprehension and pretend play operate on a similar underlying mechanism (i.e., dual representation) (Vosniadou, 1987), this relationship might have been more apparent if both constructs had been assessed using ability-based tasks rather than a tendency measure for pretend play. The absence of evidence to support the hypothesis could also be influenced by the characteristics of the sample. The sample exhibited a high tendency to participate in pretend play, with minimal scope for deviation. Observing the proposed effects might be attainable with a more diverse sample.

The findings did not support the hypothesis that children's complex language use is related to their metaphor comprehension. These results contradict existing work that views metaphor comprehension as a linguistic ability (Özçalışkan, 2005; Vosniadou, 1987). Based on these accounts, it is anticipated that different linguistic abilities will exert an influence on the comprehension of metaphors. For instance, a positive correlation was identified by Carriedo et al. (2016) between the ability to comprehend metaphors and verbal reasoning. Similarly, Pouscoulous and Perovic (2023) noted a positive association between proficiency in vocabulary and metaphor comprehension. The absence of a correlation in the present study may be a result of the task's attributes, as well. In the current task, participants' preferred language use is assessed, rather than their proficiency in complex language. A stronger link between a task that assesses linguistic skill and metaphor comprehension is more likely to be discovered. Furthermore, a closer link between abstract language proficiency and metaphor comprehension could also be expected since metaphors are a part of abstract language and could be more closely related to abstract language proficiency rather than general language proficiency. Moreover, in line with prior research (e.g., Lecce et al., 2019), the children who participated in the present study not only exhibited a reluctance to offer verbal explanations but also tended to prefer simple language, allowing for minimal space for interpretational variation. This observed reluctance and the simplicity of responses may be similar to the shortcomings of earlier metaphor comprehension studies (e.g. Asch & Nerlove, 1960) that interpreted the lack of response as an indication of non-comprehension. This suggests that the linguistic demands of the language complexity task of the current study might have masked the relationship between complex language skills and metaphor comprehension abilities.

In addition to exploring the effect of cognitive variables, the current study also attempted to reveal the effect of metaphor-related factors on metaphor understanding. The lack of agreement in the literature about the definition of simple and complex metaphors (Johnson & Pascual-Leone, 1989; Vosniadou et al., 1984; Waggoner & Palermo, 1989; Winner et al., 1976) has led us to the examination of several aspects that may influence the level of complexity in a metaphor. One such factor was the context in which the metaphor was presented. Previous research suggests that the complexity of contextual cues affects children's understanding of metaphors. For instance, Vosniadou et al. (1984) showed that in a comparable study, children's comprehension performance improved when they were provided with more predictable story endings, as opposed to those that were less predictable. However, our findings did not support this hypothesis as we did not find a relationship between the context and metaphor comprehension. This discrepancy may be attributed to the task characteristics which can be improved in follow-up studies. For instance, whether a context is easy or hard can be normed with an a priori pilot study.

The findings did not provide consistent support for the hypothesis proposing a better metaphor comprehension when a metaphor is more familiar. However, this study is important as it takes the familiarity of the metaphors into account while the recent research has mainly concentrated on children's understanding of novel metaphors (e.g., Deamer, 2013; Pouscoulous & Tomasello, 2020; Waggoner & Palermo, 1989). Despite the lack of a main effect, as reviewed above, the data suggest that levels of familiarity interact with the cognitive flexibility skill of children. Specifically, higher cognitive flexibility might have predominantly contributed to the comprehension of less familiar metaphors rather than more familiar metaphors.

The study also acknowledges certain limitations, particularly in the use of visual materials in assessing metaphor comprehension. As analyses revealed, not all of the metaphor comprehension stories worked in the intended direction. Performance on different stories was expected to differ by their complexity levels; nevertheless, for instance, the poor performance in the second story seems to be due to the visual material used in the study. The phrase “eli ayağına dolaştı” in the story set A has the literal translation of “she got caught up in her feet”. The psychological meaning of the metaphor suggests that the person is anxious or worried. Participants tended to choose the incorrect picture, which is the only picture where the girl’s hands are in a different position, standing with her hands on her hips rather than a neutral position with her hands down. This position might be interpreted as her arms being tangled in some way. Furthermore, it is also possible that the correct picture does not accurately portray an anxious person (see supplementary material for pictures of stories).

Another limitation of the images is that, although they were created based on previous research (Ekman, 1992), they were not normed according to the emotions they were meant to represent or the complexity of the context they were presented. This could be problematic, as the emotional valence and the complexity of the context might be interpreted subjectively. This highlights the challenge in designing effective metaphor comprehension tests and suggests a need for refining visual and textual elements in future research with a norming study.

Despite its limitations, the current study is significant and novel in several aspects. To our knowledge, this is the first study to assess preschool children's metaphor comprehension using a behavioral paradigm in a Turkish sample. The findings are crucial in demonstrating that preschool children possess a certain level of comprehension of psychological metaphors. Another strength of the study is its examination of both individual and metaphor-related variables, as well as their interactions, which is essential for a comprehensive understanding of the factors that contribute to metaphor comprehension. Further research could expand these findings by investigating different factors such as Theory of Mind (ToM), dual representation skills, and abstract language competence, which may differentially influence the development of metaphor comprehension across various developmental stages and metaphor types. For example, Tonini et al. (2023) found that better ToM skills improved psychological metaphor comprehension but not physical metaphor comprehension in early middle school children. Furthermore, the impact became non-significant for older middle school children. These findings suggest that ToM skills may play a crucial role in the early development and use of psychological metaphors, which warrants further research.

In conclusion, this study demonstrates that preschoolers possess the ability to understand psychological metaphors, with cognitive flexibility skills playing a significant role in this process. The results suggest that cognitive flexibility may facilitate metaphor comprehension, particularly for less familiar metaphors, while pretend play and language complexity did not show significant associations. These findings challenge earlier assumptions about young children's metaphor comprehension capabilities and highlight the need for further research with more refined methodologies considering both cognitive and metaphor-related factors. Identifying the cognitive factors influencing metaphor comprehension and their interaction with the metaphor-related factors in early childhood can help us understand how children may employ the "meaning-forming" function of metaphors to regulate their own emotions (Faranda, 2014; Karaırmak, 2015; Lapsekili & Yelboğa, 2014). Specifically, grasping metaphorical language, which often conveys emotional or abstract content, may help children make sense of their emotional experiences and develop strategies for emotion regulation. Consequently, this study provides a foundational basis for future research exploring the interplay between cognitive development, psychological metaphor comprehension, and emotional growth in early childhood.

