4- and 5-Year-Olds' Comprehension of Functional Metaphors

Rebecca Zhu (rebeccazhu@berkeley.edu)
Mariel K. Goddu (goddu@berkeley.edu)
Alison Gopnik (gopnik@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94720 USA

Abstract

Previous work suggests that children's ability to understand metaphors emerges late in development. Researchers argue that children's initial failure to understand metaphors is due to an inability to reason about shared relational structures between concepts. However, recent work demonstrates that causal framing facilitates preschoolers' relational reasoning. Might causal framing also facilitate preschoolers' metaphor comprehension? In Experiment 1, we presented 128 4- to 5-year-olds with a novel metaphor comprehension task, following a causal warm-up task, control warm-up task, or no warm-up task. In the novel comprehension task, preschoolers rated functional metaphors and nonsense statements as smart or silly, and provided explanations. Preschoolers ranked metaphors as "smarter" than nonsense statements, and a quarter of preschoolers provided functional explanations. There was no effect of warm-up tasks. In Experiment 2, we validated the metaphor comprehension task with adults. Overall, the current work presents a new paradigm that demonstrates preschoolers' capacity to understand functional metaphors.

Keywords: metaphor; relational reasoning; language acquisition

Introduction

A metaphor is a figurative utterance that directly compares a concept from one domain to another concept in an unrelated domain. Metaphors are found in everyday speech (e.g. "I got lost in a *sea* of people") as well as famous creative works (e.g. Shakespeare's "if music be the food of love, play on"). Metaphors facilitate communication and provide frameworks for reasoning about abstract concepts (Camp, 2009; Thibodeau, Hendricks, & Boroditsky, 2017), influencing how humans attend to, remember, and process information. Metaphors are also a force for creative change: metaphors can facilitate the development of scientific theories (Kuhn, 1993) and the creation of new word meanings (Bowdle & Gentner, 2005; Camp, 2006; Holyoak & Stamenkovíc, 2018).

However, metaphors appear to be difficult for children to understand and use competently. Previous work on the development of metaphor comprehension argues that children understand metaphors in an adult-like fashion late in development, possibly not until adolescence (Demorest, Silberstein, Gardner, & Winner, 1983; Silberstein, Gardner, Phelps, & Winner, 1982; Winner, Engel, & Gardner, 1980). In this literature, some researchers argue that *relational reasoning* - the ability to attend to similarities based on abstract relations between objects, rather than to the features of individual objects - underlies metaphor comprehension (Bowdle & Gentner, 2005; Gentner, 1988; Roberts & Kreuz, 1994; Wolff & Gentner, 2011; though for a review of metaphor comprehension theories, including alternatives to the relational reasoning account, see Holyoak & Stamenkovíc, 2018). On the relational reasoning account, children's failure to understand metaphors may be attributed at least in part to their inability to notice relational structures, such as functional similarities, between concepts (Gentner, 1988). For example, Silberstein and colleagues (1982) show that 6-year-olds prefer metaphors based on perceptual dimensions such as shape and colour, whereas adults prefer deeper conceptual metaphors. In this study, participants were asked to complete sentences (e.g. "The volcano is…). Children tended to select featural completions (e.g. "a bright firetruck") and adults tended to select conceptual completions (e.g. "a very angry man"). Similarly, Gentner (1988) asked 4- to 5-year-olds, 7- to 8-year-olds, and adults to select an explanation for why two objects were alike (e.g. "Why is a cloud like a sponge?"), and found that preference for functional explanations (e.g. "they both hold water") over perceptual explanations (e.g. "they are both soft and fluffy") increased with age.

Though some previous work demonstrates early metaphor comprehension in preschoolers (Pouscoulous & Tomasello, in press; Vosniadou & Ortony, 1983), these tasks involve metaphors based on perceptual similarities (e.g. "Moons are cookies"; "Eyes are buttons"). However, adults find metaphors based on perceptual similarities unsatisfying, preferring metaphors based on abstract functional similarities instead (Gentner & Clement, 1988). Thus, there is little evidence from the metaphor literature demonstrating that preschoolers can reason about abstract relations and understand metaphors in an adult-like fashion.

