CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

That kid is a grasshopper! Metaphor development from 3 to 9 years of age

Isabel Martín-González[1] (iD), Camilo R. Ronderos[2], Elena Castroviejo[1], Kristen Schroeder[2], Ingrid Lossius-Falkum[2] and Agustín Vicente[1,3]

[1]Department of Linguistics and Basque Studies, Micaela Portilla research centre, University of the Basque Country – UPV/EHU (Vitoria-Gasteiz, Spain); [2]University of Oslo, Department of Philosophy, Classics, History of Art and Ideas (Oslo, Norway) and [3]Ikerbasque, Basque Foundation for Science
Corresponding author: Isabel Martín-González; Email: isabel.martin@ehu.eus

Abstract

Two major trends on children's skills to comprehend metaphors have governed the literature on the subject: the *literal stage* hypothesis vs. the *early birds* hypothesis (Falkum, 2022). We aim to contribute to this debate by testing children's capability to comprehend novel metaphors ('X is a Y') in Spanish with a child-friendly, picture selection task, while also tracking their gaze. Further, given recent findings on the development of metonymy comprehension suggesting a U-shaped developmental curve for this phenomenon (Köder & Falkum, 2020), we aimed to determine the shape of the developmental trajectory of novel metaphor comprehension, and to explore how both types of data (picture selection and gaze behavior) relate to each other. Our results suggest a linear developmental trajectory with 6-year-olds significantly succeeding in picture selection and consistently looking at the metaphorical target even after question onset.

Keywords: metaphor development; eye-tracking; figurative language; experimental pragmatics

Introduction

For the past decades there has been extensive research on the development of metaphorical abilities (e.g., Di Paola et al., 2020; Pouscoulous, 2011; Vosniadou, 1989; Vosniadou et al., 1984; Winner et al., 1976). At the core of this research is the question of the earliest age at which children have the necessary skills to appropriately interpret a metaphor. While older studies claimed that this ability developed relatively late (10 to 12 years old) (Asch & Nerlove, 1960; Cometa & Eson, 1978; Smith, 1976; Winner et al., 1976), recent studies have challenged this idea and shown that even 3-year-olds are able to understand (certain types of) metaphors (Di Paola et al., 2020; Pouscoulous &

Tomasello, 2020).[1] Pouscoulous (2014) argues that what brought about the results in the earlier literature was a mix of task-induced complexity (e.g., experimental paradigms that were not child-friendly), uneven test materials (e.g., mixing idioms and metaphors or testing both novel and conventional metaphors together), uneven conceptual knowledge requirements (e.g., testing both psychological and perceptual metaphors together), and a lack of adequate context. This reasoning led Pouscoulous and Tomasello (2020) to show that simple perceptual metaphors such as *the tower with the hat* (said of a tower with a pointy roof) are mostly interpreted in the correct metaphorical way by 3-year-olds when using an age-appropriate task.

Despite this critical finding, several questions remain unanswered. For example, can the non-perceptual novel metaphors typically investigated in the older literature such as *the prison guard is a rock* (so-called predicative nominal metaphors of the form 'X is a Y') also be grasped at a younger age when using child-appropriate tasks and materials, or is metaphor interpretation at a younger age bound to simple perceptual metaphors? Further, beyond investigating the earliest point of success, what does the developmental trajectory of metaphor interpretation look like? Few studies have systematically tested an age range sufficient to probe a developmental trajectory of metaphor. This question is of renewed importance considering recent findings on the interpretation of novel metonymy (e.g., *The hat* is sitting in the corner) suggesting a non-lineal (U-shaped) developmental trajectory, in which 3-year-olds apparently outperform more literally-oriented 4- and 5-year-olds when tested on this particular type of figurative language use (Köder & Falkum, 2020). Finally, beyond selecting the correct (metaphorical) image at the end of a trial, how does the online processing of metaphor interpretation develop with age, and what is the relationship between processing and the explicit behavioral response? The importance of this question also relates to the findings of Köder and Falkum (2020) on the comprehension of metonymy, where they reported a disassociation in the pattern of results for an online processing measure (gaze behavior time-locked to an unfolding utterance) and the explicit behavioral response at the end of a trial (i.e., picture selection).[2]

We believe that answering these three questions will bring us closer to understanding not just *when* children are able to successfully deal with metaphorical meaning, but also *which* types of metaphors (or metaphorical mappings) are the easiest for them and *how*

---

[1]It is customary to talk about metaphor understanding or metaphor comprehension in relation with methodologies as different as multiple choice, act-out, or verbal explanation. In what follows, we will refer to the kind of metaphor processing that our methodology tests as "metaphor comprehension", although we are aware that the methodology cannot determine whether participants have what we can call "full comprehension" of the metaphors. By "full comprehension of the metaphor" we refer to grasping the metaphor in a way that it can be explained, re-used (e.g., "oh! so that child is also a grasshopper"), and used in reasoning (e.g., "if this kid is a grasshopper, that one there (who jumps even more) is a bigger grasshopper"). As said, we cannot know if the children in the experiment reach this level of comprehension. However, we consider that the children who look longer at the metaphorical picture and/or choose the right picture, display some level of comprehension. We had also considered talking about "grasping" or "apprehending" the metaphor, but using this terminology would run counter to the predominant terminological choice in similar studies, and might create confusion.

[2]A dissociation between gaze behavior and picture selection may suggest a tacit understanding, or at least sensitivity to, the contextually appropriate meaning of the referring expression despite a subsequent erroneous picture selection choice. Incorporating gaze behavior in study design may therefore provide a richer assessment of children's developing interpretation abilities than methods that focus on outcomes alone (i.e., indirect methods).

their interpretation changes ontogenetically. The goal of the current study is thus to tackle these three issues. We developed a child-friendly experimental task to investigate the development of comprehending novel predicative nominal metaphors in Spanish that go beyond perceptual similarities, such as the one illustrated in (1) below.

(1) Grasshoppers jump a lot. That kid is a grasshopper.[3]

By using an eye-tracking, Visual World paradigm (Cooper, 1974; Huettig et al., 2011; Tanenhaus et al., 1995), we were able to capture both children's (ages 3-9) online metaphor processing as well as their explicit response in a picture selection task. The current work therefore builds a bridge between the older and the more recent studies in metaphor research and provides further avenues of investigation going forward. Before presenting our experiment, we first briefly review the existing literature on the development of metaphor interpretation.

Metaphor interpretation: a developmental overview

The literature on the development of metaphor interpretation is made up of vast variability in experimental tasks and types of metaphors studied. The wide range of results can therefore be challenging to interpret. A recent review of the literature (Falkum, 2022) has done so by distinguishing between (at least) two positions with regards to how young children (younger than 10) interpret novel metaphors: the 'literal-first' and the 'early-bird' positions.

The literal-first view claims that children have a tendency to interpret metaphors literally, and that it is not until a later age (typically 10 or 12) that they are capable of correctly understanding them. This view is based on various empirical findings (e.g., Asch & Nerlove, 1960; Cometa & Eson, 1978; Smith, 1976; Vosniadou et al., 1984; Winner et al., 1976). For example, in Vosniadou et al. (1984), an action metaphor like *the little girl was a bird flying to her nest* was claimed to be interpreted literally as a girl pretending to be flying. Winner et al. (1976) asked children to explain psychological metaphors (such as *the prison guard was a hard rock*) and reported that young children understood them literally (e.g., as describing a fantasy world in which rocks could be prison guards). These and further similar conclusions from the above cited studies are for the most part based on explanation/paraphrasing methodologies. This methodology has since been criticized because it is arguably a task which recruits skills above and beyond those that might be strictly necessary for successful metaphor interpretation (see Pouscoulous, 2011). This is particularly true of tasks involving verbal reasoning, which young children typically struggle with. To this effect, Vosniadou (1987) cites several studies in which children younger than 10 show some grasp of metaphorical language (usually based on perceptual or physical similarity) when tested with multiple choice tasks (Nippold et al., 1984) or enactment tasks (Vosniadou et al., 1984) that do not rely heavily on verbal reasoning abilities. Further, children seem to even be able to explicate the meaning of metaphors when these are based on physical or perceptual similarity (Malgady, 1977). This latter

---

[3]Perceptual metaphors are usually opposed to *psychological* metaphors (e.g., *That lawyer is a shark*, meaning, inter alia, that the lawyer is aggressive). The metaphors we tested are not perceptual, but they are not psychological either. They might be called *behavioral*, as the analogy between source and target is based on some typical behavior of the source that is also displayed by the target.

finding is in line with more recent studies suggesting that psychological metaphors such as *Dad was a volcano* seem to be harder to interpret relative to perceptual metaphors (Lecce et al., 2019).

