The Development of Simile Comprehension: From Similarity to Scalar Implicature

Article *in* Child Development · January 2021

DOI: 10.1111/cdev.13507

3 authors, including:

Madeleine Long
Max Planck Institute for Psycholinguistics
18 PUBLICATIONS **133** CITATIONS
SEE PROFILE

Paula Rubio-Fernández
Massachusetts Institute of Technology
62 PUBLICATIONS **727** CITATIONS
SEE PROFILE

The development of simile comprehension: From similarity to scalar implicature

Madeleine Long[1], Vishakha Shukla[1] and Paula Rubio-Fernández[1,2]*

[1]University of Oslo, Oslo, Norway; [2]Massachusetts Institute of Technology, Cambridge, USA.

Similes require two different pragmatic skills: appreciating the intended similarity and deriving a scalar implicature (e.g., 'Lucy is like a parrot' normally implies that Lucy is not a parrot), but previous studies overlooked this second skill. In Experiment 1, preschoolers (N=48; ages 3-5) understood 'X is like a Y' as an expression of similarity. In Experiment 2 (N=99; ages 3-6, 13) and Experiment 3 (N=201; ages 3-5 and adults), participants received metaphors ('Lucy is a parrot') or similes ('Lucy is like a parrot') as clues to select one of three images (a parrot, a girl or a parrot-looking girl). An early developmental trend revealed that 3-year-olds started deriving the implicature 'X is not a Y', while 5-year-olds performed like adults.

Keywords: simile, metaphor, scalar implicature, scalar terms, pragmatic reasoning.

*Correspondence to: paula.rubio-fernandez@ifikk.uio.no

RUNNING HEAD: *Scalar implicature in simile comprehension*

Acknowledgements

This research was supported by a FRIPRO Grant from the Research Council of Norway (Ref. 275505) awarded to PRF. The authors gratefully acknowledge this funding.

Children's ability to compare members of different categories has been studied in various areas of communicative development, ranging from early overextensions (e.g., referring to a horse as a 'doggy'; Clark, 1973; Bloom, 2002; Gershkoff-Stowe et al., 2006) to pretend play (e.g., playfully referring to a bucket as a 'hat'; Winner et al., 1979, Hudson & Nelson, 1984; Vosniadou, 1987a) and, centrally, to the comprehension of similes and metaphors (Vosniadou et al., 1984; Seidenberg & Bernstein, 1986; Di Paola et al., 2019). Both similes and metaphors require comparing unlike things (as in examples (a) and (b) below), whereas literal comparisons and categorization statements apply to more comparable entities (as in (c) and (d)):

a. Fred is like a lion.

b. Fred is a lion.

c. Fred is like his brother.

d. Fred is an engineer.

Studies on metaphor and simile comprehension report that for young children, similes are easier to interpret than metaphors, probably because the comparison is made explicit (Reynolds & Ortony, 1980; Vosniadou & Ortony, 1983; Vosniadou et al., 1984; Seidenberg & Bernstein, 1986; Siltanen, 1990; Happé, 1995; Norbury, 2005). Children's understanding of similes is related to the development of figurative language (Vosniadou & Ortony, 1983) and analogy (Vosniadou, 1995), both of which require pragmatic reasoning: what is the similarity or the parallel intended in a simile or an analogy? There are, for example, different ways in which Fred could be like a lion: his hair may be long and messy, resembling a lion's mane, or he may be strong like a lion (Rubio-Fernandez & Grassmann, 2016).

Rubio-Fernandez et al. (2017) have recently proposed that similes (or comparison statements, more generally) involve another form of pragmatic reasoning that has been

overlooked in the literature: namely, the derivation of a scalar implicature. Thus, when someone compares Fred to a lion (as in 'Fred is like a lion'), or makes a literal comparison (as in 'Fred is like an engineer'), they are normally understood to imply that Fred is not a real lion (or a real engineer); otherwise, the speaker should have used the corresponding categorization statement (i.e. 'Fred is a lion' or 'Fred is an engineer'). In this view, categorization statements of the form 'X is a Y' and comparisons of the form 'X is like a Y' form a scale in which comparisons are weaker than categorizations, such that their use may imply that the stronger statement does not apply (i.e. describing X as being 'like a Y' would imply that X is not a Y). The aim of this study was to investigate preschoolers' interpretation of similes not only as expressions of similarity (in line with traditional studies of figurative language), but also as scalar expressions that can license the derivation of quantity implicatures (Rubio-Fernandez et al., 2017).

The development of simile interpretation

The relation between metaphor and simile has been intensely contested in rhetorical theory and psycholinguistics, with accounts in the Aristotelian tradition claiming that metaphors are implicit or elliptical similes, whereas categorization accounts argue that nominal metaphors of the form 'X is a Y' (e.g., 'My love is a rose') are interpreted as categorization statements, not as comparisons (e.g., Glucksberg & Keysar, 1990; Gentner et al., 2001; Glucksberg, 2001, 2003, 2011; Sperber & Wilson, 2008; Carston, 2002, 2010; Barndern, 2012). According to the latter theoretical accounts, metaphors of the form 'X is a Y' are interpreted as categorization statements, in which Y stands for an *ad hoc concept* that is looser than the literal concept, but is characterized by the relevant properties that the speaker is attributing to X (for a computational model of these accounts, see Kao et al., 2014). For example, the ad hoc concept LION* may

refer to both animals and people with mane-like hair, and could be used humorously to refer to Fred on a bad hair day. Importantly, the above theoretical accounts do not treat figurative language as fundamentally different from literal language. In particular, Relevance Theory argues that the interpretation of both literal and figurative expressions is guided by the assumption that the speaker produced a maximally relevant, or maximally informative description (Sperber & Wilson, 2008; Carston, 2002, 2010).

While there has always been controversy concerning the nature of metaphor, a general consensus seems to hold regarding simile, with theoreticians adopting the standard definition (Israel et al., 2004): a simile is 'a figure of speech in which two essentially unlike things are explicitly compared, often in a phrase introduced by *like* or *as*' (American Heritage College Dictionary, online edition). According to Israel and colleagues, the dictionary definition captures the three essential properties of similes: '(i) that they involve some form of comparison, (ii) that this comparison is explicit, and (iii) that the comparison involves entities which are not normally considered comparable — that it is, in some sense, figurative' (pp. 124-125).

Carston and Wearing (2011) argue that, just as there is no sharp distinction between literal and figurative language, the distinction between literal comparisons and similes is not easy to draw either, although clear examples of each type of comparison are not hard to find (e.g., Literal: 'Nina is like an old woman' vs Figurative: 'Nina is like a swan'; p.297). As a test of this distinction, Glucksberg (2001) proposes that similes can be paraphrased as metaphors, whereas literal comparisons cannot be felicitously paraphrased as categorization statements. Thus, both 'My lawyer is like a shark' and 'My lawyer is a shark' are acceptable (if interpreted figuratively), whereas 'Barracudas are like sharks' is acceptable, but 'Barracudas are sharks' is not. According to all theoretical accounts, however, both literal comparisons and similes are

interpreted through the same comparison process, with the literal-figurative distinction depending on the semantic distance between the two concepts being compared rather than on the interpretation mechanism involved.

As mentioned earlier, developmental studies since the 1980s have shown that young children understand similes earlier than they understand the corresponding metaphors (for a review of early studies, see Vosniadou, 1987a). Supporting the view that literal comparisons and similes are interpreted through the same mechanisms, Vosniadou and Ortony (1983) showed that 3-6-year-old children, like adults, did not show a preference for literal or figurative continuations in a comparison task. When given statements of the form 'X is like Y', all age groups were as likely to complete the comparison with a literal continuation (e.g., 'Rain is like snow') than with a figurative continuation (e.g., 'Rain is like tears'), only dispreferring anomalous comparisons (e.g., 'Rain is like a chair').

Özçalışkan and colleagues (2006, 2009) investigated the emergence of comparisons of the form 'X is like Y' in the spontaneous speech of 40 English-speaking children during a period of 2 years: from ages 1;2 to 2;10. The 'X is like Y' construction had an early onset, with children routinely using it by age 2;2. Interestingly, early similarity comparisons were holistic, often highlighting strong overall similarity between objects of the same category. However, by 30 months, most similarity comparisons were between objects that belonged to different categories and focused on a single dimension (e.g., 'Brown crayon is brown like my hair'). Özçalışkan et al. (2006, 2009) interpret this developmental trajectory as evidence that feature-based comparisons act as precursors to more abstract analogical reasoning (Gentner, 1983, 2003).

Children's pragmatic abilities with similes and other figurative uses of language have been investigated for more than four decades. However, in the last 20 years, experimental pragmatics studies have increasingly focused on the derivation of scalar implicatures.

Developmental pragmatics and scalar implicatures

Scalar implicatures are a type of pragmatic inference whereby a speaker uses the weaker of two terms and the listener is entitled to assume that the stronger term does not apply. Consider the following examples:

e. Wilma ate some of the cookies in the jar.

f. Wilma ate all of the cookies in the jar.