However, recent work on relational reasoning shows evidence for the capacity to representing abstract relations quite early in ontogenesis, in preschoolers (Christie & Gentner, 2014; Goddu, Lombrozo, & Gopnik, in press; Hochmann et al., 2017), toddlers (Walker, Bridgers, & Gopnik, 2016; Walker & Gopnik, 2017), and even infants (Anderson, Chang, Hespos, & Gentner, 2018; Hochmann, Mody, & Carey, 2016). This work suggests that the cognitive mechanisms underlying metaphor comprehension might be in place much earlier than previously supposed, and that children might show earlier competence at metaphor comprehension given different experimental methods. We designed a new method that asks children to provide absolute judgments about the validity of metaphors and nonsense statements, rather than assessing children's

metaphor comprehension through a preference task (e.g. Gentner, 1988). Moreover, causal frameworks can induce a relational mindset in young children. (Goddu et al., in press; Walker et al., 2016; Walker & Gopnik, 2018). Thus, we explored whether causal framing might facilitate preschoolers' metaphor comprehension.

Experiment 1

In Experiment 1, we designed a novel metaphor comprehension task and tested whether causal framing would facilitate performance on this task. Children in the causal framing condition received a warm-up task involving the causal transformation of objects on a conveyor belt, whereas children in the control conditions received a similar non-causal warm-up task or no warm-up task. Then, all children were given a novel metaphor comprehension task, in which they must make absolute judgments – that is, "smart" or "silly" ratings – of functional metaphors and nonsense statements. Moreover, given that some previous research suggests that preschoolers prefer literal to non-literal language (Reynolds & Ortony, 1980; but see also Winner et al., 1976 for counterevidence), we ran a causal condition with similes (e.g. "Roofs are *like* hats") as well as a causal condition with metaphors (e.g. "Roofs are hats").

Methods

Participants. We adhered to a stopping rule of 32 children per condition, leading to a total of 128 4- to 5-year-olds who participated in the study (M = 4.86 years; SD = .51 years; 67 females). Researchers tested an additional two children, whose data were excluded due to failure to complete the study (one child) and external interference (one child). Children were recruited and tested in a quiet preschool or museum setting. All experiments reported in this paper were approved by the university's Committee for the Protection of Human Subjects. All parents of child participants provided informed consent.

Stimuli and Procedure. The experimenter presented participants with short stories on a laptop computer. Each child participated in one of four conditions: the Causal Metaphor condition, the Causal Simile condition, the Control Simile condition, and the Baseline Simile condition. Each condition presented participants with different kinds of training trials. Then, during the test trials, all participants were presented with metaphors and nonsense statements, and had to differentiate between the two kinds of utterances.

Causal Metaphor Training Trials. In the Causal Metaphor training trials, participants saw the components of the metaphor in a causal context, specifically as objects undergoing causal transformations. The experimenter introduced the training trials by saying, "Hi! I'm going to tell you about a person named Annie! Annie works in a factory with a super cool purple machine. Let's watch Annie use the purple machine and see what happens."

Each training trial presented participants with two metaphors. During the first part of the training trial, participants saw an object (e.g. a bird) on the left side of a purple conveyor belt. The experimenter pointed and named the object (e.g. "Look! Annie has a bird!") The object traveled down the conveyor belt, and in the middle of the conveyor belt, a purple box came down and covered the object. When the purple box went up again, it revealed another object (e.g. a hot air balloon). The second object then traveled to the right side of the conveyor belt. Finally, the experimenter used the two objects from the conveyor belt in a metaphoric utterance (e.g. "Annie says, 'Birds are hot air balloons!'")