References

Asch, S., & Nerlove, H. (1960). The development of double function terms in children: An exploratory study. In B. Kaplan & S. Wapner (Eds.), *Perspectives in Psychological Theory* (pp. 47–60). International Universities Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A cross-linguistic developmental study.* Erlbaum.

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

Blank, G. D. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity, 3*, 21–36. https://doi.org/10.1207/s15327868ms0301_2

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*(1), 193-216. https://doi.org/10.1037/0033-295X.112.1.193

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032-1053. https://doi.org/10.1111/1467-8624.00333

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), e0150289. https://doi.org/10.1371/journal.pone.0150289

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology, 33*(2), 218-236. https://doi.org/10.1111/bjdp.12075

Cometa, M. S., & Eson, M. E. (1978). Logical operations and metaphor interpretation: A Piagetian model. *Child Development, 49*(3), 649-659. https://doi.org/10.2307/1128232

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

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

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99*(3), 550–553. https://doi.org/10.1037/0033-295X.99.3.550

Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbol, 2*(4), 239-250. https://doi.org/10.1207/s15327868ms0204_2

Faranda, F. (2014). Working with images in psychotherapy: An embodied experience of play and metaphor. *Journal of Psychotherapy Integration, 24*(1), 65-77. https://doi.org/10.1037/a0035967

Fetterman, A. K., Bair, J. L., Werth, M., Landkammer, F., & Robinson, M. D. (2016). The scope and consequences of metaphoric thinking: Using individual differences in metaphor usage to understand how metaphor functions. *Journal of Personality and Social Psychology, 110*(3), 458-476. https://doi.org/10.1037/pspp0000067

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483-527. https://doi.org/10.1016/0885-2014(95)90024-1

Gardner, H. (1974). Metaphors and modalities: How children project polar adjectives onto diverse domains. *Child Development, 45*(1), 84-91. https://doi.org/10.2307/1127753

Garvey, C. (1990). *Play* (Vol. 27). Harvard University Press.

Gentner, D. (1977). Children's performance on a spatial analogies task. *Child Development, 48*(3), 1034-1039. https://doi.org/10.2307/1128356

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170. https://doi.org/10.1207/s15516709cog0702_3

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8*, 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*(1), 3-18. https://doi.org/10.1037/0033-295X.97.1.3

Johnson, J., & Pascual-Leone, J. (1989). Developmental levels of processing in metaphor interpretation. *Journal of Experimental Child Psychology, 48*(1), 1-31. https://doi.org/10.1016/0022-0965(89)90038-6

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language, 55*(1), 18-32. https://doi.org/10.1016/j.jml.2006.02.004

Karaırmak, Ö. (2015). Çocukla psikolojik danışmada metaforik süreç. *Turkish Psychological Counseling & Guidance Journal, 5*(43), 115-127. https://dergipark.org.tr/en/pub/tpdrd/issue/21463/230070

Kaugars, A. S., & Russ, S. W. (2009). Assessing preschool children's pretend play: Preliminary validation of the Affect in Play Scale-Preschool version. *Early Education and Development, 20*(5), 733-755. https://doi.org/10.1080/10409280802545388

Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development, 1*(1), 73-96. https://doi.org/10.1016/S0885-2014(86)80024-7

Kovecses, Z. (1988). *The language of love: The semantics of passion in conversational English.* Bucknell University Press.

Kovecses, Z. (2000). The scope of metaphor. *Topics in English Linguistics, 30,* 79-92.

Küçükerdoğan, P. (2024). *Metaphor Comprehension in Preschool Children* [unpublished master's thesis]. Boğaziçi University.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*(2), 195-208. https://doi.org/10.1207/s15516709cog0402_4

Lapsekili, N., & Yelboğa, Z. (2014). Psikiyatride metafor kullanımı. *Bilişsel Davranışçı Psikoterapi ve Araştırmalar Dergisi, 3,* 116-125.

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

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

Marriott, S. (2002). Red in tooth and claw? Images of nature in modern picture books. *Children's Literature in Education, 33*(3), 175–183. https://doi.org/10.1023/A:1019677931406

McCrindle, C. M. E., & Odendaal, J. S. J. (1994). Animals in books used for preschool children. *Anthrozoös, 7*(2), 135–146. https://doi.org/10.2752/089279394787001998

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Nelson, C. A., Morse, P. A., & Leavitt, L. A. (1979). Recognition of facial expressions by seven-month-old infants. *Child Development, 50*(4), 1239-1242. https://doi.org/10.2307/1129358

Nippold, M. A., Leonard, L. B., & Kail, R. (1984). Syntactic and conceptual factors in children's understanding of metaphors. *Journal of Speech, Language, and Hearing Research, 27*(2), 197-205. https://doi.org/10.1044/jshr.2702.197

Özçalışkan, Ş. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language, 32*(2), 291–318. https://doi.org/10.1017/S0305000905006884

Pearson, B. Z. (1990). The comprehension of metaphor by preschool children. *Journal of Child Language, 17*(1), 185-203. https://doi.org/10.1017/S0305000900013179

Piaget, J. (1926). *The language and thought of the child.* Harcourt, Brace.

Pomareda, C., Devine, R. T., & Apperly, I. A. (2024). Mindreading quality versus quantity: A theoretically and empirically motivated two-factor structure for individual differences in adults' mindreading. *PLoS One, 19*(6), e0305270. https://doi.org/10.1371/journal.pone.0305270

Pouscoulous, N. (2014). The 'elevator's buttocks': Metaphorical abilities in children. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 239–259). John Benjamins Publishing.