If Wilma had eaten all of the cookies, a cooperative speaker should utter (f), from which it follows that if the same speaker chose to utter (e), that must mean Wilma did not eat all of the cookies. Pragmatic theories often treat this meaning as a pragmatic inference and not as part of the semantic meaning of 'some', which is compatible with 'all' (see Geurts (2010) and references therein; cf. Levinson (2000) and Chierchia et al. (2008) for alternative views). For example, if the speaker was not sure whether or not Wilma had eaten all of the cookies, they would utter (e), but in that situation, the listener would not be entitled to infer that the stronger statement did not apply (for empirical evidence for the *competence assumption*, see Goodman & Stuhlmuller, 2013; Rubio-Fernandez et al., 2017).

There is an extensive experimental literature on children's derivation of scalar implicatures (for recent reviews, see Papafragou & Skordos, 2016; Horowitz et al., 2018), which aims to tap into the emergence of the semantics/pragmatics distinction. In other words, when do children start deriving meaning beyond what is literally said and understand what is implied? In

line with early studies in the 1970s and 1980s (e.g., Smith, 1980; Braine & Rumain, 1981), Noveck (2001) observed that 8- and 10-year-old French-speaking children accepted statements such as 'Some elephants have trunks' as true, suggesting a logical interpretation of the quantifier (gloss: some and maybe all elephants have trunks). More recent studies have tried to overcome the limitations of earlier paradigms (e.g., the lack of context and motivation to derive a pragmatic interpretation), revealing an earlier sensitivity to pragmatic inference: under certain experimental conditions, children as young as 5 have been shown to derive scalar implicatures, although not at adult-like levels (e.g., Chierchia et al., 2001; Gualmini et al., 2001; Papafragou & Musolino, 2003; Guasti et al., 2005; Pouscoulous et al., 2007; Katsos & Bishop, 2011).

One argument that has been put forward to explain young children's limitations with scalar implicatures is the need to consider alternative expressions, which in turn requires learning the corresponding scale mates (e.g., 'or' vs 'and': 'Barney will have cake or soda' vs 'Barney will have cake and soda'). Supporting this hypothesis, recent studies have shown an increase in preschoolers' derivation of scalar implicatures when lexical alternatives are clearly contrasted (e.g., Miller et al., 2005; Barner et al., 2011; Skordos & Papafragou, 2016).

The large majority of studies on the acquisition of scalar implicatures have focused on the scale mates 'some' and 'all'. However, recent studies with adults have revealed that this canonical pair is not representative of how scalar terms are interpreted across the board (see Doran et al., 2009, 2012; Degen, 2015; van Tiel et al., 2016, 2019). Given that most developmental studies have focused on 'some' vs 'all', the adult data on scalar diversity call for more work on children's understanding of other scalar terms. This was one of the aims of this study. In particular, we focused on a new type of scalar implicature, which has been overlooked in previous studies on children's acquisition of literal comparisons and similes.

Comparison statements can license scalar implicatures

Scalar terms giving rise to implicatures are characterized by a relation of *unilateral entailment*; that is, a sentence using the weaker term is true whenever the equivalent sentence using the stronger term is true, but not vice-versa. This is clear when comparing canonical examples with 'some' and 'all': whenever Wilma eats all the cookies, it is true that she ate some cookies; but it is not always the case that when Wilma eats some cookies, she ate all the cookies. Rubio-Fernández et al. (2017) have recently argued that the same relation holds between comparison and categorization statements:

g. Betty is like a nurse.

h. Betty is a nurse.

Thus, if Betty is a nurse, she is surely similar to a nurse. However, if Betty was like a nurse, it would not necessarily hold that she was an actual nurse (for semantic diagnostics and experimental evidence with adults, see Rubio-Fernández et al., 2017).

When the construction 'X is like Y' is used to express physical similarity, it may be interchangeable with the expression 'X looks like Y'. This latter construction has been used in developmental studies investigating children's understanding of the appearance-reality distinction (e.g., 'What does this look like?' vs 'What is this really?'; Hansen & Markman, 2005). The present study was not concerned with the appearance-reality distinction, which is a cornerstone of Theory of Mind development. However, what is interesting about the parallels between the constructions 'X is like Y' and 'X looks like Y' is that both may be used in situations where the degree of similarity between X and Y does not warrant the use of a categorization statement (e.g., 'That watch looks like a Rolex, but it's not a Rolex'). Thus, both

expressions of similarity may be used and interpreted as alternatives to categorization statements, resulting in a similarity scale that ranges from 'mere similarity' to class inclusion.

While early studies have compared children's interpretation of similes and metaphors (Reynolds & Ortony, 1980; Vosniadou & Ortony, 1983; Vosniadou et al., 1984; Seidenberg & Bernstein, 1986; Siltanen, 1990; Happé, 1995; Norbury, 2005), this is the first study to investigate whether preschool children appreciate that similes and metaphors (or categorization and comparison statements, more generally) form a scale, which can be used to derive scalar implicatures. In this view, investigating young children's understanding of similes is a promising research avenue since it allows us to compare the emergence of two different pragmatic skills in the comprehension of the same linguistic expression: appreciating the similarity intended by the use of 'X is like a Y', and deriving the scalar implicature 'X is not a Y'.

According to Özçalışkan et al. (2006, 2009), children as young as 2;2 regularly produce comparisons of the form 'Y is like Y' in their spontaneous speech. This suggests that the use and comprehension of similarity expressions has an earlier onset than the derivation of scalar implicatures, which different studies report between 4-10 years of age (Chierchia et al., 2001; Gualmini et al., 2001; Papafragou & Musolino, 2003; Guasti et al., 2005; Miller et al., 2005; Pouscoulous et al., 2007; Barner et al., 2011; Katsos & Bishop, 2011; Skordos & Papafragou, 2016; Horowitz et al., 2018). Given these two developmental trajectories, we predict that young preschool children will first appreciate the similarity expressed by 'X is like a Y' descriptions, before they come to understand that they imply 'X is not a Y'. For example, if someone says that Betty is like a nurse, a young child would understand that Betty is similar to a nurse, but they may not yet understand that the speaker is implying that she is not an actual nurse.

The present study

The main aim of this study was to investigate preschoolers' ability to understand that statements of the form 'X is like a Y' may imply 'X is not a Y'. We hypothesize that the scale formed by categorization and comparison statements may be particularly accessible to preschool children because these two expressions are probably highly frequent in child-directed speech because of their pedagogical function (e.g., 'Whales are mammals' or 'A clementine is like an orange'). In the Gelman corpus from the CHILDES database (MacWhinney, 2000), the frequency count of non-verbal uses of 'like' is 1,035, whereas 'some' (the weaker term of the canonical 'some' vs 'all' scale) has a frequency count of 367. It must be noted, however, that not all non-verbal uses of 'like' are comparison markers (for sample uses of 'like' from the Gelman corpus, and a discussion of polysemy and cross-linguistic issues around the expression of similarity, see the Supplementary Materials). It is therefore possible that an earlier sensitivity to pragmatic inference may be observed with categorization and comparisons than with 'some' and 'all'.

The secondary aim of the study was to test preschoolers' appreciation of the ambiguity between literal and figurative meanings. For example, taken out of context, the sentence 'Fred is a lion' could literally refer to a lion (with a person's name) or metaphorically to a person (with lion-like properties). Analogously, 'Fred is like a lion' could be a simile (figuratively comparing a person with an animal) or a literal comparison (between a specific lion and the lion kind, as in 'Dumbo is like an elephant'). Experimental studies on the development of figurative language have not yet established at what age children become aware of these ambiguities (cf. Vosniadou & Ortony, 1983).

In our experimental tasks, categorization statements such as 'Fred is a lion' could be interpreted literally, as referring to a lion, or metaphorically, as referring to a boy with mane-like

hair. In the absence of context, these descriptions would be ambiguous between the literal and metaphorical readings since both would be maximally informative of their respective referents. Early metaphor studies have shown that, in the absence of context, young children tend to interpret metaphors literally, not showing sensitivity to their figurative meaning – what is known as a *literal bias* (for a review, see Vosniadou, 1987a). Descriptions such as 'Fred is a lion' or 'Lucy is a parrot' may counter children's literal bias by using people's names, rather than pets' names (which in Kao et al.'s (2014) model would be formalized as *prior expectations* over proper names).

Regarding our experimental hypotheses, we made three predictions. First, we predicted that even the youngest preschoolers in our sample (i.e. 3-year-olds in their first year of nursery) would understand comparisons of the form 'X is like a Y' as expressions of similarity. Second, we predicted that literal interpretations of similes (e.g., taking 'Lucy is like a parrot' as referring to a parrot; see Fig. 1) would decrease with increasing age, revealing a developing ability to derive scalar implicatures (i.e. Lucy is not a parrot). Third, in the metaphor condition, we predicted a literal bias across the preschool years (e.g., understanding 'Lucy is a parrot' as a literal description of a parrot, rather than a metaphorical description of a parrot-looking girl). The last prediction is in line with early developmental studies on figurative language comprehension (see Vosniadou, 1987a, 1987b, 1989), although prior knowledge of proper names (gloss: 'Lucy' is a girl's name) may sway children towards the metaphorical interpretation. These experimental predictions are exploratory since children's derivation of scalar implicatures in simile interpretation is a novel research area.