During the second part of the training trial, participants saw a new object (e.g. a sleeping bag) on the right side of the conveyor belt. Two objects appeared below the conveyor belt, one that was a functional match (e.g. a glove) and one that was an object match from the previous trial (e.g. a hot air balloon). The experimenter pointed to and named the object, and prompted participants to find a match for the object on the conveyor belt (e.g. "Look! Annie has a sleeping bag! This time, Annie is going to use the machine on the sleeping bag. Do you think the sleeping bag is going to turn into a glove or a hot air balloon?") After the participant made a prediction by selecting one of the objects below, the participant received feedback: the new object (e.g. the sleeping bag) went down the conveyor belt, which always causally transformed the object into its function-matched counterpart (i.e. a glove), regardless of what object the participant chose. To end the trial, the experimenter used the two objects from the conveyor belt in a metaphoric utterance (e.g. "Annie says, 'Sleeping bags are gloves!'").

Each participant received four training trials with a total of eight metaphors. Each trial's structure followed the design described above, in which the participant watched an object go down the conveyor belt, and then was asked to predict, and received feedback on, what the novel object on the conveyor belt will turn into. The order of the four training trials was randomized and the left-right placement of the function match and the object match was counterbalanced across participants. The experimenter pointed to the objects on the screen (e.g. bird, hot air balloon, glove, sleeping bag) as she named them.

Causal Simile Training Trials. The Causal Simile training trials were identical to the Causal Metaphor training trials, except all utterances were similes (e.g. "Annie says, 'Birds are *like* hot air balloons'") rather than metaphors. Given that some previous work suggests that young children may have difficulty with non-literal language (Reynolds & Ortony, 1980), we ran the Causal Simile condition as well as the Causal Metaphor condition to see whether literal, as opposed to non-literal, statements might increase the accuracy of participants' responses.

Control Simile Training Trials. The Control Simile training trials were identical to the Causal Simile training trials, except that the objects were not presented in a causal context. Thus, there was no conveyor belt. Rather, Annie simply uttered statements about objects that appeared on the screen, providing participants with the same statements about objects, but without causal framing. During the

second part of the training trial, when prompting participants to match the initial object with either a function match or an object match, the experimenter asked what the object was more similar to rather than what the object would turn into (e.g. "Do you think the sleeping bag *is like* a glove or a hot air balloon?"), since the objects did not causally transform into one another.

Baseline Simile Condition. In the Baseline Simile condition, participants were not presented with training trials. Instead, participants in this condition participated in the test trials without any previous training.

Test Trials. The experimenter introduced the test trials by saying, "Now let's play a new game. In this game, we're going to play with Annie's friend Meg. Meg is going to say things and we need your help figuring out whether what Meg said is smart or silly!" The experimenter pointed at a green happy face on the computer screen while saying "smart" and a red sad face on the computer screen while saying "silly". Then, the experimenter showed Meg with two objects (e.g. a roof and a hat) and said, "Meg says, 'Roofs are hats!' Is what Meg said smart or silly?". The experimenter pointed to the objects on the screen as she named them, and to the happy face and the sad face while saying "smart" and "silly" respectively. Once the participant answered by providing a verbal response (e.g. "I think it's smart") or pointing at the happy or sad face, the experimenter began the next trial. No feedback was provided.

Additionally, the last trial was always a metaphor. On the last trial, after participants had provided a smart/silly response, the experimenter asked for an open-ended explanation about the similarity between the two components of the metaphor (e.g. "How are windows like eyes? How are these two things alike?").

There were sixteen test trials total, and two kinds of test trials: eight metaphors (e.g. "Clouds are sponges"; "Tires are shoes") and eight nonsense statements (e.g. "Dogs are scissors"; "Pennies are sunglasses"). We counterbalanced whether participants received a metaphor or nonsense statement first. In order to minimize executive function demands that could influence metaphor comprehension (Ballestrino et al., 2016), the "smart" option (happy face) was always on the right and the "silly" option (sad face) was always on the left. Trial order was semi-randomized, such that no more than three of the same kind of trial appeared consecutively, and the last trial was always a metaphor.

In the Causal Metaphor condition, all statements were presented non-literally (e.g. "Clouds are sponges") whereas in the Causal Simile, Control Simile, and Baseline Simile conditions, all statements were presented literally (e.g. "Clouds are *like* sponges").