Pouscoulous, N. (2023). More than one path to pragmatics? Insights from children's grasp of implicit, figurative and ironical meaning. *Cognition, 240,* 105531. https://doi.org/10.1016/j.cognition.2023.105531

Pouscoulous, N., & Perovic, A. (2023). Linguistic skills and socioeconomic status: Two oft forgotten factors in child metaphor comprehension. *Children, 10*(12), 1847. https://doi.org/10.3390/children10121847

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.pragma.2019.05.021

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rubio-Fernández, P., & Grassmann, S. (2016). Metaphors as second labels: Difficult for preschool children? *Journal of Psycholinguistic Research, 45*, 931-944. https://doi.org/10.1007/s10936-015-9386-y

Smith, J. W. A. (1976). Children's comprehension of metaphor: a Piagetian interpretation. *Language and Speech, 19*(3), 236-243. https://doi.org/10.1177/002383097601900304

Spiegel, J. A., Goodrich, J. M., Morris, B. M., Osborne, C. M., & Lonigan, C. J. (2021). Relations between executive functions and academic outcomes in elementary school children: A meta-analysis. *Psychological Bulletin, 147*(4), 329–351. https://doi.org/10.1037/bul0000322

Taggart, J., Eisen, S., & Lillard, A. S. (2019). The current landscape of US children's television: Violent, prosocial, educational, and fantastical content. *Journal of Children and Media, 13*(3), 276-294. https://doi.org/10.1080/17482798.2019.1605916

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

Turner, N. E., & Katz, A. N. (1997). The availability of conventional and of literal meaning during the comprehension of proverbs. *Pragmatics and Cognition, 5*, 199–233. https://doi.org/10.1075/pc.5.2.02tur

Vosniadou, S. (1987). Children and metaphors. *Child Development, 58*(3), 870-885. https://doi.org/10.2307/1130223

Vosniadou, S., Ortony, A., Reynolds, R. E., & Wilson, P. T. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development, 55*(4), 1588-1606. https://doi.org/10.2307/1130028

Waggoner, J. E., Messe, M. J., & Palermo, D. S. (1985). Grasping the meaning of metaphor: Story recall and comprehension. *Child Development, 56*(5), 1156-1166. https://doi.org/10.2307/1130230

Waggoner, J. E., & Palermo, D. S. (1989). Betty is a bouncing bubble: Children's comprehension of emotion-descriptive metaphors. *Developmental Psychology, 25* (1), 152-163. https://doi.org/10.1037/0012-1649.25.1.152

Winner, E. (1979). New names for old emergence of metaphoric language. *Child Language, 6,* 469-491. https://doi.org/10.1017/S030500090000249X

Winner, E., McCarthy, M., & Gardner, H. (1980). The ontogenesis of metaphor. In R.P. Honeck & R. R. Hoffman (Eds.), *Cognition and figurative language* (pp. 335-360). Erlbaum.

Winner, E., McCarthy, M., Kleinman, S., & Gardner, H. (1979). First metaphors. In D. Wolf (Ed.), *Early symbolization: New directions for child development* (pp. 29-41). Jossey-Bass.

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

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297-301. https://doi.org/10.1038/nprot.2006.46

Zhu, R., & Gopnik, A. (2023). Preschoolers and adults learn from novel metaphors. *Psychological Science, 34*(6), 696-704. https://doi.org/10.1177/09567976231165267

Data, Code, and Materials Availability Statement

The authors confirm that the data, code and materials supporting the findings of this study are available within the article and its supplementary materials in a permanent repository: Metaphor Comprehension in Preschool Children at https://osf.io/rukwf/.

Ethics Statement

The research was conducted following the human ethics guidelines and was approved (2023-38T, Date: 22.12.2023) by the human ethics committee of the Boğaziçi University: Institutional Review Board for Research with Human Subjects.

Acknowledgements

The current project is part of the first author's thesis (Küçükerdoğan, 2024), which is financially supported by the TÜBİTAK 2210-A program. We would also like to thank Dicle Gül for illustrating the metaphor stories used in the study.

Authorship and Contributorship statement

Pelin Küçükerdoğan conceived of the study, designed the study, collected data, and wrote the first draft of the manuscript. **Deniz Tahiroglu** contributed to the design of the study and revised the manuscript. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

License

'Broccoli is candy': the role of metaphors in children's persuasive communication

Chiara Pompei
Laboratory of Neurolinguistics and Experimental Pragmatics (NEPLab), Department of Humanities and Life Sciences, University School for Advanced Studies IUSS, Pavia, Italy

Serena Lecce
Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

Paola Del Sette
IRCCS Ospedale Policlinico San Martino, Genova, Italy

Elena Didoni
Luca Bischetti
Valentina Bambini
Laboratory of Neurolinguistics and Experimental Pragmatics (NEPLab), Department of Humanities and Life Sciences, University School for Advanced Studies IUSS, Pavia, Italy

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.

Keywords: language acquisition; persuasive skills; communication; figurative language; metaphor production

Corresponding author(s): Chiara Pompei, Laboratory of Neurolinguistics and Experimental Pragmatics (NEPLab), Department of Humanities and Life Sciences, University School for Advanced Studies IUSS Pavia, Piazza della Vittoria 15, 27100 Pavia, Italy. Email: chiara.pompei@iusspavia.it

ORCID ID(s): Chiara Pompei: https://orcid.org/0000-0002-4381-3947; Serena Lecce: https://orcid.org/0000-0002-2745-3574; Paola Del Sette: https://orcid.org/0000-0001-6696-2139; Elena Didoni:

https://orcid.org/0009-0007-8970-538X; Luca Bischetti: https://orcid.org/0000-0001-8875-1092; Valentina Bambini: https://orcid.org/0000-0001-5770-228X

Citation: Pompei, C., Lecce, S., Del Sette, P., Didoni, E., Bischetti, L. & Bambini, V. (2025). 'Broccoli is candy': the role of metaphors in children's persuasive communication. *Language Development Research, 5*(2), 204–236. http://doi.org/10.34842/ldr2025-836

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 also depends 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, 2002a; 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 accounts in the pragmatics of language, metaphors are loose uses of words (Carston, 2010), 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, 2019; 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 & Stamenkovíc, 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 & Stamenkovíc, 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).