---------- INSERT FIG. 1 AROUND HERE ----------

We tested the first hypothesis in Experiment 1, which was based on the studies by Rubio-Fernandez et al. (2017) with adults. These authors observed that when adults were asked 'Is this animal like a tiger?' when presented with a picture of a tiger, they almost unanimously agreed, responding according to the logical interpretation of 'X is like a Y' (i.e. not deriving the scalar implicature 'X is not a Y'). In order to test whether preschool children also give logical responses in their similarity judgements, Experiment 1 presented preschoolers with a series of three images showing an animal, a child with their faces painted like the animal and a child without face paint (see Fig. 2), and they were asked 'Is this animal/child like a [animal name]?' (e.g., 'Is this animal/girl like a panda?'). If children understand 'X is like a Y' as an expression of similarity, and appreciate the similarities in our visual materials, they should respond positively when asked about the animal-looking child, but negatively when asked about the child without face paint. Crucially, if children respond positively when asked about the animal, they will be interpreting the question logically, like the adults in the study by Rubio-Fernandez et al. (2017).

---------- INSERT FIG. 2 AROUND HERE ----------

In order to test the two remaining predictions, Experiment 2 used a 3-choice task (see also Stiller et al., 2015; Horowitz et al., 2018) embedded in the game 'I spy with my little eye', in which the Experimenter gave the child a clue either in the metaphor or simile form (e.g., 'Lucy is a parrot' or 'Lucy is like a parrot') from which the child had to guess the intended referent (e.g., a parrot, a girl or a parrot-looking girl; see Fig. 1). Experiment 2 used the same design but giving simile clues in all trials. This game was chosen because it is a popular, traditional game that was likely to elicit natural responses without requiring metalinguistic awareness (which has been found to hinder young children's performance in figurative language tasks; see Gibbs, 1994; Pouscoulous, 2011).

Relatedly, the similes and metaphors used in this study were based on *physical similarity*, rather than *psychological similarity* (e.g., the parrot-looking girl had a nose that resembled the parrot's beak, rather than being chatty) because early studies have reported that metaphors that can be appreciated perceptually are easier to comprehend at a younger age. In a classic study, Asch and Nerlove (1960) showed that polysemous words such as 'hard', 'sweet' or 'bright' are first understood as physical properties, and it is not until age 11 that children appreciate their figurative meaning as psychological properties. Likewise, in their longitudinal studies, Özçalışkan et al. (2006, 2009) observed that the majority of young children's spontaneous comparisons using 'like' expressed physical similarity between two objects. Therefore, in order to make our task accessible to preschoolers, we used metaphors and similes that relied on physical similarities between people and animals.

We also followed Skordos and Papafragou (2016) in using a block design, such that the simile trials followed the metaphor trials. This block design was intended to highlight the contrast between the categorization and comparison clues (e.g., 'Lucy is a parrot' vs 'Lucy is like a parrot') so that children more readily appreciated that the two expressions can function as alternatives. According to Skordos and Papafragou (2016), appreciating that the speaker is using one of two alternative terms is fundamental to the derivation of scalar implicatures.

Both the metaphor and simile clues were ambiguous between a literal and a figurative interpretation (gloss: Is Lucy a parrot with a girl's name, or a girl who looks like a parrot?). However, whereas the metaphor clues were completely ambiguous between the two interpretations, in the second block of trials children could disambiguate the simile clue by deriving a scalar implicature of the form 'X is not a Y'. Thus, when interpreted literally, 'Lucy is like a parrot' would be true of both the parrot and the parrot-looking girl (for experimental

evidence with adults, see Rubio-Fernandez et al., 2017). However, from a pragmatics point of view, such a description would only be maximally informative of the parrot-looking girl: if the Experimenter was referring to the parrot, she should have said 'Lucy is a parrot'. The fact that the Experimenter chose the weaker description would imply that the stronger one does not hold (i.e. it is not the case that Lucy is a parrot). Thus, if preschool children are able to derive the scalar implicature 'Lucy is not a parrot' when interpreting the simile 'Lucy is like a parrot', they should select the girl who looks like the parrot, rather than the parrot – even though both are similar to a parrot, literally speaking.

Experiment 1

Methods

Participants

90% of children in the study came from middle- and upper-class Caucasian families living in a small beach town in Asturias, Northern Spain. 5% came from a Roma camp on the outskirts of the town, and 5% came from North African immigrant families (first generation). The town is safe and quiet, with child activities and facilities (e.g., a local park, bicycle trails, a beach and a promenade). The study was conducted between 2016-2019.

Forty-eight children were recruited for the first experiment: 16 3-year-olds in their first year of preschool (M = 3;5, range = 2;9-3;8, 10 girls), 16 4-year-olds in their second year (M = 4;5, range = 3;10-4;7, 9 girls) and 16 5-year-olds in their third and final year of preschool (M = 5;6, range = 4;9-5;9, 8 girls). All children were native Spanish speakers recruited from a school in Northern Spain that serves middle-class families. Ethical approval for the task was obtained

from the University of Oslo and minders of children in each grade were informed of the study. The number of children recruited was determined by the permissions that were obtained.

Materials and procedure

Five sets of 3 photographs were put together including one photograph of an animal, one of a child and one of a similar child whose face was painted like the animal. The face paintings were of clearly recognizable animals selected through piloting: panda, tiger, cat, zebra and rabbit. The pilot study included a sixth item: a butterfly and a girl with a butterfly painted on her face. However, some of the children thought the girl was disguised as a fairy, and so this item was discarded to avoid confusion. The photograph of the child with face paint was always placed in the middle in order to highlight the resemblances with both the animal and the other child, whereas the positions of the animal and the other child were counterbalanced across trials.

Children were asked whether the animal or the child in the picture was like the relevant animal. For example, the Experimenter would point at the picture of a panda and ask the child 'Is this animal like a panda?' The Experimenter would then point at the picture of the panda-looking girl and ask the child 'Is this girl like a panda?', and repeat the question while pointing at the second picture of a girl (see Fig. 2). The Experimenter moved from left to right when pointing at the three pictures in each trial, and noted the children's responses on a score sheet.

Results

Children's responses ('Yes, it is like the animal' or 'No, it is not like the animal') are plotted in Figure 3. To test whether there were any differences in children's responses to the pictures of the animal and the animal-looking child, we used logistic mixed effects regression in R (version 3.5.1; 2018), modelling the outcome variable of response (1=Yes, 0=No) with Picture (Animal,

Animal-looking Child) and Age (3, 4 and 5 years) as fixed effects and Participants and Items as random effects (Model 1). The maximal random effect structure was used for participants and items (Barr et al., 2013), including by-participant random intercepts and slopes for Picture and by-item random intercepts and slopes for Picture and Age and their interaction. Deviation coding was used for Picture (Animal=-.5, Child=.5), while participant age was entered as a scaled continuous predictor. Results showed no main effect of Picture (p=.755) or Age (p=.748), nor was there a Picture x Age interaction (p=.115) (see Table 1 in the Supplementary Materials).

---------- INSERT FIG. 3 AROUND HERE ----------

For each age group (3, 4 and 5-year-olds), we tested the reliability of the positive responses relative to chance for the critical pictures (Animal and Animal-looking Child) and the control items (Neutral Child), accounting for multiple comparisons with Bonferroni corrections (adjusted significance level of p=.005). Results revealed that for each age group, positive responses were reliably above chance for the Animal and Animal-Looking Child pictures, whereas they were significantly below chance in the control items (all p's <.001).

The results of Experiment 1 show that, like the adults in the study by Rubio-Fernandez et al. (2017), preschool children between 3-5 years respond positively to the question 'Is this animal like a panda?' when asked about a panda, and also to the analogous question 'Is this girl like a panda?' when asked about a girl whose face is painted like a panda, revealing no difference between the two picture types. Importantly, children responded negatively when asked about a girl who did not look like a panda, which confirms they were not simply agreeing with the Experimenter across the board. These results show that Spanish-speaking children as young as 3 years understand 'X is like a Y' as an expression of similarity, responding logically to the test question. What remains to be seen is whether preschool children interpret 'X is like a Y' as a

maximally informative description, rather than an underinformative one, when used as a clue in a guessing game. In other words, would pre-schoolers select the panda-looking girl, rather than the panda, when given the clue 'Sonya is like a panda'? That was the aim of the second experiment in the study.

Experiment 2

Methods

Participants

Ninety-nine children were recruited for the second experiment. The first three groups were recruited from preschool: 24 3-year-olds (M = 3;7, range = 3;1-3;11, 15 girls), 18 4-year-olds (M = 4;7, range = 4;3-5;0, 8 girls) and 18 5-year-olds (M = 5;5, range = 5;1-5;11, 7 girls). The other two groups were recruited from primary and middle school, respectively: 19 6-year-olds (M = 6;6, range = 6;0-7;0, 7 girls) and 20 13-year-olds (M = 13;2, range = 12;6-14;9, 10 girls). Children were recruited from the same preschool as in Experiment 1 and from the corresponding primary and middle schools.

Materials and design

Six drawings consisting of 3 figures each were designed for the task and printed in individual booklets. An animal, a child (gender balanced) and an identical child with a feature resembling the animal were included in each drawing. As in Experiment 1, the figure of the child that shared features with the other two figures was always placed in the middle in order to highlight the resemblance with the other two, and the positions of the other child and the animal were counterbalanced across trials. The animal features of the middle figures included: big frog-like eyes, pronounced monkey-like ears, a long giraffe-like neck, sharp shark-like teeth, messy lion-

like hair and a curvy beak-like nose resembling a parrot. The six drawings were presented in the same gender-alternating order to all children. The first block of three trials was administered in the Metaphor condition (i.e. the clue was a categorization statement) and the second block in the Simile condition (i.e. the clue was a comparison statement).