Results

Training Trials. First, we examined whether presenting objects in a causal context changed children's likelihood of selecting the functional match or the object match during the training trials. A between-subjects ANOVA with Condition (Causal Metaphor, Causal Simile, Control Simile) as the independent variable and Response (Functional Match, Object Match) as the dependent variable yielded a main effect of Condition, $F(2,94) = 12.72$, $p < .001$. Specifically, children in the Causal Metaphor condition selected the functional match significantly more frequently than children in the Control Simile condition, $t(62) = 4.28$, $p < .001$. Similarly, children in the Causal Simile condition also selected the functional match significantly more frequently than children in the Control Simile condition, $t(62) = 4.19$, $p < .001$. There was no difference in children's performance between the Causal Metaphor and Causal Simile conditions, $t(62) = .14$, $p = .89$. Thus, we found that children in the two causal conditions selected the functional match more frequently than in the control condition.

Additionally, we examined whether children were significantly above chance at selecting the functional match over the object match in each condition. Since there were three experimental groups being compared to chance, we used a Bonferroni correction for multiple comparisons, leading to an adjusted alpha of .017. (We analyzed all results with multiple comparisons in this paper using Bonferonni corrections, but only report adjusted alphas when they impact interpretations of significance or non-significance in the results.) We found that children selected the functional match at above chance levels in the Causal Metaphor condition, $M = 85.94\%$, $SE = 3.71\%$, $t(31) = 9.68$, $p < .001$, and the Causal Simile condition, $M = 86.72\%$, $SE = 4.20\%$, $t(31) = 8.75$, $p < .001$. However, children were at chance selecting between the functional match and the object match in the Control Simile condition, $M = 60.16\%$, $SE = 4.74\%$, $t(31) = 2.14$, $p = .04$.

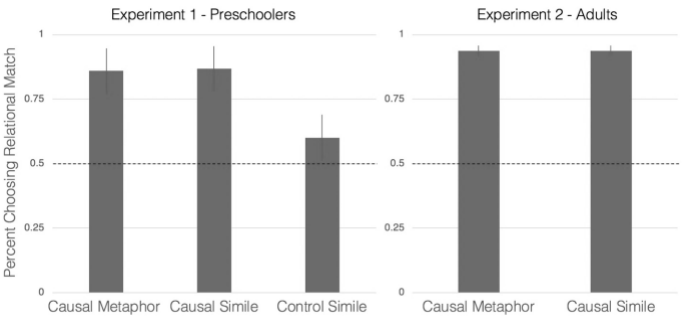

Figure 1: Training trial data from preschoolers and adults. Error bars show 1 standard error.

Test Trials. In order to determine whether children were able to differentiate between metaphors and nonsense statements, we created a Composite Score (percentage of metaphors rated as "smart" subtracted by percentage of nonsense statements rated as "smart") for each child. A child who rated all metaphors as "smart" and all nonsense statements as "silly" would have a score of 1, whereas a child who rated all metaphors *and* nonsense statements as "smart" would have a score of 0. Thus, the Composite Score

assessed children's performance on both metaphor and nonsense statement trials.

In order to investigate whether causal framing would facilitate performance on the metaphor task, we ran a between-subjects ANOVA with Condition (Causal Metaphor, Causal Simile, Control Simile, Baseline Simile) as the independent variable and Accuracy, as measured by Composite Score, as the dependent variable. There was no effect of Condition on Accuracy, $F(3,125) = .30$, $p = .82$. Similarly, a linear regression comparing Accuracy (measured by Composite Scores) in the three training conditions to the baseline condition showed no significant difference between the Baseline Simile condition, M = 10.94%, SE = 7.96%, and any of the other conditions, including the Causal Metaphor condition, M = 15.63%, SE = 5.38%, $\beta = .05$, $p = .57$, Causal Simile condition, M = 17.19%, SE = 5.32%, $\beta = .06$, $p = .45$, and Control Simile condition, M = 10.94%, SE = 4.75%, $\beta < .001$, $p = 1.00$.