In contrast, the more recent 'early-birds' view states that young children (even 3-year-olds) are capable of apprehending metaphors (Deamer, 2013; Di Paola et al., 2020; Pouscoulous & Tomasello, 2020). Pouscoulous and Tomasello (2020) showed that in a simple object selection task, 3-year-olds consistently preferred the metaphorical interpretation of *the tower with the hat* (a tower with a red pointy roof) over the distractor (a tower with a red balcony). Importantly, Pouscoulous and Tomasello made sure that children had the relevant world knowledge skills to interpret these metaphors as such. This design did not include a literal option, and so does not tap into a possible tendency for literal interpretation in young children when facing metaphors. It shows, alongside with the other cited studies, that very young children already possess some of the inferential skills necessary to interpret metaphors.

A critical feature of this study as well as that by Di Paola et al. (2020) is the type of metaphors used, which are based on simple perceptual features (e.g., *the tower with the hat*, *the car with the backpack*). These are conceptually similar to the type of overextensions and loose uses of language characteristic of toddlers and young children (such as a toddler referring to every old man they meet as *grandpa*, Vosniadou, 1987; or referring to all round things as *ball*, Clark, 1973; Rescorla, 1980). Some theoretical accounts treat overextensions as being in a continuum of phenomena together with metaphor which involves lexical broadening resulting in a communicated ad hoc concept (Wałaszewska, 2011; see Wilson & Carston, 2007, for an outline of the theory of metaphor as ad hoc concept construction). Others see overextensions as early precursors to fully-fledged metaphors – in both cases a broadened concept is constructed based on properties associated with the conventional denotation of the chosen expression (which are typically perceptual in the case of overextension and early metaphors) – the main difference between the two being whether the speaker's violation of the conventional category boundary is intentional or not (Falkum, 2019; Pouscoulous, 2011). An important question is how children cope with metaphors that are not directly based on perceptual features, and which have less of an affinity with overextension – will they be more difficult for children to handle?

In a recent study looking at perceptual metaphors, children were asked to choose between the picture of a girl who looks like a parrot (metaphorical target), a distractor, and a picture of a parrot (literal competitor) when hearing an utterance such as *Lucy is a parrot* (Long et al., 2021). When including such a literal competitor, results point to a literal bias from 3 to 6 years old, which is not observed for thirteen year olds, who served as an older comparison group. Importantly, this is still a child-friendly task which does not involve the complexities mentioned in the previous section, tentatively suggesting that there might be differences brought on either by the morphosyntactic properties of the metaphorical expression or by methodological differences, such as the presence of a literal competitor (Winner, 1997).

Besides the earliest age of comprehension, it is important to elucidate the developmental trajectory of metaphor interpretation. Few studies to date have assessed a wide range of ages sufficient for evaluating the developmental trajectory of metaphor comprehension. Deamer (2013) evaluated three-, four-, and five-year-old children and Long et al. (2021) evaluated children aged 3-6 and thirteen-year-olds. Performance of participants in Deamer (2013) suggest a linear maturation, in which metaphor understanding strengthened in each tested age group. Young participants aged 3-, 4-, 5- and 6-years old

in Long et al. (2021) did not succeed on the metaphor task, potentially due to reasons discussed above. As such the developmental trajectory of predicative metaphors remains underexplored. Recently, Falkum and collaborators (Falkum et al., 2017; Köder & Falkum, 2020) have suggested the possibility of a U-shaped development of understanding metonymic expressions such as *the helmet rides her bike and goes home*, or *the beard is happy*. According to their findings, 3-year-olds show some signs of novel metonymy comprehension, while 4- and 5-year-olds display a preference for literal interpretations of the same utterances, with 6-year-olds reverting to a metonymic interpretation. Falkum and collaborators speculate that these results could be due to an increased sensitivity to the conventional meanings of the constituents of the metonymies. Younger children have flexible category boundaries (Vosniadou, 1987), which might allow them to accept metonymic or metaphoric interpretations. As children grow older, they become more sensitive to conventional meanings, i.e., to the fact that there are "correct" and "incorrect" uses of words, rendering the literal option more salient as the "correct" answer. As they continue to develop, children become more aware of how people may use language creatively (and often tend to purposely deviate from the literal-conventional), thus becoming able to grasp figurative meaning. This raises the question of whether such a developmental trajectory would also be present for metaphor interpretation.[4]

Importantly, Köder and Falkum (2020) did not rely on a simple object selection task in their study but used a Visual World paradigm. Here, children's eye-movements were tracked while they heard the target sentences unfold. With this paradigm, they were able to find that 4- and 5-year-old children show a preference for the metonymic interpretation (judging by the proportion of looks to the metonymic picture relative to looks to the literal one), but end up choosing the literal one. This amounts to a degree of dissociation between their online processing (indexed by their gaze behavior) and their end-result interpretation (indexed by the picture they point to at the end of a trial). Considering how such a paradigm has never (to our knowledge) been used to study metaphor development, it is important to investigate differences between processing and final interpretation of metaphor comprehension in order to understand not just the earliest age of adult-like metaphorical interpretation, but also how children's processing strategies more generally develop with age.

The present study

In the current study, we investigate the developmental trajectory of comprehension of predicative nominal ('X is a Y') metaphors that involve an explicit category clash, such as *that kid is a grasshopper*. We deployed a Visual World, eye-tracking paradigm, which allowed us to test both children's final interpretation of a metaphor as well as their developing online metaphor processing skills. The goal of our study was three-fold. First, we aimed to investigate how children aged 3-9 comprehend such metaphors when the task is age-appropriate. Are the youngest children able to comprehend them, or is the degree of difficulty something that only older children can master? Second, we aim to describe the developmental trajectory of metaphor interpretation. Do children get better

---

[4]Metonymy and metaphor (when they are novel) share relevant similarities in that the child must accommodate a departure from the conventional sense of a word. These devices differ, *inter alia*, with regard to how the relationship to the referent is established, via contiguity relations in the case of metonymy or resemblance relations in the case of metaphor.

at comprehending metaphors in a lineal way, or can their development best be described as a U-shape, as has been claimed to be the case for metonymy? Third, we seek to determine how explicit behavioral measures (that tap into the final stages of interpretation) map onto implicit behavioral measures that index online sentence processing. Do these two measures go hand in hand, or do they vary in a way that can be revealing of a change in processing strategy with age?

Participants

We recruited 80 participants between the ages of 3 and 9. Distribution of participants for each age group was as follows: 3 year olds (5 children), 4 year olds, (8), 5 year olds (22), 6 year olds (16), 7 year olds (14), 8 year olds (10), 9 year olds (3). Participants were selected from two public elementary schools in the Vitoria-Gasteiz (Álava, Basque Country) area. Legal guardians of the participants provided their explicit informed consent prior to testing. All participants were bilingual Basque-Spanish. Only children without special educational needs, diagnoses or behavioral issues were included in the study. No participants were excluded based on our pre-determined exclusion criteria. The distribution of the number of participants per age bracket is shown in Table 1 below. Ethical approval was issued by the University of the Basque Country's Ethics Committee for research with human beings.

Materials and design

Our experimental items contained purely novel metaphors. To ensure these metaphors were novel in Spanish we conducted a small survey in which we included several novel metaphors and a Likert Scale (1 to 7). We asked adult natives (N = 21) whether the metaphoric expression was familiar to them. We intertwined control items with conventional metaphors in Spanish. For this experiment, we selected metaphors whose mean score was under 3.5. Items used in this study and the corresponding novelty ratings are included in Appendix 1.

The study used a Visual World, eye-tracking paradigm. In each trial, participants heard two utterances: A context and a critical utterance (see Figure 1). After this, they heard a question prompting them to select one of the four pictures that they saw on the computer screen in front of them. The pictures represented a literal and a metaphorical interpretation of the final word in the critical sentence, together with two distractors (one human figure and one animal/object figure, depending on the type of metaphor)[5] (see Figure 2). Participants' eye-movements were recorded as they heard both utterances in a trial.