Procedure

The Experimenter played the Spanish equivalent of the game 'I spy with my little eye' individually with children 3-6 years. The Experimenter started the game with the phrase 'I see, I see' (in Spanish: 'Veo, veo'), to which the child is supposed to answer 'What do you see?' ('¿Qué ves?'). The Experimenter then responded with a name and a clue: 'I see Lucy. Let me give you a clue: Lucy is a parrot' or 'Lucy is like a parrot' ('Veo a Luci. Y te doy una pista: Luci es un loro' o 'Luci es como un loro'), depending on the condition. Importantly, the descriptions were not stressed contrastively (for a recent study on the role of contrastive prosody on pragmatic inference, see Kurumada & Clark, 2017). The child then had to guess who Lucy was and point to one of the three figures, which the Experimenter marked on the paper.

The task was administered as a class exercise for the 13-year-olds, who were told that their responses would serve as control data for younger children. These children were given booklets with the drawings and for each drawing, the Experimenter gave the corresponding clue out loud and they marked their choice on the paper. The 13-year-olds were only tested in the Metaphor condition because they were expected to be at ceiling in the Simile condition, whereas a baseline was needed for children's appreciation of the ambiguity inherent in the metaphor trials (gloss: was Lucy a girl who looked like a parrot, or was she a parrot with a girl's name?).

Responses were coded as (a) Literal, (b) Pragmatic and (c) Neither. The label 'literal' referred to the non-figurative interpretation of the categorization statement (e.g., 'Lucy is a

parrot' as referring to the parrot) and the logical interpretation of the comparison (e.g., 'Lucy is like a parrot' as referring to the parrot). The label 'pragmatic' referred to the figurative interpretation of the categorizations (e.g., 'Lucy is a parrot' as referring to the parrot-looking girl) and the scalar-implicature interpretation of the comparisons (e.g., 'Lucy is like a parrot' as referring to the parrot-looking girl, and not the parrot).

Results

Children's responses in the Metaphor and Simile conditions are plotted in Figures 4 and 5, respectively. Given the negligible percentages of Neither responses (M: 3% in the Metaphor condition and 1% in the Simile condition), statistical analyses focused on Literal and Pragmatic responses. Overall, children's preference for Literal and Pragmatic responses were taken as evidence that they understood the comparison to the animal in both the Metaphor and Simile conditions, and therefore disregarded the dissimilar figure. We consider the following analyses exploratory, rather than confirmatory, given the novelty of this research area.

---------- INSERT FIG. 4 AROUND HERE ----------

We analyzed the data from the subset of children presented with both the Metaphor and Simile conditions (ages 3-6, N=79) using logistic mixed effects regression in R (version 3.5.1; 2018). We modelled the outcome variable of interpretation (1=Literal, 0=Pragmatic), with Condition (Metaphor, Simile) and Age (3, 4, 5 and 6 years) as fixed effects and Participants and Items as random effects (Model 2). The maximal random effect structure was used for participants and items (Barr et al., 2013), including by-participant random intercepts and slopes for Condition and by-item random intercepts and slopes for Condition and Age and their interaction. Deviation coding was used for Condition (Metaphor=-.5, Simile=.5), while

participant age was entered as a scaled continuous predictor. Results revealed a main effect of Condition (β= -3.734, p<.001): more literal interpretations were observed in the Metaphor condition, and Age (β=-1.175, p=.0404): literal interpretations decreased with increasing age (see Fig. 6 and Table 2 in the Supplementary Materials). There was also a Condition x Age interaction (β=-3.728, p<.001). To follow-up on this interaction, we conducted separate analyses for the Metaphor condition (Trial block 1) and the Simile condition (Trial block 2), and found no main effect of Age on Metaphor interpretation (β=.717, p=.292), but a main effect of Age on Simile interpretation (β=-3.913, p<.001) whereby literal interpretations of similes decreased with increasing age. Parallel results were observed at the subject level when analysing the three response patterns in the data (i.e. All Literal, All Pragmatic and Mixed; see Model 3 and Table 3 in the Supplementary Materials).

---------- INSERT FIG. 5 AROUND HERE ----------

For each age group (3, 4, 5, 6 and 13-year-olds), we first tested the reliability of the literal responses in the Metaphor condition against chance, accounting for multiple comparisons with Bonferroni corrections (adjusted significance level of p=.01). Two-tailed binomial tests revealed a significant preference for the literal interpretation in children ages 3-6 years (all p's<.001). By contrast, the 13-year-olds were at chance between the literal and pragmatic interpretations, revealing sensitivity to the ambiguity between the two readings.

---------- INSERT FIG. 6 AROUND HERE ----------

We then tested the reliability of the literal responses in the Simile condition against chance for each age group (3, 4, 5 and 6-year-olds), accounting for multiple comparisons with Bonferroni corrections (adjusted significance level of p=.0125). We found that 3-year-olds were significantly above chance (p<.001), revealing a reliable preference for the literal interpretation; 4-year-olds

were at chance between the literal and pragmatic interpretations (p=1.0), and 5- and 6-year-olds were below chance (all p's<.001), showing a reliable preference for the pragmatic interpretation. These results suggest a developmental trend in pre-schoolers' derivation of scalar implicatures when interpreting similes.

The results of the Metaphor condition replicated previous findings of a literal bias in young children, supporting our predictions: preschoolers (ages 3-5) and first graders (age 6) showed a reliable preference for the literal interpretation of the categorization clues, whereas older children (age 13) gave a comparable number of literal and metaphorical responses. It is interesting that preschoolers and first graders did not prefer the parrot-looking girl over the parrot simply because the name in the clue was 'Lucy'. This suggests that the association between proper names and people was not a strong prior in children's reasoning (Kao et al., 2014), possibly because they are used to animals having people's names in children's books or in their own pretend play.

Children in middle school showed more sensitivity to the two possible meanings of the clues. At the individual level, 11 of the 20 middle schoolers gave mixed responses, while 5 selected the literal response and 4 the figurative response in all trials. It must be noted, however, that the fact that some 13-year-olds selected the same response across trials cannot be taken as evidence that they did not appreciate the two possible readings of the description, since they could have adopted an interpretive strategy in the first trial and applied it consistently throughout the task. Therefore, children's appreciation of the ambiguity inherent in the clues is better established at the group level, with 13-year-olds clearly revealing a more mixed pattern of responses than the younger children (see Fig. S1 in the Supplementary Materials).

It must be noted that the clues used in this game were not contextualized and were therefore not a fair assessment of young children's pragmatic abilities with metaphorical language (see Di Paola et al., 2019; Pouscoulous & Tomasello, 2019). What can be concluded from these results, however, is that in the absence of context, it was not until after age 6 (and perhaps not even until middle school) that children revealed sensitivity to the literal-metaphorical ambiguity inherent in the clues.

Regarding the Simile condition, there was an early preference for the literal interpretation that decreased with increasing age, also as predicted. Subject-level analyses revealed that, overall, most responses in Experiment 2 were either consistently literal or consistently pragmatic, revealing a significant Age x Condition interaction. In the Metaphor condition, most responses were consistently literal, with comparable rates across ages 3-6 years. By contrast, in the Simile condition, most responses were consistently pragmatic, revealing a growing trend across these age groups. Subject-level analyses therefore confirm that children responded differently to the categorization clues in the Metaphor condition than to the comparison clues in the Simile condition.

However, the interpretation of these patterns of results requires careful examination. The older children's preference for the character who looked like the animal in the Simile condition might be explained as an effect of the names used in the clues – which were people's names, rather than animals'. However, all preschool children interpreted people's names as referring to the animal in the Metaphor condition, making it unlikely that their performance in the Simile condition was simply determined by the kind of names used in the clues.

The older children's preference for the character who looked like the animal in the Simile condition could have also been driven by a preference for the figure in the middle position, rather

than by a developing ability to derive scalar implicatures. However, a comparison with the first block of trials is also relevant here: given that all children 3-6 years pointed to the animal in the Metaphor condition and its position was counterbalanced across trials, it seems unlikely that what developed with age in the second block of trials was a preference for a specific position in the display. If anything, older children should be more flexible in their responses and more sensitive to the linguistic clues, rather than developing interpretive strategies during the task based on location rather than meaning.

Having said that, it is still possible that the older children appreciated the change from the metaphor condition in the first block to the simile condition in the second, but started selecting the middle figure because they were uncertain about the correct answer to the similes. Thus, whereas placing the picture of the animal-looking child in the middle position may have increased the chances that children noticed the intended similarities with both the animal and the other child in the display, such an experimental design leaves open the possibility that low-level factors affected the results.

Another alternative interpretation of the developmental trend observed in the Simile condition is that it may have been accentuated by the metaphor trials administered in the first half of the task, which may have *primed* the literal response in the younger age groups (a sort of perseverance error), while highlighting the two alternative expressions for the older children (as intended). We therefore ran a follow-up experiment testing kids only in the simile condition to see if they would show an earlier preference for the pragmatic interpretation. We also used photographs, in addition to drawings, in case some of the young children in Experiment 2 had understood the drawings to be similar to the real animals they depicted (gloss: a drawing of a parrot is like a real parrot).