Since we did not find a significant difference between any of the conditions, we aggregated data across conditions and analyzed them together. From the aggregated Composite Scores, we find that children performed significantly above chance on the test trials, M = 13.67%, SE = 2.91%, $t(127) = 4.70$, $p < .001$. However, while children rated nonsense statements as "silly" significantly more frequently than chance, M = 59.28%, SE = 2.78%, $t(127) = 3.34$, $p < .001$, their ratings of the metaphors were not different from chance, M = 54.39%, SE = 2.43%, $t(127) = 1.81$, $p = .07$.

Explanations. We examined the explanations that children gave for how the two components of a metaphor were alike (e.g. "How is a roof like a hat?"). There were 128 explanations total, as each child provided an explanation on the final trial. Explanations were coded blind to participants' responses in the training and test trials. Explanations fell into three categories: irrelevant, perceptual, and functional. Irrelevant explanations were non-responses (e.g., "I don't know") or irrelevant (e.g. "I have a tire swing") and comprised 49% of all explanations. Perceptual explanations were based on perceptual similarities (e.g. "they're both fluffy", "because they look the same") comprised 25% of all explanations. Functional explanations were based on functional similarities (e.g. "because you can see through a window and that's why they're like eyes"; "because they both protect your head") and comprised 26% of all explanations. Two coders coded all explanations. Intercoder reliability was 95%, converging on the same category for 122 out of 128 explanations. The categorization of the remaining 6 explanations was resolved through discussion.

We analyzed data from the children who provided functional explanations, perceptual explanations, and irrelevant explanations separately, examining whether the composite scores, metaphor ratings, and nonsense ratings were significant for each group of explanations. Since there were a total of nine comparisons against chance, we used a Bonferroni correction for multiple comparisons, leading to an adjusted alpha of .006. We find that the children who provided functional explanations (n = 33) were able to distinguish between metaphors and nonsense statements: the functional explainers had Composite Score above chance levels, M = 32.58%, SE = 5.24%, $t(32) = 6.21$, $p < .001$, and were significantly likely to rate metaphors as "smart", M = 62.88%, SE = 3.79%, $t(32) = 3.40$, $p = .002$ and nonsense statements as "silly", M = 69.70%, SE = 5.13%, $t(32) = 3.84$, $p < .001$. In contrast, the perceptual explainers (n = 32) had an average Composite Score that was not significantly different from chance levels, M = 11.33%, SE = 6.11%, $t(31) = 1.86$, $p = .07$, and performed at chance on ratings for both metaphors, M = 47.27%, SE = 5.04%, $t(31) = .54$, $p = .59$, and nonsense statements, M = 64.07%, SE = 4.72%, $t(31) = 2.98$, $p = .006$. The irrelevant explainers (n = 63) also had an average Composite Score that was not significantly different from chance levels, M = 4.96%, SE = 3.74%, $t(62) = 1.33$, $p = .19$, and performed at chance on ratings of both metaphors, M = 53.57%, SE = 3.64%, $t(62) =$