---

[5]The distractors needed to resemble each critical option (metaphorical/literal) in both the structure and the content. If we were including people doing things, objects working in a particular way, or animals doing things, the corresponding distractors would have to depict the same but not exactly.

This was a way to ensure that children were obligated to process all words included in the sentence to choose correctly. In other words, if the child only grasps that the critical sentence talks about humans (that child) or that it talks about doing math/jumping, it is not enough, and they'll be confused between the distractors and the correct choices. Thus, our distractors not only distract per se but also ensure that errors are not due to a lack of comprehension of the literal meanings of the words included. If there are errors, they must be due to something else.

Table 1. Familiarity rating of novel metaphor test items and conventional metaphor control items

Novel metaphor(Test items)		English translation	Mean	SD	Median
1	Luis es una tijera	*Luis is a scissor*	1.08	0.27	1
2	Juan Carlos es un pez espada	*Juan Carlos is a swordfish*	1.12	0.33	1
3	Guillermo es un río	*Guillermo is a river*	1.27	0.72	1
4	Corina es una fresa	*Corina is a strawberry*	1.35	0.85	1
5	Laura es un cocodrilo	*Laura is a crocodile*	1.42	1.03	1
6	Lola es una oruga	*Lola is a caterpillar*	1.42	0.76	1
7	Pepa es un leopardo	*Pepa is a leopard*	1.65	1.35	1
8	Javi es un globo	*Javi is a balloon*	1.65	1.62	1
9	Berta es un rinoceronte.	*Berta is a rhinoceros*	1.81	1.36	1
10	Carolina es un coche de carreras	*Carolina is a race car*	1.92	1.6	1
11	Carmen es una ardilla	*Carmen is a squirrel*	2.27	1.69	1
12	Manuel es una oveja	*Manuel is a sheep*	2.31	1.81	1
13	Vicente es un hipopótamo	*Vicente is a hippopotamus*	2.35	1.87	1
14	Fernanda es una aspiradora	*Fernanda is a vacuum*	2.35	2.15	1
15	Daniel es un flamenco.	*Daniel is a flamingo*	2.38	2.14	1
16	Maikel es una peonza	*Maikel is a spinning top*	2.46	1.98	1
17	Tomás es un saltamontes	*Tomás is a grasshopper*	2.69	1.98	2
18	Jesús es un pollito	*Jesús is a chick*	2.85	2.15	2
20	María es un ordenador	*María is a computer*	3	2.15	2
Novel metaphor average			**2.02**	**1.50**	**1.2**
Novel metaphor(Excluded items)		English Translation	Mean	SD	Median
1	Natalia es una cebra	Natalia is a zebra	1.27	0.83	1
2	Alejandro es un pingüino	Alejandro is a penguin	1.27	0.6	1
3	Clara es un oso polar	Clara is a polar bear	1.42	0.9	1
4	Elena es un bisonte	Elena is a bison	1.73	1.54	1
5	Gorka es un búfalo	Gorka is a buffalo	1.96	1.64	1
6	Carmen es un avestruz	Carmen is a stork	2.27	1.8	1
7	Ignacio es un ratón	Ignacio is a mouse	3.08	2.1	3
8	Teresa es un altavoz	Teresa is a loudspeaker	3.12	1.97	3
9	Alberto es una pantera	Albert is a panther	3.35	2.19	3
10	José Antonio es una serpiente	José Antonia is a snake	3.58	2.27	4
11	Gisela es una flor	Gisela is a flower	3.81	1.98	4
12	Andrés es un libro	Andrés is a book	3.88	2.21	4
13	José es un mono	José is a monkey	4.62	2.02	5
14	Ander es una gacela	Ander is a gazelle	4.96	2.01	5

Table 1. *(Continued)*

	Conventional metaphor (Control items)	English translation	Mean	SD	Median
1	Julián es un cordero	*Julián is a sheep*	2.85	2.5	1
2	Gaizka es un oso	*Gaizka is a bear*	3.54	2.39	3
3	Julen es una hormiga	*Julen is an ant*	3.96	2.34	4
4	José Antonio es un gusano	*José Antonio is a worm*	4.31	2.41	5
5	Romina es una jirafa	*Romina is a giraffe*	4.58	2.4	5.5
6	Teresa es una tabla	*Teresa is a table*	4.81	2.3	5
7	Jon es una seta	*Jon is a mushroom*	4.88	2.39	6
8	Aitor es un perro	*Aitor is a dog*	5.35	2.19	6
9	Garazi es una tortuga	*Garazi is a turtle*	5.54	2.08	6
10	Carlos es un buitre	*Carlos is a vulture*	5.65	2.04	7
11	Iñigo es un gallina	*Iñigo is a chicken*	5.69	2.26	7
12	Catalina es una leona	*Catalina is a lion*	5.77	1.86	7
13	Jorge es una rata	Jorge is a rat	6.15	1.46	7
14	Lucía es un loro	*Lucía is a parrot*	6.31	1.38	7
15	Natalia es un trozo de pan	*Natalia is a piece of bread*	6.42	1.65	7
16	Laura es una víbora	*Laura is a viper*	6.54	0.86	7
17	Samanta es una bestia	*Samantha is a beast*	6.62	0.64	7
18	Timoteo es un burro	*Timoteo is a donkey*	6.69	0.62	7
19	Juan es un cerdo	*Juan is a pig*	6.77	0.59	7
20	Jaime es una máquina	*Jaime is a machine*	6.77	0.43	7
21	Marina es un lince	*Marina is a lynx*	6.77	0.51	7
	Conventional metaphor average		**5.52**	**2.66**	**5.98**

The study had two independent variables. The first one was the categorical variable UTTERANCE TIME with two levels: 'metaphorical' and 'literal'. This factor refers to whether the critical utterance that participants heard in a trial was intended to be interpreted metaphorically or literally. The context utterance was the same in both conditions; it is the target utterance that comes right afterwards that makes the difference. In the literal condition, the target utterance refers to either an animal, or an object, and predicates a feature about them. In the metaphorical condition, the target utterance refers to a human being, and (metaphorically) predicates a feature about them. In the metaphorical condition, the context sentence provided children with information about which property associated with the meaning of the metaphorical vehicle was relevant to appropriately interpret the metaphor.

The literal condition served as a baseline for children's capability to both integrate two sentences together when the meaning expressed is not figurative, as well as to ensure that children understood the literal meaning of the critical metaphorical vehicles (e.g.,

Utterance type	Context utterance	Target utterance		Question
			Target word	
METAPHORICAL	Los saltamontes saltan mucho.	Ese **niño** es un	**saltamontes**.	¿Cuál es?
	DET.PL grasshoppers jump much	that kid is a	grasshopper	which is
	'Grasshoppers jump a lot.	That **kid** is a	**grasshopper**.	Which one is it?'
LITERAL	Los saltamontes saltan mucho.	Ese **animal** es un	**saltamontes**.	¿Cuál es?
	DET.PL grasshoppers jump much	that animal is a	grasshopper	which is
	'Grasshoppers jump a lot.	That **animal** is a	**grasshopper**.	Which one is it?'

CRITICAL
time window

QUESTION
time window

Figure 1. Example of a critical trial in the metaphorical and critical conditions.

Figure 2. Example of the visual display for a critical trial.

'grasshopper'). An example of a critical trial in both conditions is shown in Figure 1, including the distribution of relevant time-windows.

The second independent variable was the continuous variable AGE, measured in days. This referred to the age of the participant on the day of testing.

We created 20 items, with two versions each (for each condition). Items in both conditions were counterbalanced across two experimental lists. All items, as well as a description of the pictures accompanying them, are included in Appendix 2.

Procedure

All participants were tested in a quiet room. They 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. After the experimental session, they were given stickers as compensation.

Analysis

We analyzed two types of data collected from the experiment: The gaze behavior of participants during a trial (i.e., which of the pictures they were looking at as a function of the unfolding utterance) as well as the explicit picture selected at the end of each trial. Our analysis script, data and list of materials are freely accessible on the project's OSF page: https://osf.io/2qfpx/?view_only=8de767c6dfb348b0af66afbe181a2565.