Experiment 3

Methods

Participants

Two-hundred and one participants were recruited for the third experiment. They were divided into two groups: (a) children tested with drawings, including 27 3-year-olds (M = 3;7, range = 2;10-3;11, 14 girls), 31 4-year-olds (M = 4;5, range = 4;0-4;10, 17 girls) and 29 5-year-olds (M = 5;8, range = 5;0-6;5, 14 girls), and (b) individuals tested with photographs, including 29 3-year-olds (range = 2;9-3;8, mean = 3;2, 17 girls), 32 4-year-olds (range = 3;9-4;8 , mean = 4;3, 10 girls), 35 5-year-olds (range = 4;9-5;8, mean = 5;3, 20 girls) and 18 adults (range = 20-23, mean = 20;8, 10 women). Children were recruited from the same preschool as in Experiments 1 and 2. Adults were Spanish university students contacted via email to volunteer as control subjects.

Materials and procedure

The same photographs used in Experiment 1 and the same drawings used in Experiment 2 were used again in Experiment 3, but the clues were all formulated as comparison statements (e.g., 'Lucy is like a parrot'). Children were randomly allocated to one of the two versions of the task (Photographs vs Drawings).

The testing procedure was the same one used in the Simile condition of Experiment 2. For the adults, a PDF was built with the slides used with the children. The clue was written above the photographs and the photographs were labelled A-B-C (see Fig. 2). Adult participants returned their responses to the Experimenter by email. Responses were coded the same as before.

Results

Children's responses in the Drawing and Photograph versions of the task are plotted in Figures 7 and 8, respectively. Given the negligible percentages of Neither responses (M = 6% in the Drawing condition and 1% in the Photograph condition), statistical analyses focused on Literal and Pragmatic responses. Once again, we consider the following analyses exploratory.

---------- INSERT FIG. 7 AROUND HERE ----------

Using logistic mixed effects regression, we modelled the outcome variable of interpretation (1=Literal, 0=Pragmatic) with Presentation (Drawing, Photograph) and Age (3, 4 and 5 years) as fixed effects and Participants and Items as random effects (Model 4). The maximal random effect structure was used for participants and items (Barr et al., 2013), including only by-participant random intercept (as Presentation did not vary by participant), and by-item random intercepts and slopes for Presentation and Age and their interaction. Deviation coding was used for Presentation (Drawing=-.5, Photograph=.5), while participant age was entered as a scaled continuous predictor. There was a main effect of Age (β=-1.613, p<.001): literal interpretations of similes decrease with increasing age (see Table 4 in the Supplementary Materials). There was also a main effect of Presentation (β= -2.339, p= 0.003), with fewer literal interpretations in the Photograph condition. However, the Age x Presentation interaction did not approach significance (β=-0.178, p=.815). Parallel results were observed at the subject level when analysing the three response patterns in the data (i.e. All Literal, All Pragmatic and Mixed; see Model 5 and Table 5 in the Supplementary Materials).

---------- INSERT FIG. 8 AROUND HERE ----------

We tested the reliability of children's literal responses in the Drawing condition against chance for each age group (3, 4 and 5 years), accounting for multiple comparisons with Bonferroni corrections (adjusted significance level of p=.016). Two-tailed binomial tests

revealed that 3-year-olds' performance was not different from chance (p=.060), whereas a reliable rejection of the literal interpretation was observed in 4- and 5-year-olds (all p's<.001).

We then looked at the literal responses in the Photograph condition for each age group (3, 4, 5 years, plus adults), accounting for multiple comparisons (adjusted significance level of p=.0125). In the Photograph condition, all preschool groups were below chance, revealing a reliable preference for the pragmatic interpretation (all p's<.001). Adults always chose the pragmatic interpretation (with the exception of a single response).

The results of Experiment 3 revealed a lower rate of literal interpretations in the younger age groups than Experiment 2. While the results of these two experiments are not directly comparable, the lower rates of literal interpretations in Experiment 3 suggest that the first block of metaphor trials in Experiment 2 might have primed younger children to continue selecting the literal response in the simile trials. Subject-level analyses support this interpretation of the results (see Fig. S2 and Fig. S3 in the Supplementary Materials): in Experiment 2, around 60% of 3-year-olds and 45% of 4-year-olds systematically interpreted similes literally, whereas in Experiment 3, around 15% of 3- and 4-year-olds systematically selected the literal interpretation of the simile clues. These response patterns suggest that the Metaphor condition administered in the first block of trials of Experiment 2 may have primed the younger age groups to select the literal interpretation of the similes in the second block of trials, rather than highlighting the contrast between the two types of clues (as we intended).

The type of visual materials used in the task had a main effect in Experiment 3, with children performing better with photographs than with drawings. Subject-level analyses revealed an increase in consistently-pragmatic responses vs. mixed responses in the Photograph condition compared to the Drawing condition. Interestingly, the 3-year-olds went from chance level in the

version with drawings to a reliable preference for pragmatic interpretations in the version with photographs. These results suggest that some children may have taken the drawings of the animals to be similar to the actual animals depicted. Such interpretation of the results, while clearly speculative, suggests a different type of literal interpretation of the simile clues (gloss: which of these drawings looks more like the real animal?). Overall, preschoolers' derivation of scalar implicatures increased with age, with 5-year-olds showing adult-like performance when tested with photographs.

We interpret children's performance with similes (e.g., 'Lucy is like a parrot') as evidence that they used a scalar implicature (i.e. 'Lucy is not a parrot') to choose between the animal and the child who resembled the animal (i.e. the parrot and the parrot-looking girl). The results of Experiment 1 support this interpretation since preschool children of all ages agreed that both the parrot and the parrot-looking girl were 'like a parrot'. Therefore, selecting the parrot-looking girl over the parrot when interpreting the clue 'Lucy is like a parrot' suggests that children were deriving the scalar implicature 'Lucy is not a parrot' when interpreting the simile as a maximally informative clue.

Discussion

Young children are known to find similes easier than metaphors (Reynolds & Ortony, 1980; Vosniadou & Ortony, 1983; Vosniadou et al., 1984; Siltanen, 1990; Seidenberg & Bernstein, 1986; Happé, 1995; Norbury, 2005). Here preschool and first-grade children (ages 3-6) showed a stronger preference for the literal interpretation of potentially metaphorical statements ('Lucy is a parrot') than for that of similes ('Lucy is like a parrot'), unlike their middle school counterparts (age 13), whose responses reflected the literal-metaphorical ambiguity. These results support the

view that young children suffer from a literal bias when interpreting metaphors out of context, although their abilities with figurative language can be more sophisticated when metaphorical meanings are more accessible in the context (for a review, see Pouscoulous, 2011).

More central to the aim of this study, preschool children's initial preference for the logical interpretation of similes decreased with increasing age, revealing a growing preference for the pragmatically enriched interpretation: if Lucy is described as being 'like a parrot', that means she is not a parrot (otherwise the clue would have been 'Lucy is a parrot'). When testing children with photographs (rather than drawings), even 3-year-olds showed a reliable preference for the pragmatic interpretation, while 5-year-olds' performance was comparable to that of adults. These results confirm that when tested with appropriate materials and protocols, even young preschool children are sensitive to scalar implicatures (see Pouscoulous et al., 2007; Stiller et al., 2015; Skordos & Papafragou, 2016; Horowitz et al., 2018).

Recent metaphor studies employing more child-friendly procedures have also revealed improved performance in preschoolers (Rubio-Fernandez & Grassmann, 2016; Di Paola et al., 2019; Pouscoulous & Tomasello, 2019) relative to the poor metalingustic judgements that were observed in early metaphor studies (see Vosniadou, 1987a). In the present study, the enhanced pragmatic performance observed in the simile condition relative to the metaphor condition might suggest that preschoolers are able to derive scalar implicatures before they can interpret metaphors. However, such an interpretation is likely to underestimate preschoolers' abilities with metaphorical language, which clearly improve when figurative uses are properly contextualized (see Falkum et al., 2017; Köder & Falkum, 2019). Future studies should therefore compare preschoolers' pragmatic abilities when deriving scalar implicatures and interpreting metaphors

using equally suitable tasks that prevent young children from defaulting to a literal interpretation in the absence of context.

The block design of Experiment 2 intended to contrast the two alternative expressions under investigation (e.g., 'Lucy is a parrot' vs 'Lucy is like a parrot'), potentially making the pragmatic interpretation of the similes more accessible (Skordos & Papafragou, 2016). This contrast, however, seems to have gone unnoticed by the younger groups, who continued selecting the literal response in the second block of trials. Interestingly, when the categorization-comparison scale was not highlighted in Experiment 3, preschoolers' performance was not compromised, suggesting scalar implicatures were highly accessible in this task. However, future studies should try to make the contrast between the two alternative expressions more salient (ideally without priming the literal response, as it happened with our block design) to see if they observe even higher rates of pragmatic interpretations in the younger age groups.

Another methodological consideration when interpreting the results of this study is that in all three experiments, the animal-looking figure was placed in the middle of the display, so that children could more easily appreciate its similarity with both the animal and the neutral figure. This experimental design leaves open the possibility that children in the simile condition might have selected the pragmatic response because they had a preference for the middle position in the display, and not because they derived a scalar implicature (that ruled out the animal figure as a possible referent). While not impossible, two patterns of results undermine this alternative interpretation. First, in Experiment 2, children 3-6 years revealed a significant preference for the literal interpretation of the metaphors in the first block of trials, even though the position of the animal figure (i.e. the literal response) was counterbalanced across trials. Second, in both Experiments 2 and 3, children's preference for the pragmatic response in the simile condition (i.e.

the animal-looking figure) increased with age, making it unlikely that what developed with age was a blind preference for the middle position in the display, rather than a pragmatic ability to derive scalar implicatures.