The *metaphorical advantage* in persuasive communication was confirmed by a meta-analysis of 29 studies comparing literal versus metaphorical uses in persuasive statements, which showed that metaphors ensured attitude changes to a greater extent. This advantage over literal language was explained in terms of a facilitatory effect of metaphors in integrating the current message (i.e., persuasive statements) into the interlocutor’s prior knowledge (Sopory & Dillard, 2002b). Consistently, the ability to master metaphors was shown to be key in promoting the understanding of scientific concepts (i.e., climate change-related) in children, while modulating psychological attitudes toward environmental issues and proactive behavior (Pompei et al., 2024). 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, attitudes, and behaviors toward the topic. Specifically, metaphorical skills seem to provide a flexible conceptual platform through which ideas can be broadened, narrowed, and creatively associated to strengthen persuasive effects. However, despite the numerous connections between persuasive strategies and metaphors highlighted above, a clear understanding of whether and how persuasive and metaphorical abilities are related to one another during development is still missing.

The study of persuasion is a relatively underexplored area within the field of developmental social psychology. Investigations on children's 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 et al., 2018; Lonigro et al., 2017). Studies have also shown that persuasive skills mature together with other skills, including Theory of Mind (ToM), 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 et al., 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 et al. (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 et al., 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.

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.

Method

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, having hearing deficits, or 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 meta-analysis 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.

Design and 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.

Measures

Vocabulary. The Italian version of the Peabody Picture Vocabulary Test - Revised (PPVT-R; Stella et al., 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 test interruption. The total score consists of the number of correct answers (score range: 0-175).

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 of difficulty (2-word; 3-word; 4-word; 5-word trials) are administered; failure in both items corresponds to the interruption of the test. The total score consists of the number of correctly achieved levels (score range: 0–8).

Metaphor Comprehension. To assess metaphor comprehension, we used the newly developed multiple-choice version of the Physical and Mental Metaphor task (Lecce et al., in prep.), originally developed 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 $.37 < r\text{s} < .57$). 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 an appropriate interpretation or incorrect (0) for literal and unrelated interpretation (score range: 0-10).

Metaphor Production. To assess metaphor production abilities, we adapted the elicitation task by Cortés et al. (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 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, children 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).

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.1. for the schematic representation of the task).

Following the recommendations proposed by Peterson et al. (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 the 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], $p < .001$.

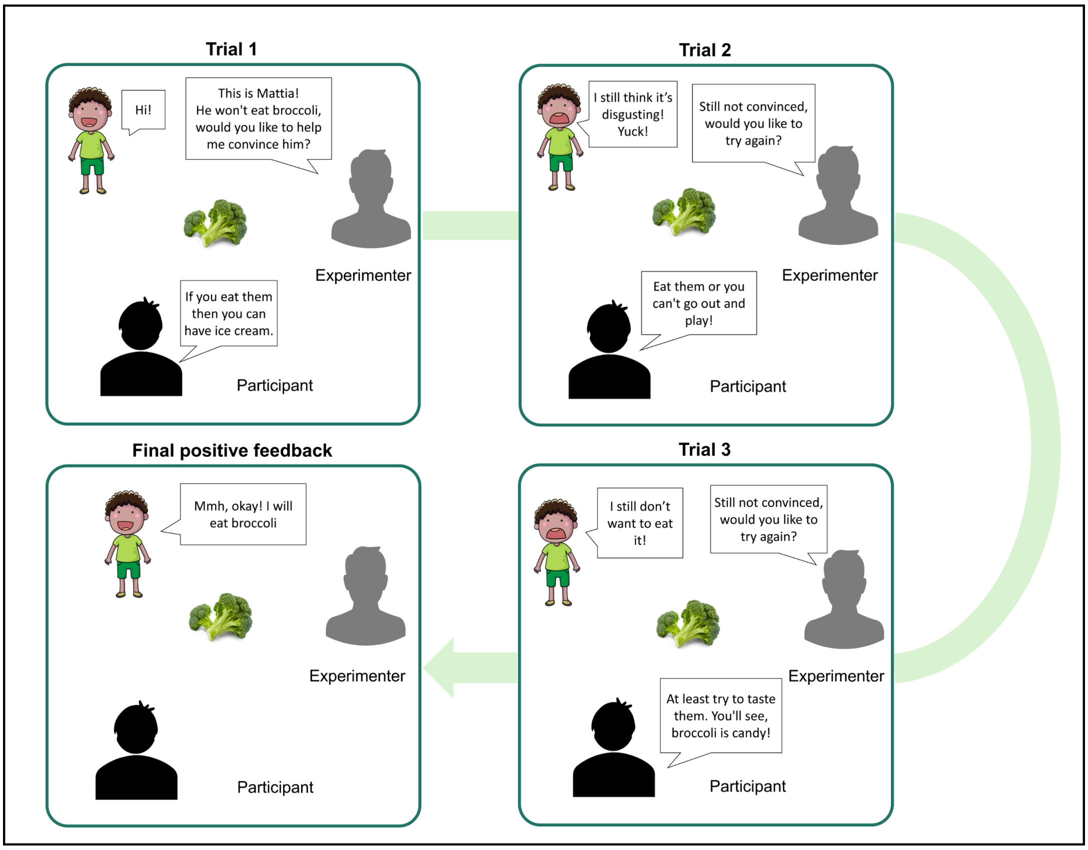


Figure 1. Schematic representation of the Persuasion task adapted from Peterson and colleagues (2018). Examples are offered in English translation from Italian.

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 materials' 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 ~ 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, whose formula was: Metaphor count ~ 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 Odds 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 et al., 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 et al., 2017), and the *emmeans* (v. 1.10.6-090001; Lenth, 2024) packages.

Results

Children's performance in each variable, across Age Groups, is shown in Table 1 and graphically represented in Figure 2.