Gaze behavior

Gaze behavior was analyzed using log-gaze probability ratios (Arai et al., 2007) of proportion of looks to metaphorical picture divided by the proportion of looks to the literal picture, as previously done in a similar experiment on adult metaphor comprehension (Ronderos et al., 2022). We analyzed log-gaze probabilities across two time-windows. The first was time-locked from the onset to the offset of the metaphorical vehicle (e.g., *grasshopper*), i.e., the critical word (CRITICAL region). The second was time-locked from the onset until the offset of the final question (QUESTION region). These two analyses can tell us whether there is a preference for viewing the literal or metaphorical picture immediately upon hearing the metaphorical vehicle (CRITICAL time window) or shortly after, but before explicitly selecting a final interpretation (QUESTION time window).

To analyze the gaze data we fitted a linear mixed-effects model including UTTERANCE TIME, AGE and their interaction as predictors. This model was meant to assess the developmental trajectory of metaphor processing. To better understand the potential interaction between AGE and UTTERANCE TIME, we coded UTTERANCE TIME using treatment contrast coding, fitting the model first with the 'metaphorical' condition as the intercept of the model, and then re-fitting it with the 'literal' condition as the baseline. This way, we could answer the question of whether AGE affected the comprehension of metaphorical or literal utterances (or both). The random effects structure of the model was maximal (in the sense of Barr et al., 2013), with random intercepts and slopes for AGE, UTTERANCE TIME, and their interaction by items and random intercepts and slopes for UTTERANCE TIME by participants.

Picture selection

Picture selection behavior was analyzed by first removing instances where participants selected one of the distractor images at the end of a trial (which occurred in 3% of all trials). We then coded the remaining trials based on whether participants had accurately selected the target image (coded as 1) or incorrectly selected the competitor image (coded

as 0). We then fitted two different types of models to this resulting data. The first and main model was a logistic mixed-effects regression model including UTTERANCE TIME, AGE, and their interaction as predictors. Similar to the model used for analyzing gaze data, this model evaluated the developmental trajectory of metaphor comprehension. Critically, this model included AGE as a centered linear predictor as well as a quadratic predictor (using second-order orthogonal polynomials), a procedure known as growth curve analysis (Mirman, 2017). This was done to investigate whether metaphor processing in our task could be best described as having a linear or a non-linear (U-shaped) developmental trajectory. UTTERANCE TIME was treatment-contrast coded with the 'metaphorical' condition coded as the baseline, and the model had a 'maximal' random effects structure, with random intercepts and slopes for AGE, UTTERANCE TIME, and their interaction by items and random intercepts and slopes for UTTERANCE TIME by participants.

The second type of model was identical to the first but included AGE as a categorical predictor with four levels: 3-4 year-olds, 5-year-olds, 6-year-olds and 7-9 year-olds.[6] This complementary model was used in order to obtain a coarse estimate on the earliest age bracket in which, on average, metaphors are understood correctly. Importantly (given how the age groups have varying number of participants), the groups are not compared to each other statistically, and are used only to evaluate whether for each of the groups the probability of selecting the correct image is significantly higher than chance. To do this, we fitted this model three times, varying the baseline level of the categorical predictor UTTERANCE TIME in order to evaluate if the intercept is significantly different than zero in the 'metaphorical' condition.

Results

Gaze data

The results are shown in Figure 3 and the relevant model output summary can be seen in tables 2 and 3 and tables I and II in Supplementary materials (displaying results with a different reference level in the variable CONDITION). Our plots depict the preference for one picture or the other in both conditions. The preference for the metaphorical picture would be represented in the positive area (0 to 1) whereas the preference for the literal picture would be represented in the negative area (0 to -1). The younger the children, the closer both lines are to zero, therefore, the less clear their preference for the metaphorical picture in the metaphorical condition. In terms of statistical comparisons, both time windows show a similar pattern: First, there is a significant effect of UTTERANCE TIME, with more looks to the metaphorical picture (i.e., larger log-gaze probability ratios) in the metaphorical condition relative to the literal condition (tables 2 and 3; CRITICAL time-window: t-value =13.36, $p < 0.001$; QUESTION time-window: t-value = 10.32, $p < 0.001$). This means that children of all ages preferred to look at the metaphorical picture in the metaphorical condition compared to the literal one. Further, there is a significant effect of AGE on the metaphorical condition (tables 2 and 3; CRITICAL time-window: t-value = 2.37, $p < 0.05$; QUESTION time-window: t-value =2.45, $p < 0.05$), but not on the literal condition (tables I and II in supplementary materials; CRITICAL time-window: t-value = 0.7, $p = 0.45$; QUESTION time-window: t-value = 1.34, $p = 0.18$). This resulted in an interaction between

---

[6]The ages were grouped in this way to strive for a more comparable number of participants in each group for the purpose of statistical analysis.
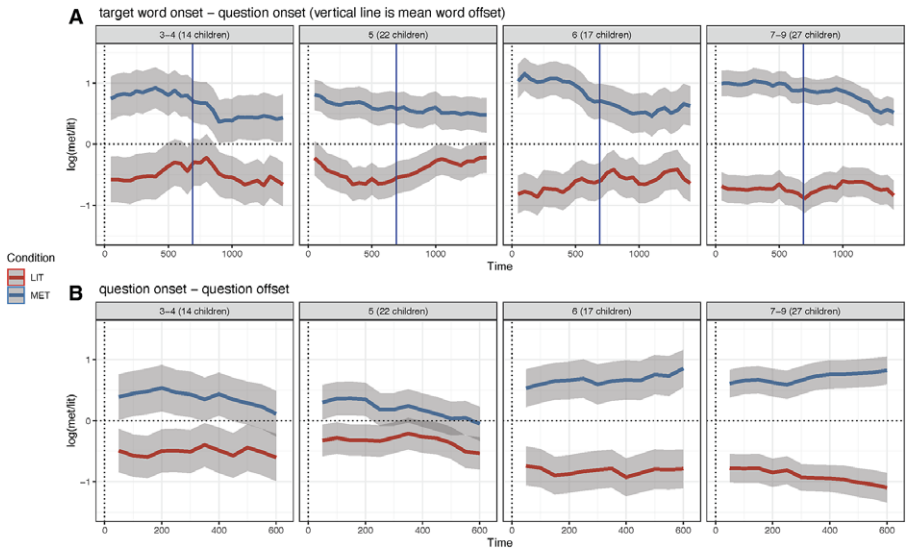
Figure 3. Time-course of critical (panel A) and question (panel B) time windows. Error ribbons show 95-percent confidence intervals. Age groups are shown for the purpose of clarity: statistical analyses were conducted using age as a continuous variable.

Table 2. Eye-tracking results with metaphorical condition as baseline, critical time-window

term	B	95% CI	*t*	*df*	*p*
MET	0.73	[0.55, 0.91]	7.85	59.83	< .001
MET v. LIT	−1.33	[−1.53, −1.14]	−13.36	80.10	< .001
AGE (centered)	0.18	[0.03, 0.33]	2.37	87.72	.020
UTTERANCE TIME X AGE	−0.24	[−0.44, −0.04]	−2.30	68.88	.025

Note. First row shows condition coded as intercept.

Table 3. Eye-tracking results with metaphorical condition as baseline, question time-window

term	B	95% CI	*t*	*df*	*p*
MET	0.50	[0.29, 0.71]	4.69	70.70	< .001
MET v. LIT	−1.18	[−1.40, −0.96]	−10.32	77.79	< .001
AGE (centered)	0.25	[0.05, 0.45]	2.45	74.70	.017
UTTERANCE TIME X AGE	−0.39	[−0.64, −0.14]	−3.08	61.43	.003

Note. First row shows condition coded as intercept

UTTERANCE TIME and AGE in both time-windows (tables 2 and 4; CRITICAL time-window: t-value = 2.3, p < 0.05; QUESTION time-window: t-value = 3, p < 0.005). This result suggests that, as they get older, participants in our experiment show an increased preference for viewing the metaphorical picture in the metaphorical condition. However, the preference for viewing the literal picture in the literal condition remains relatively constant with age.