A low-level explanation of these results which cannot be ruled out at present is that children's preference for the middle figure in the simile condition may reveal uncertainty (i.e. a sort of compromise between the two extreme responses). Future studies should therefore determine whether children are able to appreciate the similarities between the animal-looking figure and the other two figures while fully counterbalancing their position in the display (see, e.g., Stiller et al., 2015), and whether such a design compromises their performance in the task.

One methodological reason why the task used in this study may have been particularly easy for preschoolers is that, unlike earlier studies on the acquisition of scalar implicatures (e.g., Noveck, 2001; Papafragou & Musolino, 2003; Barner et al., 2011), it did not require that underinformative descriptions be *rejected* as pragmatically infelicitous (e.g., children did not have to reject 'Lucy is like a parrot' when a puppet predicated that of a parrot). Instead, children had to select the character that *better fit* the description by favoring a pragmatic interpretation over a logical reading (see also Stiller et al., 2015; Horowitz et al., 2018).

While sentence judgement tasks may be harder for preschoolers than reference disambiguation tasks, it should be noted that they are not necessarily a good test of children's abilities with scalar implicatures (for discussion, see Katsos & Bishop, 2011). As for the present results, children seemed to appreciate that both the parrot and the parrot-looking girl were similar to a parrot (Experiment 1), but that the parrot-looking girl was the one who better fit the description 'Lucy is like a parrot' in the guessing game (Experiments 2 and 3). Future studies should try to

establish at what age children start rejecting 'Lucy is like a parrot' as an underinformative description of an actual parrot (e.g., by penalizing the puppet who uses such descriptions).

A number of recent studies have also observed young preschoolers' comprehension of implicatures using arbitrary or *ad hoc* scales (e.g., inferring 'Fred did not eat the whole cupcake' from the utterance 'Fred ate the frosting'; Papafragou & Tantalou, 2004; Stiller et al., 2015; Horowitz et al., 2018). Because these studies used child-friendly paradigms that did not require rejecting underinformative statements, and tried to make alternative expressions available at testing, it is difficult to determine whether the earlier age of success is related to the arbitrary nature of the scales, to the improved experimental methods, or to a combination of the two. However, these studies make an important contribution to the developmental pragmatics literature since they investigated children's comprehension of scalar terms beyond 'some' vs 'all'. Given that experimental research with adults has recently revealed that the interpretation of this canonical pair does not generalize to other scalar terms (Doran et al., 2009, 2012; Degen, 2015; van Tiel et al., 2016, 2019), developmental studies need to start testing children on a wider range of scalar terms, including the scale formed by comparison and categorization statements (Rubio-Fernandez et al., 2017).

The results of Horowitz et al. (2018) are generally consistent with the hypothesis that young children struggle with scalar implicatures because of their limitations with generating alternatives (see also Barner et al., 2011). However, the authors suggest that partial quantifier knowledge is another factor that could explain why young children have trouble deriving scalar implicatures. In this view, young children may have fewer problems deriving implicatures from ad hoc scales because those rely on general informativity expectations, rather than the scalarity of individual terms. Thus, rather than having to bear in mind an alternative to the actual utterance produced by

the speaker (gloss: the speaker could have said 'all of the cookies' instead of 'some of the cookies'), ad hoc scales would require that children derive a scalar implicature on the basis of the amount of information specified in an utterance (gloss: why specify that Fred ate the frosting if he actually ate the whole cupcake?). Relatedly, Papafragou and Skordos (2016) have argued that the first step in being able to derive a scalar implicature is to learn the semantics of the individual terms, and then learn that these alternatives form a scale. This suggests that deriving implicatures with ad hoc scales may be easier for young children, as it relies entirely on pragmatic reasoning, and does not require acquiring alternative meanings.

Despite having to treat categorization and comparison statements as alternatives, the preschool children in the present study were able to derive scalar implicatures at a younger age than in previous studies, with 3-year-olds showing a reliable preference for the pragmatic interpretation in the photograph version of the simile task, and 5-year-olds performing comparably to adults. A possible reason why young children are able to derive higher rates of scalar implicatures with similes than with other scalar expressions is that categorization and comparison statements are likely to be highly frequent in child-directed speech because of their pedagogical function (e.g., 'A surgeon is a doctor' or 'Leopards are like cheetahs') and their acquisition may therefore happen earlier than that of quantifiers (e.g., 'some' vs 'all') or logical connectives (e.g., 'or' vs 'and'). Longitudinal studies by Özçalışkan and colleagues (2006, 2009) have shown that as early as 2;2 years, toddlers start using similarity comparisons of the form 'X is like Y' regularly in their spontaneous speech, suggesting that they must be exposed to this construction from early on. Future research should therefore investigate children's exposure to different scalar terms and its effect on their ability to derive scalar implicatures. In the case of similes and metaphors (or comparisons and categorization statements, more generally), even children in the first year of

preschool can show a reliable preference for the pragmatically enriched meaning, supporting the view that children's documented difficulties with scalar implicatures need not stem from their limited pragmatic abilities, but from the protracted acquisition of alternative expressions and their scalarity.

References

Asch, S. E., & Nerlove, H. (1960). The development of double function terms in children. *Perspectives in Psychological Theory*, 47-60.

Barnden, J. A. (2012). Metaphor and simile: Fallacies concerning comparison, ellipsis, and inter paraphrase. *Metaphor and Symbol, 27*, 265–282. 10.1080/10926488.2012.716272

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 84-93. 10.1016/j.cognition.2010.10.010

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

Bloom, P. (2002). *How children learn the meanings of words*. MIT press. 10.7551/mitpress/3577.001.0001

Braine, M. D., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*, 46-70. 10.1016/0022-0965(81)90003-5

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell Publishing. 10.1002/9780470754603

Carston, R. (2010). Metaphor: Ad hoc concepts, literal meaning and mental images. In *Proceedings of the Aristotelian Society* (Vol. 110, pp. 295-321). Oxford, UK: Oxford University Press. 10.1111/j.1467-9264.2010.00288.x

Carston, R., & Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition*, *3*, 283–312. 10.1515/langcog.2011.010

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University Conference on Language Development* (pp. 157-168).

Chierchia, G., Fox, D. & Spector, B. (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In C. Maienborn, K. von Heusinger and P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 47-62). Mouton de Gruyter. 10.1353/lan.2013.0026

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 acquisition of language* (pp. 65-110). Academic Press.

Degen, J. (2015). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*, 11-1. 10.3765/sp.8.11

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*. 10.1016/j.pragma.2019.08.008

Doran, R., Baker, R., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-unified nature of scalar implicature: An empirical investigation. *International Review of Pragmatics*, *1*, 1-38. 10.1163/187730909X12538045489854

Doran, R., Ward, G., Larson, M., McNabb, Y., & Baker, R. E. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*, *88*, 124-154. 10.1353/lan.2012.0008

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

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170. 10.1016/S0364-0213(83)80009-3

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). MIT Press. 10.7551/mitpress/4117.003.0015

Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In Centner, D., Holyoak, K.J., & Kokinov, B.N. (Eds.), The analogical mind: Perspectives from cognitive science (pp. 199-253). Cambridge MA, MIT Press. 10.7551/mitpress/1251.003.0010

Gershkoff-Stowe, L., Connell, B., & Smith, L. (2006). Priming overgeneralizations in two-and four-year-old children. *Journal of Child Language*, *33*, 461-486. 10.1017/S0305000906007562

Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press. 10.1515/ip-2016-0020

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language and understanding*. Cambridge: Cambridge University Press.

Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford: Oxford University Press. 10.1093/acprof:oso/9780195111095.001.0001

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*, 92-96. 10.1016/S1364-6613(02)00040-2

Glucksberg, S. (2011). Understanding metaphors: The paradox of unlike things compared. In K. Ahmad (Ed.), *Affective computing and sentiment analysis: Emotion, metaphor and terminology* (pp. 1–12). Dordrecht, The Netherlands: Springer. 10.1007/978-94-007-1757-2_1

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review 97*, 3–18. 10.1037/0033-295X.97.1.3

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*, 173-184. 10.1111/tops.12007

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001, October). At the semantics/pragmatics interface in child language. In *Semantics and Linguistic Theory* (Vol. 11, pp. 231-247). 10.3765/salt.v11i0.2840

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*, 667-696. 10.1080/01690960444000250

Hansen, M. B., & Markman, E. M. (2005). Appearance questions can be misleading: A discourse-based account of the appearance–reality problem. *Cognitive Psychology*, *50*, 233-263. 10.1016/j.cogpsych.2004.09.001

Happé, F. G. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbol*, *10*, 275-295. 10.1207/s15327868ms1004_3

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development*, *89*, 572-593. 10.1111/cdev.13014

Hudson, J., & Nelson, K. (1984). Play with language: Overextensions as analogies. *Journal of Child Language*, *11*, 337–346. 10.1017/S0305000900005808

Israel, M., J. Harding & V. Tobin. (2004). On simile. In M. Achard & S. Kemmer (Eds.), *Language, culture, and mind* (pp. 123–135). Stanford: CSLI Publications.