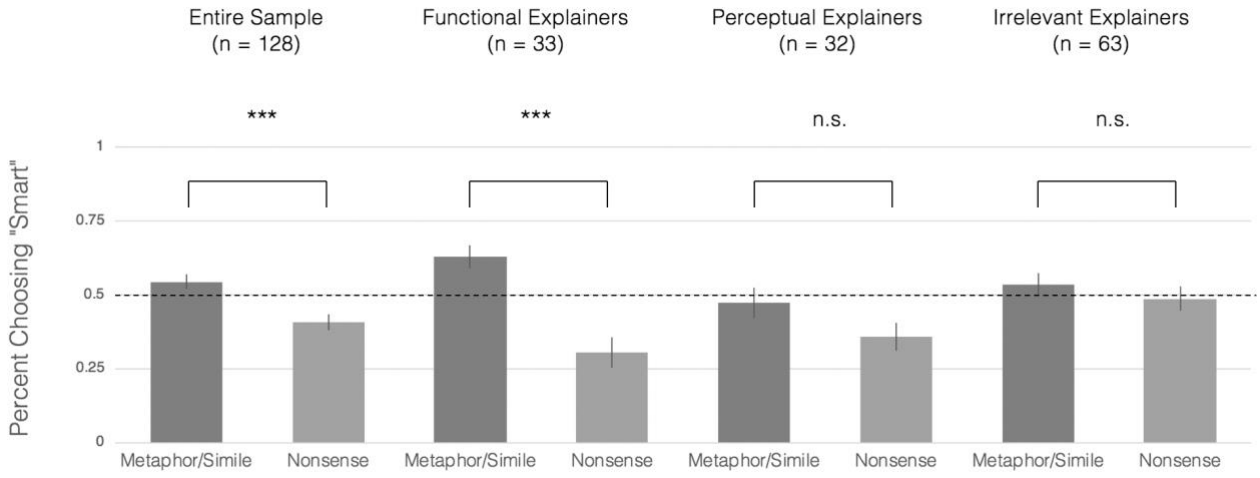

Figure 2: Test trial data from preschoolers. Error bars show 1 standard error.

.98, p = .33, and nonsense statements, M = 51.39%, SE = 4.15%, t(62) = .33, p = .74. Thus, the subset of children who provided explanations involving functional similarity performed above chance on all measures of metaphor comprehension, and their performance drove the success of the entire sample.

Discussion

Our novel paradigm showed that preschoolers already possess some competence with metaphor comprehension and relational reasoning. As a group, preschoolers distinguished between functional metaphors and nonsense statements. This effect was driven by a subset of children who explicitly noted the functional similarities between objects in their explanations. Additionally, we found no difference in children's performance on similes and metaphors, suggesting that preschoolers understand literal and non-literal language equally well.

Consistent with previous research (Goddu et al., in press; Walker et al., 2016; Walker & Gopnik, 2018), we find that introducing a causal framework encouraged preschoolers to adopt a relational mindset, such that they selected the functional matches over the object matches during the causal training trials. In contrast, without a causal framework, preschoolers tend to choose the object match over the relational match (Hochman et al., 2017). However, there was no effect of the causal framework training trials on the metaphor comprehension test trials, as the difference between test trial performance in the four conditions (causal training with metaphors, causal training with similes, control training with similes, and no training with similes) was not significant.

Experiment 2

We ran a sample of adults to validate our novel experimental paradigm. Comprehension of novel metaphors can be difficult for adults (Blasko & Connine, 1993; Bowdle & Gentner, 2005) as well as children (Demorest et al., 1983; Silberstein et al., 1982; Winner et al., 1980). Our novel paradigm may be somewhat pragmatically odd, as it may be unclear what it means for an utterance to be "smart" or "silly". Thus, we wished to demonstrate that adults could distinguish between metaphors and nonsense statements.

Methods

Participants. As with Experiment 1, we adhered to a stopping rule of 32 participants per condition, leading to a total of 64 adults who participated in the study (M = 24.70 years; SD = 5.97 years; 39 females). Researchers tested an additional three participants, whose data were excluded due to experimenter error (two participants) and external interference (one participant). Adults were recruited and tested in a university lab or other quiet on-campus setting. All participants provided informed consent.

Stimuli and Procedure. We ran adults on either the Causal Metaphor condition or Causal Simile condition. The stimuli and procedure of these two conditions are identical to those detailed in Experiment 1.

Results

Training Trials. There was no significant difference in training trial performance between conditions: adults performed identically in the two conditions, t(62) = 0, p = 1.00. Moreover, adults performed almost at ceiling in both conditions. Participants were significantly more likely to pick the functional match than the object match in the Causal Metaphor condition, M = 93.75%, SE = 1.94%, t(31) = 22.50, p < .001, and the Causal Simile condition, M = 93.75%, SE = 1.94%, t(31) = 22.50, p < .001.