Table 4. Picture selection results with metaphorical condition as baseline

term	B	95% CI	z	p
MET	3.06	[1.34, 4.77]	3.49	< .001
MET v. LIT	3.15	[0.31, 5.98]	2.18	.029
AGE (lineal)	26.25	[14.13, 38.38]	4.24	< .001
AGE (quadratic)	−0.72	[−12.85, 11.41]	−0.12	.907
UTTERANCE TIME X AGE (lineal)	−19.61	[−36.62, −2.60]	−2.26	.024
UTTERANCE TIME X AGE (quadratic)	−1.16	[−18.13, 15.82]	−0.13	.894

Note. First row shows condition coded as intercept; AGE is a continuous predictor measured in days
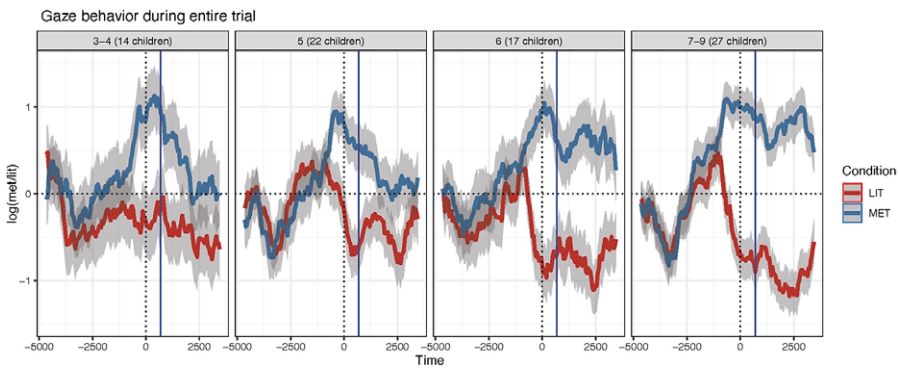
Figure 4. Gaze behavior during the entire trial ('Grasshoppers jump a lot. That child/animal is a grasshopper. Which one is it?'). The data is time-locked to the onset of the critical word ('grasshopper'). Error ribbons show 95-percent confidence intervals. Dotted line shows the beginning of the critical word, and the blue vertical line shows the offset of the critical word.

Finally, Figure 4 shows children's gaze behavior during the entire trial in both conditions. The first half of both lines (until approximately 3500-4000 ms.) corresponds to the context time window, i.e., when children hear "Grasshoppers jump a lot". In this window, we see that all participants are inclined to look towards the literal picture (i.e., to prefer it), which corresponds to the fact that they are hearing the context phrase which directs attention to the literal picture (in this example, the grasshopper). Gradually, in the metaphorical condition (indicated by the blue line) the preference trends towards the metaphorical picture, as the blue line dives deeper in the "positive" area of the graph. Importantly, children start looking more towards the metaphorical referent even before they hear the metaphorically used target word *grasshopper* (the beginning of which is depicted by the dotted line, in the figure), likely as a result of their hearing the noun phrase *That child,* and combining it with the information provided by the verb in the context utterance, 'Grasshoppers *jump* a lot'. This preference in children's gaze behavior remains present also during the first milliseconds of the question region (cf. Figure 3). However, while the older children hold on to this preference in their choice of metaphorical picture option, the younger children seem to finally disregard it in the picture selection task, as we will see in the next section.

Figure 5. Results of picture selection task with age as a continuous predictor and a superimposed regression Line. Error ribbons show 95-percent confidence intervals.

Picture selection

Results of the picture selection task can be seen in Figure 5, while the relevant model output is displayed in table 4. The model shows a significant effect of UTTERANCE TIME on picture selection, with participants choosing the target picture significantly less often in the metaphorical relative to the literal condition. Parallel to the gaze data, there was a significant effect of AGE on the metaphorical but not on the literal condition (see table III in supplementary materials), resulting in an interaction between AGE and UTTERANCE TIME. Critically, we did not find a significant effect of the second-order polynomial of AGE on picture selection (table 4). This suggests that the developmental trajectory of selecting the target picture in the metaphorical condition is better described as linear and not U-shaped[7].

The complementary models including AGE as categorical predictor showed that when the '3-4-year-olds' and the '5-year-olds' age brackets were coded as the baseline conditions of the model (table IV and table V in supplementary materials), the intercepts were not significantly different from zero ('3-4-year-olds': *beta-coefficient: -0.410, z-value: -0.232, p-value: 0.817;* '5-year-olds': *beta-coefficient: 0.595, z-value: 0.351, p-value: 0.726).* This was the case only for the older age brackets ('6-year-olds': *beta-coefficient: 5.116, z-value: 2.748, p-value<0.01;* '7-9-year-olds': *beta-coefficient: 6.877, z-value: 5.482, p-value < 0.001).*

General discussion

In the present study, we aimed to contribute to the debate on the development of metaphor interpretation in three ways. First, we set out to test the earliest age in which children comprehend predicative nominal metaphors that display an overt category clash using a child-friendly paradigm that does not impose particular world knowledge or

---

[7]This test was done due to the U Shape Köder & Falkum found in metonymy comprehension. They only found it in picture selection data, not in gaze data, and so we tested a curved description of the data only with picture selection results.

metalinguistic demands on the participants. Second, we aimed to evaluate the developmental trajectory of metaphor interpretation, in order to assess whether it can be best described as a linear or a U-shaped development. Finally, we probed both the end-result of the interpretative process (using picture selection) as well as the incremental online processing of metaphors (using eye-tracking) in order to examine potential dissociation patterns between these two measures. Throughout this section, we will discuss in detail how our results could relate to the theoretical views set out in the Introduction, introducing different hypothesis. Besides, we will mention some limitations of our proposed experimental design (i.e. the demands the task might pose), as we also recommend interpreting our results for the youngest age groups with caution, since our sample included only five 3-year-olds and eight 4-year-olds, due to availability constrains at the schools we worked with.

In terms of our first goal, the results of the picture selection task suggest that children under the age of 6 seem to not comprehend our metaphors. In the 3-4 and 5-year-old age brackets we failed to find a significant preference for either selecting the incorrect – literal or the correct – metaphorical picture in the metaphorical condition, judging by the non-significant intercept of the models for which these age groups were used as the reference level (see tables IV and V in supplementary materials). Only the older age brackets (6 and 7-8 year-olds, table VI) displayed a significant preference for selecting the metaphorical picture in the metaphorical condition, suggesting that they were overall able to grasp the figurative meaning of our test items.

In terms of our second goal, the results suggest that the developmental trajectory of metaphor comprehension is best described as being lineal and not U-shaped: Using growth curve analysis, we found that the second-order polynomial of the predictor AGE did not have a significant effect on metaphor comprehension. The first order polynomial, however, did have a significant effect on the interpretation of metaphorical but not on that of literal expressions, resulting in an interaction pattern. This suggests that, with age, participants in our experiment became better at correctly identifying the metaphorical target picture in a lineal way (and correspondingly became less likely to choose the literal picture), whereas grasping the literal meaning was something they were able to do consistently regardless of their age. This falls in line with what we may gather from the current body of literature on metaphor (e.g., Deamer, 2013; or a classic study by Vosniadou et al., 1984), who similarly found a linear maturation in picture selection.

Finally, the results of the eye-tracking experiment showed that, in general, participants preferred to look at the metaphorical picture in the metaphorical condition, and at the literal picture in the literal condition. This pattern was visible both as the target metaphorical vehicle was heard (CRITICAL time window) as well as immediately after (QUESTION time window). Further, both time windows showed an interaction between AGE and UTTERANCE TIME, suggesting that older children tended to look more at the metaphorical picture in the metaphorical condition relative to younger children, whereas looks to the literal picture in the literal condition remained relatively stable with age.

In sum, we find linear developmental trends in both gaze and picture selection data. However we see that children's 'preference' differs between measures in function with age. For children aged 6 and older there is strong evidence for successful comprehension of novel nominal predicative metaphors: these children both look towards the metaphoric referent in the unfolding of the target sentence and successfully select the correct metaphoric referent when prompted. Children aged five and under present a different pattern, looking significantly towards the metaphoric referent during the unfolding of the target sentence while ultimately mainly selecting the literal option. We will first situate

our picture selection results among relevant previous findings and then reflect potential interpretations for the discrepancy between gaze and picture selection measures for our youngest participants.