Table 1. ***Descriptive statistics of children's performance in each age group.***

Measures	Range of possible scores	Full sample (N = 167) *M* (*SD*)	4 years (*n* = 52) *M* (*SD*)	5 years (*n* = 76) *M* (*SD*)	6 years (*n* = 39) *M* (*SD*)
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 Psychology 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

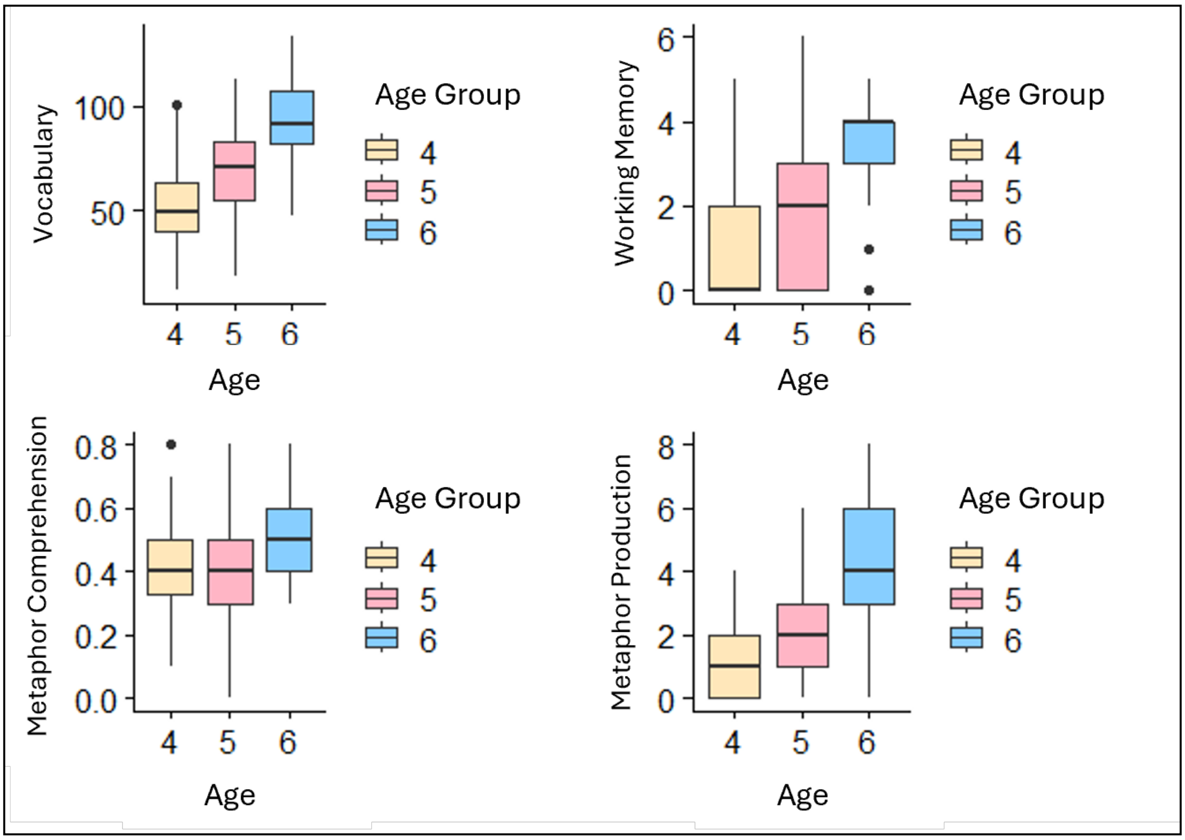


Figure 2. Box plot of the linguistic, cognitive, and socio-cognitive measures for each age group. The central mark indicates the median, the bottom edge the 25th percentile and the top edge the 75th percentile of data. The whiskers indicate 1.5 times the interquartile range.

Pearson's correlation coefficients are plotted in the correlogram in Figure 3. 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.