Test Trials. We again created Composite Scores (percentage of metaphors rated as "smart" subtracted by percentage of nonsense statements rated as "smart") for each participant. We found that Accuracy, as measured by the Composite Scores, is significantly different between the conditions, t(62) = 2.24, p = .03, with Accuracy being greater in the Causal Simile condition, M = 86.33%, SE = 2.94%, than in the Causal Metaphor condition, M = 72.27%, SE = 5.55%. Regardless, adults were able to distinguish between metaphors and nonsense statements at above-chance levels in both the Causal Metaphor condition, t(31) = 13.02, p < .001, and Causal Simile condition, t(31) = 29.41, p < .001. The difference in Accuracy across conditions was driven by differences in responses to the metaphors across conditions. While there was no significant difference between adults' ratings of the nonsense statements between the Causal Metaphor and Causal Simile conditions, t(62) = .36, p = .72, adults in the Causal Metaphor condition rated the metaphors as "smart" significantly less frequently than adults in the Causal Simile condition, t(62) = 2.21, p = .03. Four out of 32 adults in the Causal Metaphor condition rated all metaphor and nonsense statements as "silly", whereas none of the adults in the Causal Simile condition did so.

Despite differences in the metaphor ratings between the Causal Metaphor and Causal Simile conditions, we found that adults in both conditions were above chance at rating both metaphors and nonsense statements. In the Causal Metaphor condition, adults were significantly above chance at rating the metaphors as "smart", M = 79.30%, SE = 5.66%, t(31) = 5.18, p < .001, and the nonsense statements as "silly", M = 92.97%, SE = 2.24%, t(31) = 19.18, p < .001.

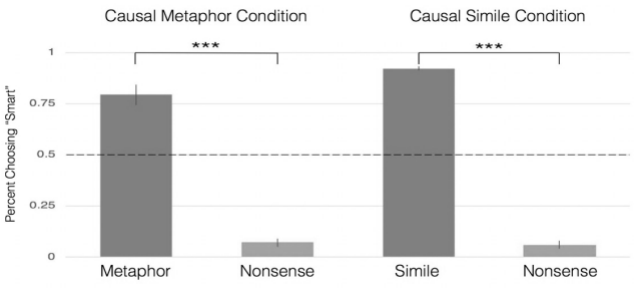

Figure 3. Test trial data from adults. Error bars show 1 standard error.

Similarly, in the Causal Simile condition, adults were significantly above chance at rating the metaphors as "smart", M = 92.19%, SE = 1.46%, t(31) = 28.93, $p < .001$, and the nonsense statements as "silly", M = 94.14%, SE = 2.38%, t(31) = 18.55, $p < .001$. Moreover, 78% of adults in the Causal Metaphor condition and 97% of adults in the Causal Simile condition provided explanations based on functional similarity on the last trial.

Discussion

The results of Experiment 2 validate our paradigm by showing that adults judge metaphors as "smart" and nonsense statements as "silly." It is worth noting that adults are not always at ceiling at the task, especially in terms of rating metaphors as "smart". This result is consistent with previous work demonstrating that novel metaphor comprehension is difficult even for adults (Blasko & Connine, 1993). Interestingly, while there was no difference between preschoolers' "smartness" ratings of metaphors and similes, adults rated similes as smarter than metaphors, consistent with previous work showing that adults prefer novel comparisons in simile form and conventional comparisons in metaphor form (Bowdle & Gentner, 2005).

General Discussion

This paper introduces a new paradigm that provides evidence of preschoolers' capacity to reason about abstract relations, such as shared function, and understand metaphors in an adult-like fashion. We find that at least some preschoolers are already capable of distinguishing between functional metaphors and nonsense statements. Over a quarter of preschoolers were able to explicitly articulate the functional similarities between two objects, and the performance of this subset of children drove the success of the entire sample. This finding provides substantiation for theoretical frameworks that emphasize the role of relational reasoning in metaphor comprehension (Bowdle & Gentner, 2005; Gentner, 1988; Wolff & Gentner, 2011). Further, the fact that some preschoolers were able to provide sophisticated explanations (e.g. "the hat shades you and the top of the roof does too"; "you can drive with wheels and you can walk with feet") is striking, yet consistent with previous work showing that preschoolers are capable of reasoning about abstract relations (Christie & Gentner, 2014; Goddu et al., in press; Hochmann et al., 2017) and the functions of objects (Diesendruck, Markson, & Bloom, 2003; Haward, Wagner, Carey, & Prasada).