Kao, J. T., Bergen, L., & Goodman, N. D. (2014). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 719–724).

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67-81. 10.1016/j.cognition.2011.02.015

Köder, F., & Falkum, I. L. (2019). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics*. 10.1016/j.pragma.2019.07.007

Kurumada, C., & Clark, E. V. (2017). Pragmatic inferences in context: Learning to interpret contrastive prosody. *Journal of Child Language, 44*, 850–880. 10.1017/S0305000916000246

Levinson, S. C. (2000). Presumptive meanings: The theory of generalized implicature. MIT Press. 10.7551/mitpress/5526.001.0001

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd Edition). Mahwah, NJ: Lawrence Erlbaum Associates

Miller, K., Schmitt, C., Chang, H., & Munn, A. (2005). Young children understand some implicatures. In *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 389–400).

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*, 383-399. 10.1348/026151005X26732

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*, 165-188. 10.1016/S0010-0277(00)00114-1

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, *86*, 161–180. 10.1037/0033-295X.86.3.161

Özçalıskan, S., & Goldin-Meadow, S. (2006). "X Is like Y": The emergence of similarity mappings in children's early speech and gesture. In G. Kristiansen, M. Achard, R. Dirven, and F. J. Ruiz, *Cognitive Linguistics: Current Applications and Future Perspectives* (pp. 229–60) Berlin: de Gruyter.

Özçalışkan, Ş., Goldin-Meadow, S., Gentner, D., & Mylander, C. (2009). Does language about similarity play a role in fostering similarity comparison in children? *Cognition*, *112*, 217-228. 10.1016/j.cognition.2009.05.010

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics/ pragmatics interface. *Cognition*, *86*, 253-282. 10.1016/S0010-0277(02)00179-8

Papafragou, A., & Skordos, D. (2016). Scalar implicature. *The Oxford Handbook of Developmental Linguistics* (pp. 611-632). Oxford: Oxfor University Press. 10.1093/oxfordhb/9780199601264.013.26

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition, 12*, 71–82. 10.1207/s15327817la1201_3

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

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*, 347 375. 10.1080/10489220701600457

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

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reynolds, R. E., & Ortony, A. (1980). Some issues in the measurement of children's comprehension of metaphorical language. *Child Development*, 1110-1119. 10.2307/1129551

Rubio-Fernandez, P., Geurts, B., & Cummins, C. (2017). Is an apple like a fruit? A study on comparison and categorisation statements. *Review of Philosophy and Psychology*, *8*, 367-390. 10.1007/s13164-016-0305-4

Rubio-Fernandez, P., & Grassmann, S. (2016). Metaphors as second labels: Difficult for preschool children? *Journal of Psycholinguistic Research*, *45*, 931-944. 10.1007/s10936-015-9386-y

Seidenberg, P. L., & Bernstein, D. K. (1986). The comprehension of similes and metaphors by learning disabled and non learning-disabled children. *Language, Speech, and Hearing Services in Schools*, *17*, 219-229. 10.1044/0161-1461.1703.219

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

Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, *30*, 191–205. 10.1016/0022-0965(80)90057-0

Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. *The Cambridge Handbook of Metaphor and Thought* (pp. 84–105). Cambridge University Press. 10.1017/CBO9780511816802.007

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad hoc implicature in preschool children. *Language Learning and Development, 11*, 176–190. 10.1080/15475441.2014.927328

Siltanen, S. A. (1990). Effects of explicitness on children's metaphor comprehension. *Metaphor and Symbol*, *5*, 1-20. 10.1207/s15327868ms0501_1

Van Tiel, B., Pankratz, E., & Sun, C. (2019). Scales and scalarity: Processing scalar inferences. *Journal of Memory and Language*, *105*, 93-107. 10.1016/j.jml.2018.12.002

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, *33*, 137-175. 10.1093/jos/ffu017

Vosniadou, S. (1987a). Children and metaphors. *Child Development*, *58*, 870–885. 10.2307/1130223

Vosniadou, S. (1987b). Contextual and linguistic factors in children's comprehension of nonliteral language. *Metaphor and Symbol*, *2*, 1–11. 10.1207/s15327868ms0201_1

Vosniadou, S. (1989). Context and the development of metaphor comprehension. *Metaphor and Symbolic Activity*, *4*, 159–171. 10.1207/s15327868ms0403_4

Vosniadou, S. (1995). Analogical reasoning in cognitive development. *Metaphor and Symbol*, *10*, 297–308. 10.1207/s15327868ms1004_4

Vosniadou, S., & Ortony, A. (1983). The emergence of the literal-metaphorical-anomalous distinction in young children. *Child Development*, *54*, 154–161. 10.2307/1129872

Vosniadou, S., Ortony, A., Reynolds, R., &Wilson, P. (1984). Sources of difficulty in children's understanding of metaphorical language. *Child Development*, *55*, 1588–1606. 10.2307/1130028

Winner, E., McCarthy, M., Kleinman, S., & Gardner, H. (1979). First metaphors. *New Directions in Child Development*, *3*, 29–41. 10.1002/cd.23219790305

Zharikov, S., & Gentner, D. (2002). Why do metaphors seem deeper than similes? In *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 976–981). 10.4324/9781315782379-203

Figures

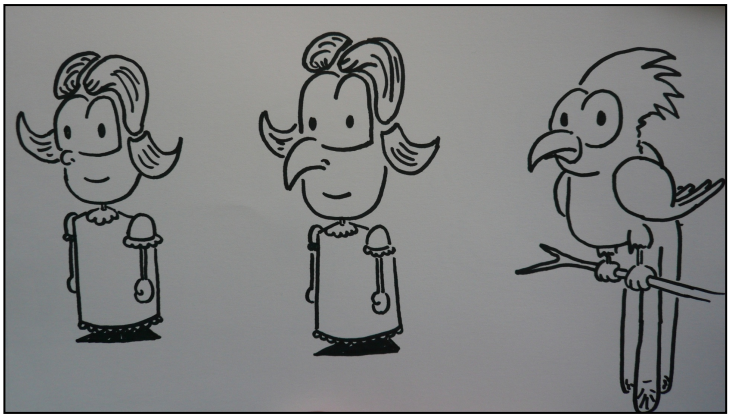

Fig. 1: Sample item for the game 'I spy with my little eye' in Experiment 2, in which the Experimenter gave the child a clue in the metaphor form ('Lucy is a parrot') or the simile form ('Lucy is like a parrot'). Rates of animal selections were used to measure a literal bias in the Metaphor condition and a logical bias in the Simile condition.

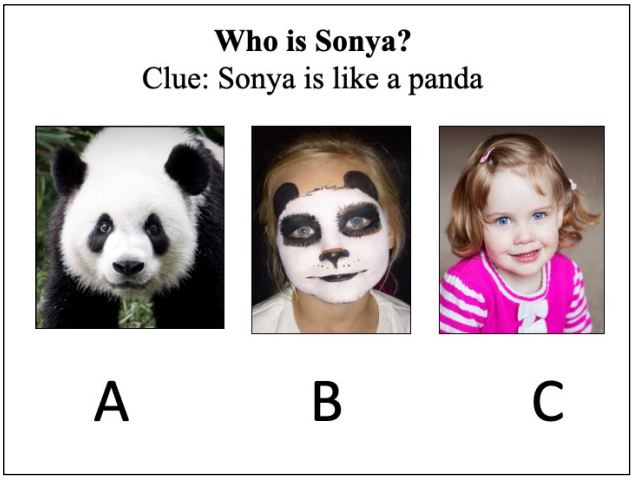

Fig. 2: Sample item from Experiments 1 and 3. The test question in Experiment 1 was 'Is this animal/child like a panda?'. In Experiment 3, the clue was 'Sonya is like a panda' and the participant had to guess who Sonya was.

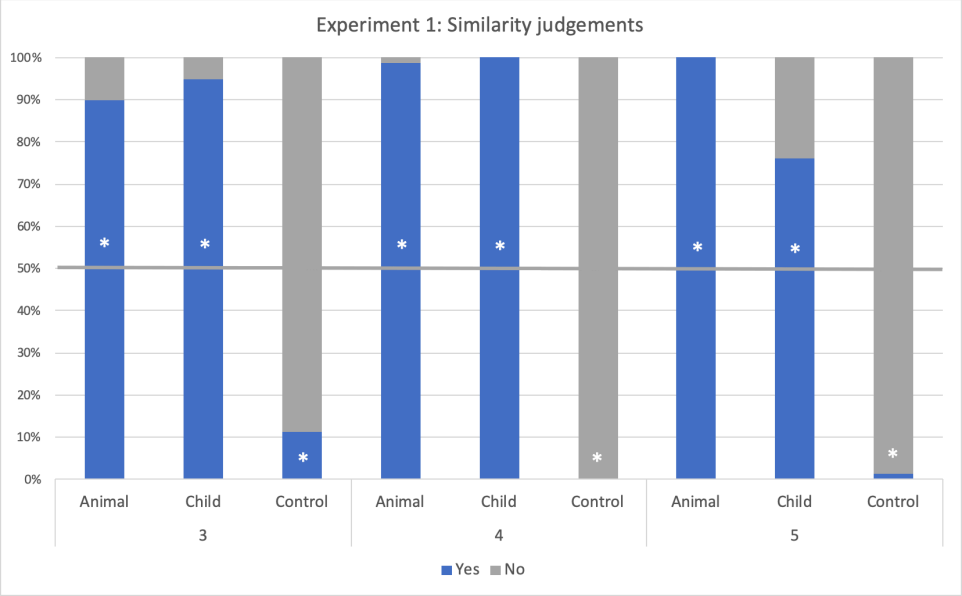

Fig. 3: Percentages of 'Yes' (*it is like an* X) and 'No' (*it is not like an* X) responses for each picture type and age group. Asterisks indicate a reliable preference for (above chance) or rejection of (below chance) the positive response ($p < .001$).