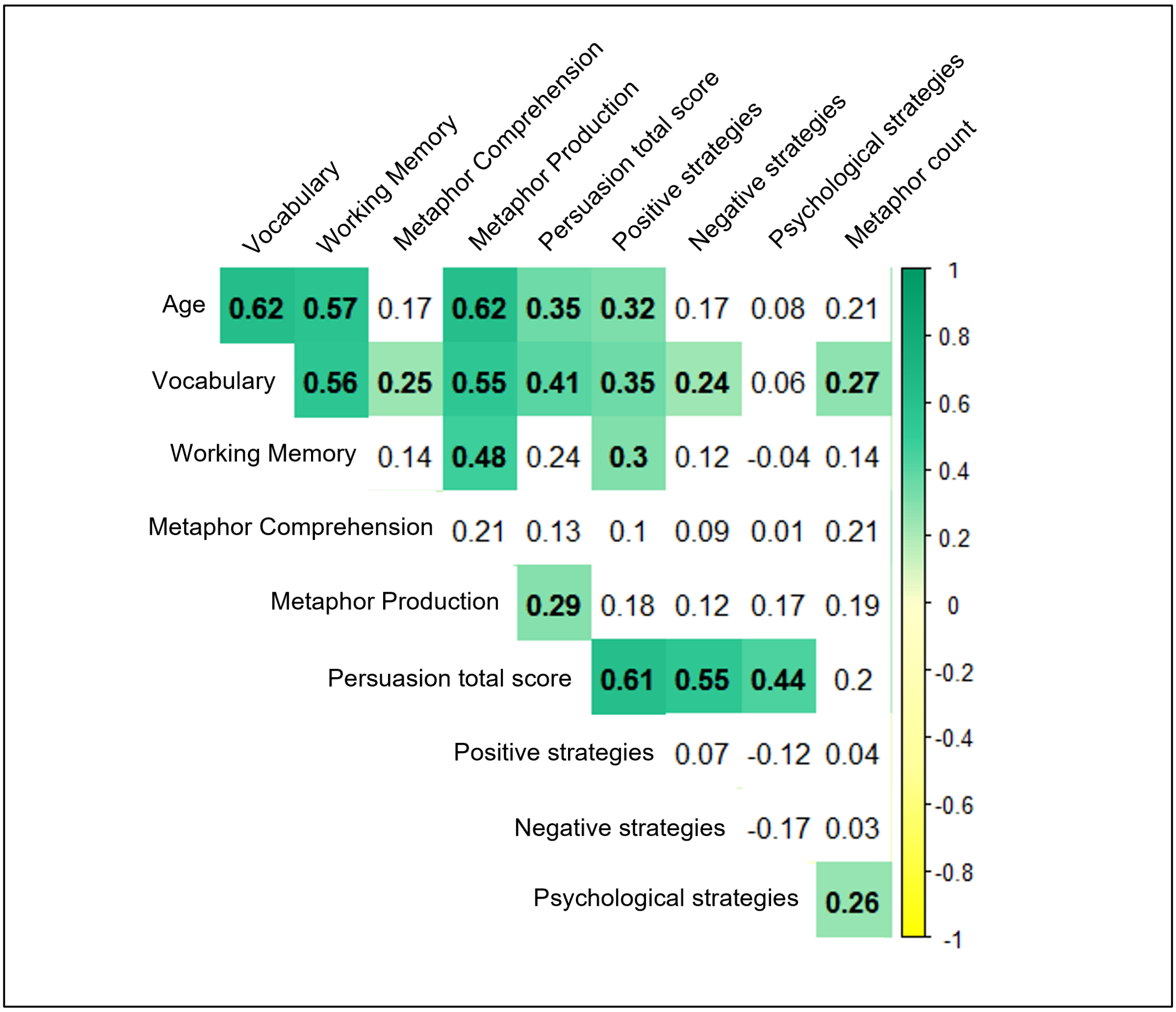


Figure 3. Correlogram between persuasion scores, linguistic and pragmatic variables. The plot shows correlations between Persuasion scores and Vocabulary, Working Memory and Metaphor Comprehension and Production, and Metaphor count. Positive correlations are displayed in green and negative correlations in yellow. The color intensity is proportional to the magnitude of the correlation coefficients. White cells indicate associations at *p*-value > .05. Age was transformed in months.

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.

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 4a). When inspecting the pairwise differences between persuasion strategies within age groups, we confirmed that 5-year-olds used to a greater extent positive over negative ($\Delta\beta = 0.39$, $\mathrm{z} = 2.79$, $p = .010$) and psychological over positive strategies ($\Delta\beta = -0.30$, $z = -2.23$, $p = .049$). Differently, comparisons in other age groups did not reach statistical significance (4: $\Delta\beta < 0.42$, $p > .096$; 6: $\Delta\beta < 0.46$, $p > .092$).

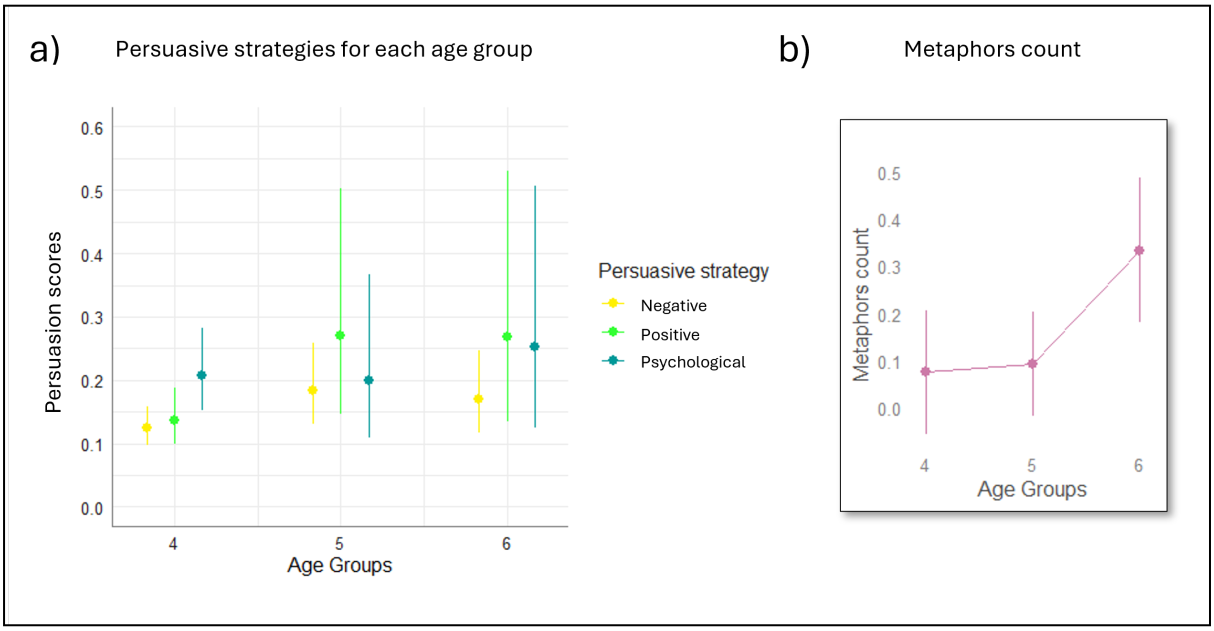


Figure 4. Development of persuasive strategies and focus on the use of metaphors. Panel a) depicts model estimates of the use of different persuasive strategies across Age Groups. Panel b) displays the use of metaphors across different Age Groups. In both panels, error bars represent the standard deviation of model estimates.

Regarding the scaled continuous predictors, the model showed a main effect of Vocabulary (*OR* = 1.21, *CI* = [1.06, 1.37], *z* = 2.88, *p* = .004): greater vocabulary skills were associated with the use of more persuasive arguments, regardless of the type of strategy (see Figure 5a). The model also highlighted a significant Working Memory by Type (Psychological vs. Positive) interaction (*OR* = 0.69, *CI* = [0.55, 0.86], *z* = –3.23, *p* = .001): the post-hoc analysis identified that higher levels of working memory were associated with a higher use of positive strategies (*β* = 0.10, *CI* = [0.00, 0.19], *z* = 2.01, *p* = .045) and a lower use of psychological arguments (*β* = –0.12, *CI* = [–0.22, –0.01], *z* = –2.24, *p* = .025; see Figure 5b). In addition, a significant Metaphor Production by Type (Psychological vs. Positive arguments: *OR* = 1.39, *CI* = [1.12, 1.73], *z* = 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 (*β* = 0.12, *CI* = [0.03, 0.20], *z* = 2.57, *p* = .010; see Figure 5c). No effect of Metaphor Comprehension on Persuasion scores was found (see Figure 5d).[1]

---

[1] We acknowledge the debate about the treatment of the age variable. To allow for 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

Table 2. ***Output of the generalized linear mixed-effects model with the persuasion scores as the dependent variable.***