Concerning the debate about development ("literal stage" or "early birds"), to which we intended to contribute; recall that advocates of the literal stage claimed that children would not be able to comprehend metaphors until the age of 10-12 years old, while the "early birds" view defended that children as young as 3 should be able to do so. Our results somewhat question the developmental claim of the literal stage hypothesis. It is clear that young children (6-year-olds and some 5-year-olds at least) are capable of comprehending novel metaphors that include violation of conventional category boundaries in a predicative nominal form, when considering children's picture selection. Importantly, this was the first study to shed light on the developmental trajectory of predicative metaphor, capturing the transitional age at which children succeed in explicit measures. Long et al. (2021) also tested a similar age range, however participants did not demonstrate an understanding of the metaphors tested. Both studies are similar in that the metaphors used are nominal predicative metaphors in Spanish (as an example from Long et al.'s: 'Lucía es un loro'; *lit: Lucy is a parrot*, for a girl looking like a parrot), and some of our chosen metaphors are also animal-human. However, there are key differences between the current study and Long *et al.*'s, besides their metaphors being more perceptually-based. One such difference is the type of literal competitor included; while the "literal" picture in Long et al.'s study depicted the interpretation of the whole statement (e.g., a picture of a parrot for 'Lucy is a parrot', which would amount to a parrot named Lucy) our literal competitor represented the literal meaning of the metaphor vehicle (e.g., a grasshopper). Compared to Long et al.'s (2021), our choice of literal competitor is in line with theoretical accounts which take metaphors to be interpreted through a process of pragmatic adjustment of the concept encoded by the metaphorical vehicle (e.g., so-called *ad hoc* concept construction; Sperber & Wilson, 2008; Wilson & Carston, 2006). In our paradigm, if this adjustment is not performed, the literal interpretation of the whole utterance is absurd or impossible (i.e, a child being literally a grasshopper), whereas in Long et al.'s design, such an interpretation is possible – and indeed, plausible). The second difference between Long *et al.*'s design and the current study is that, as the authors themselves point out, one crucial aspect to disambiguate between Lucy being a girl or a parrot in the example is that, in Spanish, 'Lucía' (*Lucy*) is not usually a name for a pet. This means that older children may have accessed a piece of knowledge about linguistic conventions that younger children do not have at their disposal. Results from the present study suggest that comprehension of predicative metaphor may develop earlier than previously known.

Regarding 3-year-olds being early birds in understanding metaphors, it seems that this may apply to a certain type of metaphorical statements. Reviewing the studies which indicated correct picture selection by the age of three (Deamer, 2013; Di Paola et al., 2020; Pouscoulous & Tomasello, 2020), we find a similar pattern in that the metaphorical mappings included are strongly based on perceptual similarity between objects (e.g., *the tower with the hat* for a tower with a pointy roof; *the car that has a bad foot,* for a car with a flat tire). As stated in the introduction, the type of metaphorical mapping arguably influences performance. Although our metaphors possess imaginable elements, our mappings are not based on perceptual similarity, but rather on actions and functions that both the metaphorical target and the vehicle can perform. Besides, our metaphors are all nominal predicates in form and as such, less similar to overextensions and renamings that children spontaneously perform. It might be that our metaphors are more difficult to

solve than the ones traditionally included in "early birds" supportive studies. Other relevant methodological differences should also be considered: in some design contexts no other potentially relevant interpretation option was available (cf. Pouscoulous & Tomasello, 2020), and some experimental contexts focused attention on the correct human referent prior to the target metaphoric statement (Deamer, 2013); both things likely posing low demands on inhibitory control. Conversely, in our task, children had to use the property of the metaphorical vehicle provided verbally in the context sentence (*Grasshoppers jump a lot*) to arrive at a metaphorical interpretation, suppressing other associated features. They lacked other clues that would make the interpretation depicted by the correct picture more salient. In fact, given that the vehicle term is heard twice (in the context and in the metaphorical statement), its literal meaning likely receives strong activation and may make feature suppression harder for some children. I.e., our task might be more demanding for children's general executive function and in particular inhibitory control. As general executive function improves with age, it has been proven to have a significant effect on metaphor comprehension (see Deamer, 2013). This might be viewed as a limitation of the current design with regard to testing the competence of our youngest participants. At the same time, this feature of the current design also provides strong evidence of older children's (but still young to the literal stage hypothesis) capability of suppressing literal meaning to reach the correct, metaphorical interpretation, regardless of the activation received through the context.

Across paradigms, evidence suggests that from a young age, children are prepared to solve and grasp metaphorical meanings. What accounts for differences among 'early birds' (i.e., whether we witness competence in picture selection from three to six years old) may depend on specifics of the methodology selected. This may be a relevant point for future research to address directly what may account for variability among the earliest of ages: type of metaphor assessed, how the relevant trait is made salient (perceptually or verbally) as well as the role of cognitive factors such as inhibitory control, which is rapidly maturing during this time span (see Deamer, 2013, for a correlation between metaphor comprehension and inhibitory control in children from these age spans).

As we note in the introduction, there are likely differing levels of metaphor comprehension and we cannot attest to whether participants in our study have a comprehension of metaphor in the fullest sense, such that the metaphor could be explained, reused or used in reasoning. That noted, taking into consideration both gaze and picture selection performance of participants aged six and older gives us confidence to infer successful comprehension for these age brackets. Moreover, there is recent evidence that children as young as 4 years old can use novel metaphorical meaning in their reasoning (Zhu & Gopnik, 2023), so that would speak in favor of our six-year-olds being capable of understanding metaphors in a deeper way. But how can we understand the performance of children who are five and under who prefer the metaphoric referent according to their gaze behavior at least until question onset, yet tend to select the literal picture? Considering both eye-tracking and picture selection results together, our findings are compatible with more than one interpretation. Both interpretations to be developed have to do with the notion of confidence and take into account how task properties might be affecting results.

One possible interpretation is that young children are sensitive to the metaphor yet for some reason exhibit a literalist picture selection behavior. This interpretation would parallel that developed in Köder and Falkum (2020) with regard to their results on novel metonymy, which share some similarities in gaze-selection patterns for children five and under: a dissociation between looks towards the non-literal referent in the gaze data while

selecting the literal option for 4- and 5-year-olds in Köder and Falkum (2020) and for our 3-4 and 5-year-old age brackets in the present metaphor study. As commented in the introduction, a similarity in metaphor and metonymy is a deviation from the conventional meaning of the metaphoric or metonymic vehicle. Köder and Falkum (2020) propose that a growing sensitivity to linguistic sense conventions may result in a literalist bias in picture selection, overriding a sensitivity to the non-literal interpretation. In this sense, one could interpret that all participants have some sensitivity to the metaphoric meaning of the utterance, yet the younger participants are less confident in their interpretation as it involves a clash with the conventional meaning of the metaphoric vehicle and thus ultimately opting for what they take to be the safest bet.

An alternative view would grant a weaker interpretation to their significant gaze towards the metaphoric referent - looking at the dynamic unfolding gaze behavior in terms of predictive processing and maturation of inhibitory control. Recall that all children first attend to the image of the grasshopper during the unfolding of the context sentence, and then shift their gaze to the metaphoric referent even prior to hearing the target word *grasshopper* in the target sentence *that child is a grasshopper.* Children may be integrating the relevant properties of the context sentence to restrict their gaze to the relevant person option, who shares the key property of jumping. However, one might take a more muted stance as to whether this unfolding sequence is indicative of a true sensitivity to the metaphoric meaning. Older children may be more confident about their choice because they are able to effectively suppress features of the metaphorical vehicle, while younger children may be in the process of learning to do it, thus they don't fully succeed in processing the metaphor, and this is the reason why they don't finally choose the correct picture. According to this interpretation, both types of data would blend in their linear trajectories, showing that with age, children get better at understanding metaphors.

In sum, our results suggest that children aged 6 and above reliably understand nominal predicative metaphors, given the specifics and constraints of our design.