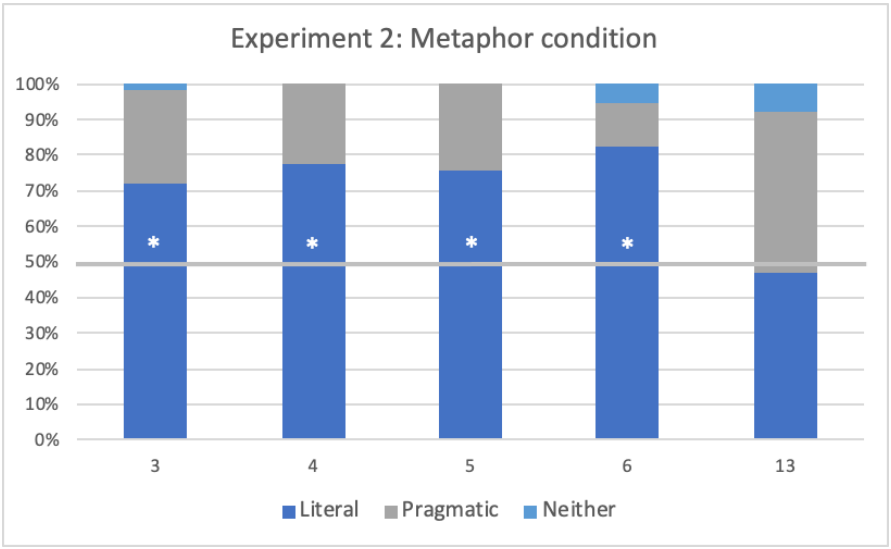

Fig. 4: Percentages of Literal, Pragmatic and Neither responses in the Metaphor condition (Trial block 1) for each age group. Asterisks indicate a reliable preference for the literal interpretation ($p < .001$).

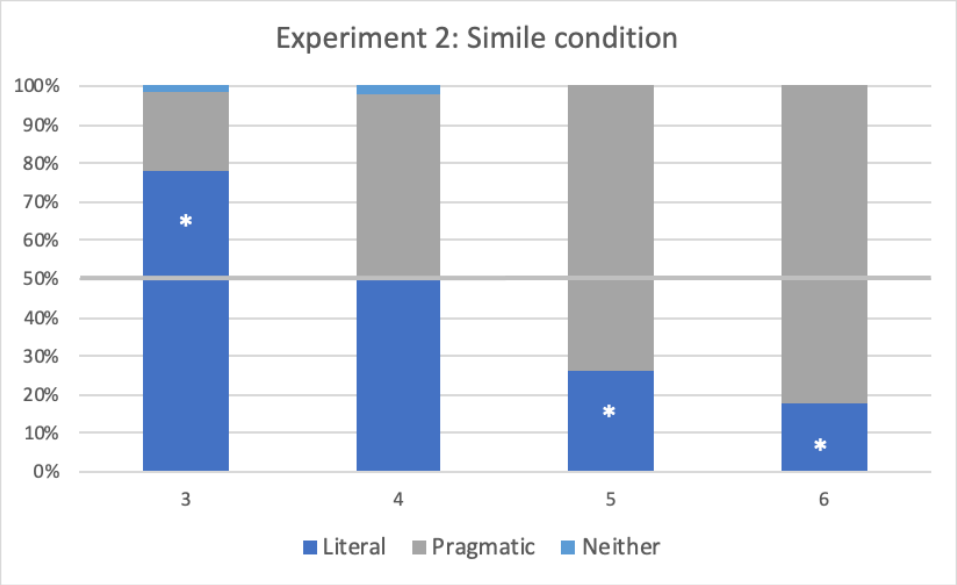

Fig. 5: Percentages of Literal, Pragmatic and Neither responses in the Simile condition (Trial block 2) for each age group. Asterisks indicate a reliable preference (above chance) or rejection (below chance) of the literal interpretation ($p < .001$).

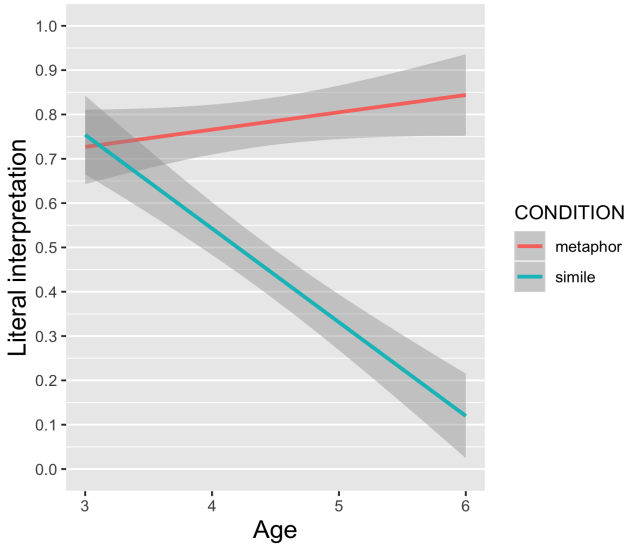

Fig. 6: Regression lines showing literal interpretations by age in the two conditions of Experiment 2. The shaded bands around the regression lines represent a 95% confidence region for the regression fit.

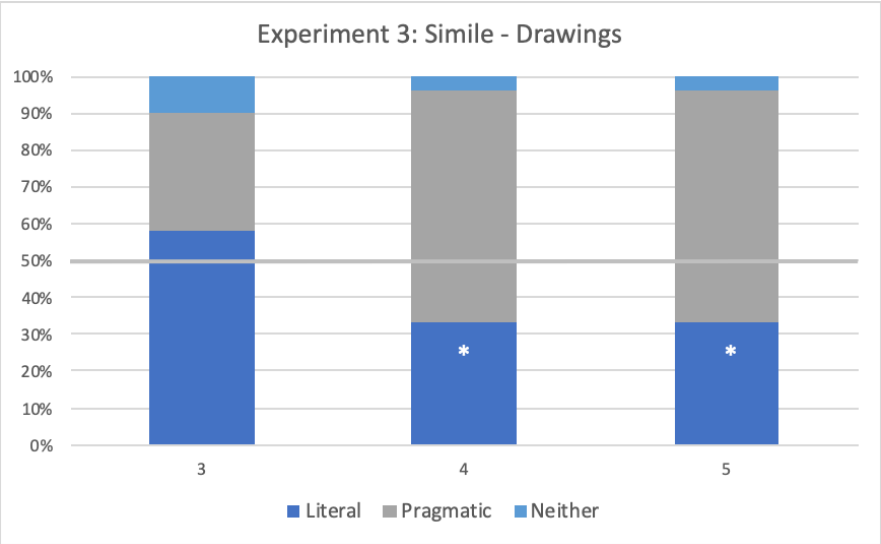

Fig. 7: Percentages of Literal, Pragmatic and Neither responses in the Drawings condition for each age group. Asterisks indicate a reliable rejection of the literal interpretation (below chance; $p <$ .001).

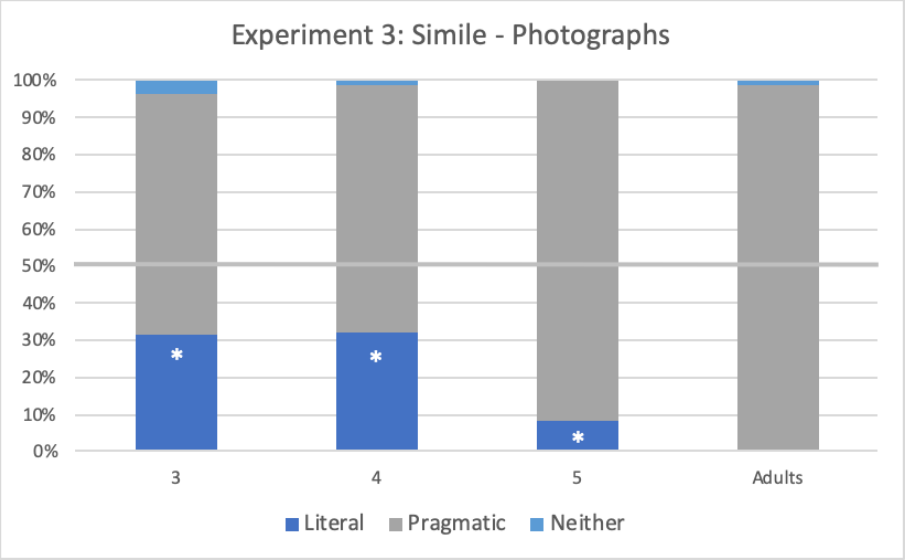

Fig. 8: Percentages of Literal, Pragmatic and Neither responses in the Photographs condition for each age group. Asterisks indicate a reliable rejection of the literal interpretation (below chance; $p < .001$).