Fixed effects	*Incidence Rate Ratios*	*CI*	*z-value*	*P value*
(Intercept)	0.19	0.15 – 0.25	−13.20	**< .001**
Age 5 vs. 4	1.42	1.10 – 1.83	2.68	**.007**
Age 6 vs. 5	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 5 vs. 4 × Positive vs Negative	1.35	0.76 – 2.37	1.03	.303
Age 6 vs. 5 × Positive vs Negative	1.08	0.64 – 1.80	0.28	.783
Age 5 vs. 4 × Psychological vs Positive	0.49	0.29 – 0.81	−2.79	**.005**
Age 6 vs. 5 × 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**

Random effects	*Variance*	*SD*
Intercept$_{\text{Subject}}$	0.10	0.32
Intercept$_{\text{Trial}}$	0.01	0.10
Intercept$_{\text{Item}}$	0.02	0.14
ICC$_{\text{SubjectTrialItem}}$	0.07	

Model fit	*Marginal*	*Conditional*
R^2	.094	.159

---

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.

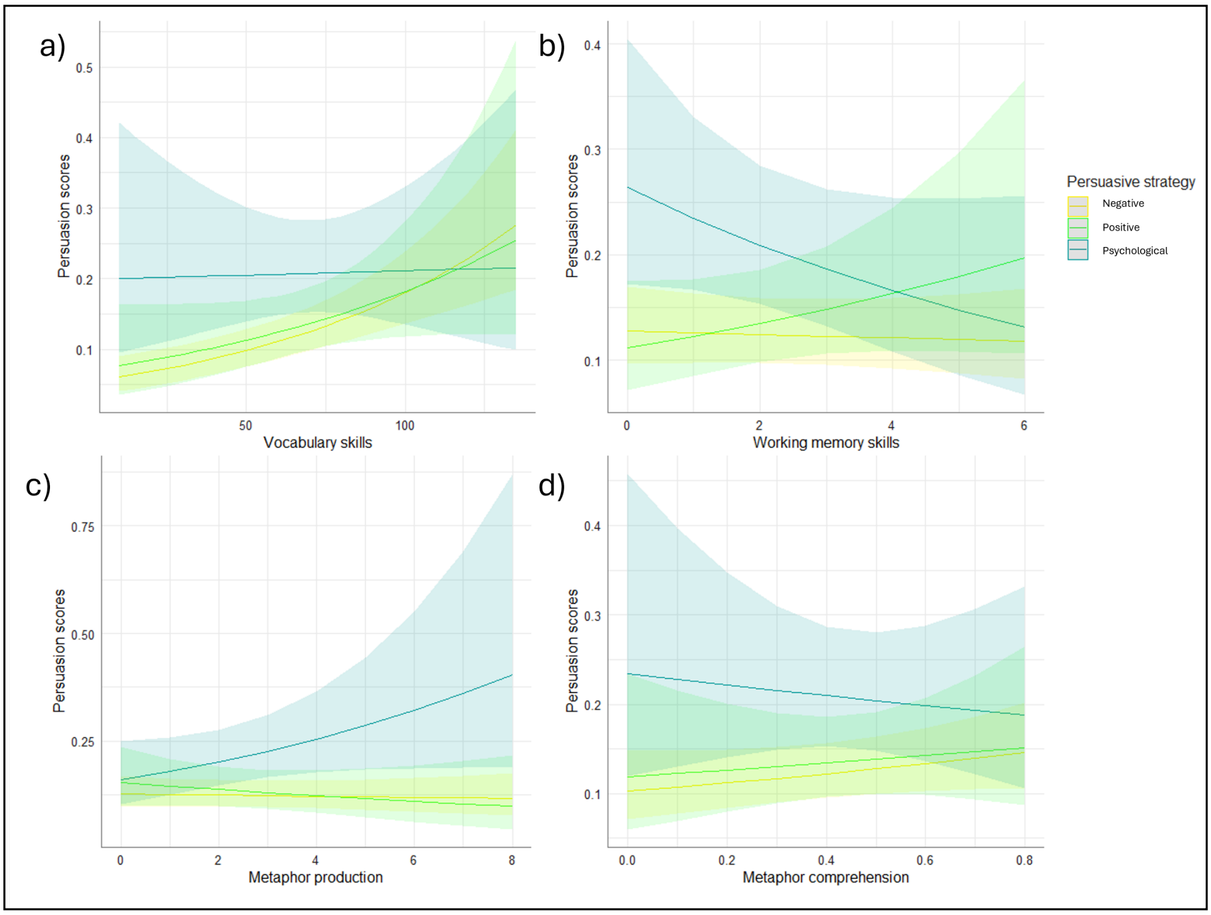


Figure 5. Association between persuasive strategies and covariates. The plot reports the relationships between persuasive strategies (Negative, Positive, Psychological) and Vocabulary (Panel a), Working Memory (Panel b), Metaphor production (Panel c), and Metaphor comprehension (Panel d).

The model on Metaphor count (see Table 3) showed a significant effect of Age Group for the 6 vs. 5 comparison showing that children's use of metaphors increased between 5 and 6 years of age ($\beta = 0.24$, $CI = [0.05, 0.43]$, $t = 2.53$, $p = .012$; see Figure 4b).

Table 3. ***Output of the linear model with the metaphor count as the dependent variable.***

Predictors	*Estimates*	*CI*	*z-value*	*p-value*
(Intercept)	0.17	0.09 – 0.24	4.31	**< .001**
5 vs. 4	0.02	–0.15 – 0.19	0.19	.848
6 vs. 5	0.24	0.05 – 0.43	2.53	**.012**
Model Fit	*Marginal*	*Marginal*		
R^2	.046	.034		

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 of 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). This suggests 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 their 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 of metaphor production 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 & Stamenkovíc, 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). In addition, 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 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 with their 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 expressive metaphorical 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 may increase the level of intimacy between interlocutors, as Bowes and Katz showed (2015), and these positive effects on social relationships are already manifest 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 the 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 experiences 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), fostering 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 to 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 persuasive strategies or other types. 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, young children are able to perform certain metaphor tasks (Pouscoulous & Tomasello, 2020), but their ability to fully articulate metaphorical meaning is still fragile until 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 5), 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., 2025; 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 the child’s actual desire. 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, emerging specifically for psychological strategies, we cannot rule out the possibility that this association is due to the shared expressive modality of both tasks. Further studies using other non-metaphorical productive tasks could help determine whether this effect is driven solely by modality compatibility or if a true association exists 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.