We witness a linear maturation in picture selection, with 3-4-year-olds overwhelmingly selecting the literal referent, 5-year-olds in a transitional phase selecting the literal or metaphoric referent equally, and 6-year-olds demonstrating their comprehension in picture selection, despite a context that emphasizes the metaphoric vehicle. When considering gaze and picture selection data together, we see that there is a general alignment in terms of trajectory – both gaze behavior and selection behavior presenting a more or less linear development. With regard to children aged 5 and under, we see an apparent dissociation between gaze and picture selection, however it is uncertain to what extent this gaze behavior implies sensitivity to the metaphor as children are shifting their gaze to the correct referent even prior to hearing the metaphor in full. This is not to discount what younger children are doing however – they indeed appear sensitive to the context to restrict their gaze to the relevant human referent who shares key traits with the metaphoric vehicle. What we may infer is that all children prefer the human referent which shares relevant traits with the referent of the context sentence, however this preference may be more fragile for younger participants, specifically when asked to make an explicit choice. Younger children's tendency towards selecting the literal picture may have differing potential causes such as growing sensitivity to conventional meanings of words which make them less confident selecting a referent which clashes with the conventional meaning of the word or potentially a less mature inhibitory control capacity which affects their ability to fully process the metaphor.

Conclusion

In this paper, we investigated the development of metaphor interpretation using an eye-tracking, Visual World paradigm. Our study allowed us to both shed light on the moment-to-moment processing of predicative nominal metaphors, as well as on how this ability develops with age. Our results reveal a lineal development in metaphor comprehension both in terms of their online processing (judging by the eye-tracking record) and of their overall interpretation (judging by the picture selection data). It seems that only after six years of age can we say that children exhibit a level of comprehension of predicative nominal metaphors that allows them to choose the right picture – at least in the current set-up – and consistently display behavior compatible with the derivation of metaphor comprehension, in contrast with previous findings showing that even 3-year-olds are able to comprehend simple perceptual metaphors also in a forced choice paradigm. Our study therefore adds to the existing literature on metaphor comprehension by providing results about the development of comprehension of non-perceptual metaphors.

References

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

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

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

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 65–110). New York: Academic Press.

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

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. doi: 10.1016/0010-0285(74)90005-X

Deamer, F. (2013). *An investigation into the processes and mechanisms underlying the comprehension of metaphor and hyperbole* (Doctoral dissertation). Retrieved from: 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. doi: 10.1016/j.pragma.2019.08.008

Falkum, I. L. (2019). Metaphor and metonymy in acquisition: A relevance-theoretic perspective. In K. Scott, B. Clark & R. Carston (Eds.), *Relevance, Pragmatics, and Interpretation* (pp.205–217). Cambridge: Cambridge University Press.

Falkum, I. L. (2022). The development of non-literal uses of language: Sense conventions and pragmatic competence. *Journal of Pragmatics*, **188**, 97–107. doi: 10.1016/j.pragma.2021.12.002

Falkum, I. L., **Recasens, M.**, & **Clark, E. V.** (2017). "The moustache sits down first": On the acquisition of metonymy. *Journal of Child Language*, **44**(1), 87–119. doi: 10.1017/S0305000915000720.

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

Köder, F., & **Falkum, I. L.** (2020). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics*, **156**, 191–205. doi: 10.1016/j.pragma.2019.07.007

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. doi: 10.1017/S030500091800048X

Long, M., **Shukla, V.**, & **Rubio-Fernández, P.** (2021). The development of simile comprehension: From similarity to scalar implicature. *Child Development*, **92**(4), 1439–1457. doi: 10.1111/cdev.13507

Malgady, R. G. (1977). Children's interpretation and appreciation of similes. *Child Development*, **48**(4), 1734–1738.

Mirman, D. (2017). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC.

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. doi: 10.1044/jshr.2702.197

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

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

Pouscoulous, N., & **Tomasello, M.** (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, **156**, 160–167. doi: 10.1016/j.pragma.2019.05.021.

Rescorla, L. (1980). Overextension in early language development. *Journal of Child Language*, **7**(2), 321–335. doi: 10.1017/S0305000900002658

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

Smith, J. (1976). Children's comprehension of metaphor: A Piagetian interpretation. *Language and Speech*, **19**(3), 236–243. doi: 10.1177/002383097601900304

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. doi: 10.1017/CBO9780511816802.007

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. doi: 10.1126/science.7777863

Vosniadou, S. (1987). Children and metaphors. *Child Development*, 870–885. doi: 10.2307/1130223

Vosniadou, S. (1989). Context and the development of metaphor comprehension. *Metaphor and Symbol*, 4(3), 159–171. doi: 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*, 1588–1606. doi: 10.2307/1130028

Wałaszewska, E. (2011). Broadening and narrowing in lexical development: How relevance theory can account for children's overextensions and underextensions. *Journal of Pragmatics*, 43(1), 314–326. doi: 10.1016/j.pragma.2010.07.017

Wilson, D., & Carston, R. (2006). Metaphor, relevance and the "emergent property" issue. *Mind & Language*, 21(3), 404–433. doi: 10.1111/j.1468-0017.2006.00284.x

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). London: Palgrave.

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

Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12(4), 289. doi: 10.1037/0012-1649.12.4.289

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

Appendix 1

Novelty control study

A total of 21 adult participants responded to an online familiarity rating task. Participants were native speakers of Spanish. They were asked to rate how familiar they were with each metaphorical expression on a 7-point scale (1: very unfamiliar and 7: very familiar).

Overall, the mean familiarity score for our 20 experimental items is 2.02 (SD=1.5; median=1). Fourteen novel metaphor items were excluded either for having a familiarity score of 3 or above or for other considerations.

The table below lists the mean familiarity score for our 20 included novel metaphor test items, 14 excluded novel metaphor items, and 21 conventionalized metaphors, which served as control items for familiarity rating.

Appendix 2

Experimental materials

Squirrels climb up trees. That boy is a squirrel.

1a. Las ardillas suben árboles. Ese niño es una ardilla. ¿Cuál es? 1
1b. Las ardillas suben árboles. Ese animal es una ardilla. ¿Cuál es? 2

1. Literal option: *a squirrel climbing up a tree.*	0. Distractor-literal: *a cat jumping to the top of a wall.*
0. Distractor-metaphorical: *a boy going up a ladder (like a metal ladder, those that we use to fix lightbulbs)*	0. Metaphorical option: *a boy climbing up a tree.*

Swordfishes swim fast. That boy is a swordfish.

2a. Los peces espada nadan rápido. Ese niño es un pez espada. ¿Cuál es? 2
2b. Los peces espada nadan rápido. Ese animal es un pez espada. ¿Cuál es? 1

5. Literal option: *a swordfish*	6. Distractor-literal: *a tiger*
7. Distractor-metaphorical: *a boy in the beach*	8. Metaphorical option: *a boy swimming*

Grasshoppers jump a lot. That boy is a grasshopper.

3a. Los saltamontes saltan mucho. Ese niño es un saltamontes. ¿Cuál es? 1
3b. Los saltamontes saltan mucho. Ese animal es un saltamontes. ¿Cuál es? 2

9. Literal option: *a grasshopper jumping*	10. Distractor - literal. *A beatle jumping*
Distractor-metaphorical: a boy running	11. Metaphorical option: *a boy who is jumping.*

Rhinos are strong. That boy is a rhino.

4a. Los rinocerontes son fuertes. Ese niño es un rinoceronte. ¿Cuál es? 2
4b. Los rinocerontes son fuertes. Ese animal es un rinoceronte. ¿Cuál es? 1

12. Literal option: *a rhino*	13. Distractor - literal. *A giraffe*
14. Distractor-metaphorical: *a boy playing football with friends*	15. Metaphorical option: *a very big boy with wide shoulders and strong arms (his biceps flexing, perhaps). We can surround him with other two skinny kids, so that the contrast is noticeable (he can be in the middle)*

Crocodiles have big teeth. That boy is a crocodile.

5a. Los cocodrilos tienen los dientes grandes. Ese niño es un cocodrilo. ¿Cuál es? 1
5b. Los cocodrilos tienen los dientes grandes. Ese animal es un cocodrilo. ¿Cuál es? 2

16. Literal option: *a crocodile with mouth open (big, sharp teeth) trying to bite a beach ball*	17. Distractor-literal: a tiger with big teeth trying to bite a beach ball
18. Distractor-metaphorical: the face and head of boy with long hair	19. Metaphorical option: the face of a boy with a big mouth smiling a lot. *Big teeth are noticeable, something like the Cheshire cat;)*

Sheeps have a lot of wool. That boy is a sheep.