References

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

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

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

Bischetti, L., Frau, F., & Bambini, V. (2024). Neuropragmatics. In M. J. Ball, N. Müller, & E. Spencer (Eds.), *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

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

Bülow-Møller, A. M. (2005). Persuasion in business negotiations. In H. Halmari & T. Virtanen (Eds.), *Persuasion Across genres: A linguistic approach* (pp. 27–58). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.130.05bue

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

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

Carston, R. (2010). Lexical pragmatics, ad hoc concepts and metaphor: A Relevance Theory perspective. *Italian Journal of Linguistics, 22*(1), 153–180.

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

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

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

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. (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

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., 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

Dillard, J. (2010). Persuasion. In C. R. Berger, E. M. Roloff, & R. D. Roskos-Ewoldsen (Eds.), *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

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.

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

Falkum, I. L. (2019). Pragmatic development: Learning to use language to communicate. In J. S. Horst & J. Von Koss Torkildsen (Eds.), *International Handbook of Language Acquisition* (Issue October, pp. 234–260). Routledge. https://doi.org/10.4324/9781315110622-13

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

Gass, R. H., & Seiter, J. S. (2022). *Persuasion: Social influence and compliance gaining (7th ed.)*. Routledge. https://doi.org/10.4324/9781003081388

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

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

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

Holyoak, K. J., & Stamenkovíc, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin, 144*(6), 641–671. https://doi.org/10.1037/bul0000145

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

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

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., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

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

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., Pompei, C., Del Sette, P., Canal, P., & Bambini, V. (In preparation). *Climbing literal and physical steps: Developmental trajectories of metaphor understanding in early childhood.*

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

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

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

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

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

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

Montanari, F. (1996). Introduzione. In M. Dorati (Ed.), *Aristotele, Retorica.* Mondadori.

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

Nippold, M. A., Ward-Lonergan, 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)

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

O'Keefe, D. J. (2007). Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. *Communication Methods and Measures, 1*(4), 291–299. https://doi.org/10.1080/19312450701641375

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

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. (2025). 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*, *51*(3), 460–477. https://doi.org/10.1037/xlm0001381

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. Paper presented at *Crossing The Borders: The Interplay of Language, Cognition, and the Brain in Early Human Development*. 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. Paper presented at *Communicating Climate Hope: Countering Eco-Anxiety and Climate Doomism in Research and Practice*. https://climatehope.sites.olt.ubc.ca/attendees-only-area/book-of-abstracts/

Pouscoulous, N. (2014). The elevator's buttocks: Metaphorical abilities in children. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 239–259). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.10.14pou

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.pragma.2019.05.021

R Core Team. (2023). *RStudio: Integrated Development Environment for R* (4.3.2).

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* (pp. 171–191). Springer. https://doi.org/10.1007/978-3-030-56437-7_12

Sbisà, M. (2013). Locution, illocution, perlocution. In M. Sbisà & K. Turner (Eds.), *Pragmatics of speech actions* (pp. 25–75). Mouton de Gruyter. https://doi.org/10.1515/9783110214383.25

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

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

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. https://doi.org/10.4135/9781412976046.n21

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. (1986). *Relevance: Communication and cognition*. Harvard University Press.

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

Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). *Il Peabody Test di Vocabolario Recettivo.* Omega.

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

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

Van Patten, J. K. (2013). Metaphors and persuasion. *South Dakota Law Review, 58*(2), 295–318. https://doi.org/10.2307/1130223

Vosniadou, S. (1987). Children and metaphors. *Child Development, 58*(3), 870–885.

Wilson, S. R., Guntzviller, L. M., & Munz, E. A. (2012). Persuasion and families. In A. L. Vangelisti (Ed.), *The Routledge Handbook of Family Communication* (pp. 471–496). Routledge. https://doi.org/10.4324/9780203848166.ch22

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

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

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

Zhu, R., & Gopnik, A. (2023). Preschoolers and adults learn from novel metaphors. *Psychological Science, 34*(6), 696–704. https://doi.org/10.1177/09567976231165267

Data, Code, and Materials Availability Statement

The dataset analyzed in the study and the code for the analysis are available in the Open Science Framework repository at the link: https://osf.io/85zu4/?view_only=90088190e23941e1806d91cbb7a6b041

Materials cleared for open access (the metaphor production task and the persuasion task) are available in the Open Science Framework repository at the link: https://osf.io/85zu4/?view_only=90088190e23941e1806d91cbb7a6b041.

Content subject to commercial licensing agreements (e.g., the PPVT-R and the word recall task) is not included among the publicly accessible resources. For the metaphor comprehension task, only the textual components of the items are provided, as licensing considerations are currently under evaluation. In the latter case, complete task materials can be obtained upon reasonable request from the corresponding author.

Ethics Statement

Ethics approval was obtained from the Local Ethical Committee of the Department of Brain and Behavioral Sciences of the University of Pavia (protocol n. 029/2019), following the principles of the Declaration of Helsinki. Both parents of each participant gave informed written consent before taking part in the study.

Authorship and Contributorship Statement

Chiara Pompei: Investigation, Data Curation, Methodology, Formal Analysis, Visualization, Writing – original draft; **Serena Lecce**: Conceptualization, Methodology, Supervision, Writing – review and editing; **Paola Del Sette**: Investigation, Data curation, Writing – review and editing; **Elena Didoni**: Data Curation, Writing – review and editing; **Luca Bischetti**: Formal Analysis, Writing – review and editing; **Valentina Bambini**: Conceptualization, Methodology, Supervision, Writing – review and editing.

Acknowledgements

This work received support in part from the Italian Ministry of Research under the PRIN 2022 programme, 'The Fragility of Metaphors (FraMe): learning, losing, and how to train them' (project number: 2022289RNA) and from the European Research Council under the EU's Horizon Europe program, ERC Consolidator Grant 'PROcessing MEtaphors: Neurochronometry, Acquisition and Decay, PROMENADE' (101045733). The content of this article is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.

License