6a. Las ovejas tienen mucha lana. Ese niño es una oveja. ¿Cuál es? 2
6b. Las ovejas tienen mucha lana. Ese animal es una oveja. ¿Cuál es? 1

20. Literal option: a sheep with lots of wool	21. Distractor-literal: a cow
22. Distractor-metaphorical: A boy wearing leather jacket and cowboy boots.	23. Metaphorical option: A boy wearing lot of white fleece clothing (fuzzy white fleece sweater, hat and mittens). It is important that the clothes make the child look like a sheep, so that the clothes are white and fluffy.

Leopards have a lot of dots. That boy is a leopard! (kid with noticeable freckles)

7a. Los leopardos tienen muchos puntos. Ese niño es un leopardo. ¿Cuál es? 1
7b. Los leopardos tienen muchos puntos. Ese animal es un leopardo. ¿Cuál es? 2

24. Literal option: A leopard	25. Distractor-literal: a zebra
26. Distractor-metaphorical: The face of a boy with sunglasses.	27. Metaphorical option: *A boy with noticeable freckles on his face (just the face in the drawing, we don't need more details)*

Hippos have a very big mouth. That kid is a hippo!

8a. Los hipopótamos tienen la boca muy grande. Ese niño es un hipopótamo! 2
8b. Los hipopótamos tienen la boca muy grande. Ese animal es un hipopótamo! 1

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

Chickens always follow their mum. That boy is a chicken!

9a. Los pollitos siempre van con su mamá. Ese niño es un pollito. ¿Cuál es? 1
9b. Los pollitos siempre van con su mamá. Ese animal es un pollito. ¿Cuál es? 2

32. Literal option: baby chicken walking next to her mother hen	33. Distractor-literal: two adult cats sleeping together
34. Distractor-metaphorical: a boy and a girl facing each other	35. Metaphorical option: *A little boy walking next to her mother, following her by the hand, resembling the hen and chicken picture*

Flamingos are pink. That kid is a flamingo!

10a. Los flamencos son rosa. Ese niño es un flamenco. ¿Cuál es? 2
10b. Los flamencos son rosa. Ese animal es un flamenco. ¿Cuál es? 1

36. Literal option: A flamingo	37. Distractor–literal: a yellow duck
38. Distractor–metaphorical: A blond–haired boy who is very pale (perhaps looks cold with bluish lips)	39. Metaphorical option: *A boy with very pink–tone skin (similar to the flamingo pink). As if he were sunburnt, but a bit hyperbolic, so that the pink is noticeable. Also pink hair. Preferable that the boy is wearing a bathing suit, that boy more of his pink skin is visible*

Caterpillars move (slither, squirm) on the ground. That kid is a caterpillar!

11a. Las orugas van por el suelo. Ese niño es una oruga. ¿Cuál es? 1
11b. Las orugas van por el suelo. Ese animal es una oruga. ¿Cuál es? 2

40. Literal option: A caterpillar inching on the ground	41. Distractor–literal: A frog jumping
42. Distractor–metaphorical: a boy stretching (standing up), maybe stretching one of his sides.	43. Metaphorical option: *a boy 'inching' like a caterpillar on the ground*

Vacuum cleaners absorb everything. That kid is a vacuum cleaner!

12a. Las aspiradoras absorben todo. Ese niño es una aspiradora. ¿Cuál es? 2
12b. Las aspiradoras absorben todo. Esa cosa es una aspiradora. ¿Cuál es? 1

44. Literal option: A vacuum cleaning being pushed by a man with some dirt or debris in its path to be vacuumed.	45. Distractor–literal: A man pushing a shopping cart.
46. Distractor–metaphorical: A boy eating a salad with a fork.It is enough to depict only the upper half of his body	47. Metaphorical option: A boy eating a plate of spaghetti. The child is eating spaghetti 'like a vacuum cleaner' and so probably with lots of noodles in his mouth at once. Surrounded by a lot of empty plates, it is obvious that he ate them. It is enough to depict only the upper half of his body, maybe sitting at a table, and the plates around him.

Skyscrapers are tall. That kid is a skyscraper!

13a. Los rascacielos son altos. Ese niño es un rascacielos. ¿Cuál es? 1
13b. Los rascacielos son altos. Esa cosa es un rascacielos. ¿Cuál es? 2

48. Literal option: a skyscraper next to several small houses	49. Distractor–literal: a fence
50. Distractor–metaphorical: three kids sitting in chairs talking	51. Metaphorical option: A very tall boy standing between two short kids

Balloons are round. That kid is a balloon.

14a. Los globos son redondos. Ese niño es un globo. ¿Cuál es? 2
14b. Los globos son redondos. Esa cosa es un globo. ¿Cuál es? 1

Literal option: (NO PICTURE NEEDED)	Distractor–literal: a soap bubble.
52. Distractor–metaphorical: An average size boy, wearing normal clothes.	53. Metaphorical option: A boy in puffy snowsuit, that makes him look quite round, for example: https://fluffouterwear.com/wp–content/uploads/2020/11/201010_Fluff_D1_2467.jpg

Computers do math very fast. That kid is a computer!

15a. Los ordenadores hacen mates muy rápido. Ese niño es un ordenador. ¿Cuál es? 1
15b. Los ordenadores hacen mates muy rápido. Esa cosa es un ordenador. ¿Cuál es? 2

54. Literal option: A desktop computer with spreadsheet on the screen	55. Distractor–literal: a blender
56. Distractor–metaphorical: A boy sitting at a desk, painting	57. Metaphorical option: A young boy sitting at a desk, working on an assignment. He can have a thought bubble with a math problem like 30 + (30 x 4) = 150

Las fresas son rojas. Ese niño es una fresa! [kid with red cheeks]. *Strawberries are red. That kid is a strawberry.*

16a. las fresas son rojas. Ese niño es una fresa. ¿Cuál es? 2
16b. Las fresas son rojas. Esa cosa es una fresa. ¿Cuál es? 1

58. Literal option: a strawberry	59. Distractor–literal: a green pepper
60. Distractor–metaphorical: The face of a boy with normal pale cheeks.	61. Metaphorical option: The face of a boy who has very red cheeks (like rosy from cold or blushing)

Los ríos tienen mucha agua. El niño es un río! (a kid soaked in water). *Rivers have tons of water. That kid is a river. [kid soaked in water]*

17a. Los ríos tienen mucha agua. Ese niño es un río. ¿Cuál es? 1
17b. Los ríos tienen mucha agua. Esa cosa es un río. ¿Cuál es? 2

62. Literal option: a river	63. Distractor–literal: a road
Distractor–metaphorical: a boy with something special about his wear, like a pirate costume.	64. Metaphorical option: A boy soaked in a lot of water (clothes soaked), standing in a puddle, drops of water falling to the floor

Spinning tops spin a lot. That kid is a spinning top.

18a. Las peonzas dan muchas vueltas. Ese niño es una peonza. ¿Cuál es? 2
18b. Las peonzas dan muchas vueltas. Esa cosa es una peonza. ¿Cuál es? 1

65. Literal option: a spinning top (it would be good to have some lines indicating that it is in motion)	66. Distractor–literal: A boy going down a slide
Distractor–metaphorical: a sewing wheel (also spins)	67. Metaphorical option: A boy doing a cartwheel.

Scissors cut stuff. That kid is a scissor!

19a. Las tijeras cortan cosas. Ese niño es una tijera. ¿Cuál es? 1
19b. Las tijeras cortan cosas. Esa cosa es una tijera. ¿Cuál es? 2

68. Literal option: a pair of scissors cutting a big piece of paper	69. Distractor–literal: a hammer, hammering a nail
70. Distractor–metaphorical: a child carrying some boxes	71. Metaphorical option: A boy tearing a big piece of paper with his hands (the tear should be quite straight, though doesn't need to be perfect).

Racing *cars go very fast. That kid is a racing car!*

20a. Los coches de carreras van muy rápido. Ese niño es un coche de carreras. ¿Cuál es? 2
20b. Los coches de carreras van muy rápido. Esa cosa es un coche de carreras. ¿Cuál es? 1

72. Literal option: A race–car (like a formula 1 car)	73. Distractor–literal: A skateboard
74. Distractor–metaphorical: A boy making a handstand	75. Metaphorical option: A boy running very fast (would be good to have a sense of movement)