Moreover, our finding that a causal framework encourages children to select relational matches over object matches during the training trials conceptually replicates previous work on relational reasoning (Goddu et al., in press; Walker et al., 2016; Walker & Gopnik, 2018). However, this facilitation effect did not transfer over to the metaphor comprehension test trials. One reason for the lack of a facilitation effect on the metaphor comprehension test trials might be that preschoolers, at baseline, were already able to notice functional similarities and understand functional metaphors under the novel "smart versus silly" paradigm. Given preschoolers' baseline competence with metaphor comprehension in this paradigm, the effect of additional training trials may be insignificant. Our new paradigm, which requires children to make absolute judgments about the validity of metaphors, might be a more sensitive measure of metaphor comprehension than previous preference-based paradigms (e.g. Gentner, 1988).

In conclusion, the current research shows that at least a subset of 4- to 5-year-olds are already capable of understanding certain kinds of adult-like non-literal language, such as functional metaphors. Outstanding questions include whether young children can also use metaphors and relational reasoning in the service of other complex cognitive processes, such as learning and conceptual change (Kuhn, 1993; Xu, 2019), and whether researchers may have also underestimated preschoolers' competence at other kinds of figurative language, such as irony (Demorest et al., 1983). These questions provide exciting possibilities for future research.

Acknowledgments

This work was supported by an NSERC Post-Graduate Doctoral Fellowship to RZ [532517-2019]. We are grateful to members of the Cognitive Development and Learning Lab at UC Berkeley, especially Zhimeng Li, Emily Demsetz, Esmerelda Herrera, and Teresa Garcia for their help with data collection. Thanks also to the Lawrence Hall of Science, Children's Creativity Museum, and the preschools, parents, and children who made this research possible.

References

Anderson, E. M., Chang, Y. J., Hespos, S., & Gentner, D. (2018). Comparison within pairs promotes analogical abstraction in three-month-olds. *Cognition, 176,* 74-86.

Ballestrino, P., Carriedo, N., Sebastián, I., Corral, A., Montoro, P. R., & Herrero, L. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLoS ONE, 11,* e0150289.

Blasko, D.G., & Connine, C.M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 19, 295-308.

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

Camp, E. (2006). Metaphor in the mind: The cognition of metaphor. *Philosophy Compass, 1,* 154-170.

Camp, E. (2009). Two varieties of literary imagination: Metaphor, fiction, and thought experiments. *Midwest Studies In Philosophy, 33,* 107-130.

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science,* 38, 383–397.

Demorest, A., Silberstein, L., Gardner, H., & Winner, E. (1983). Telling it as it isn't: Children's understanding of

figurative language. *British Journal of Developmental Psychology, 1*, 121-134.
Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14*, 164-168.
Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*, 47-59.
Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. *Psychology of Learning and Motivation, 22*, 307-358.
Goddu, M., Lombrozo, T., & Gopnik, A. (in press). Transformations and transfer: Preschool children understand abstract relations and reason analogically in a causal task. *Child Development.*
Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition, 176*, 255-268.
Hochmann, J. R., Mody, S., & Carey, S. (2016). Infants' representations of same and different in match- and non-match-to-sample. *Cognitive Psychology, 86*, 87–111.
Hochmann, J. R., Tuerk, A., Sanborn, S., Zhu, R., Long, R., Dempster, M., & Carey, S. (2017). Children's representation of abstract relations in relational/array match-to-sample tasks. *Cognitive Psychology, 99*, 17-43.
Holyoak, K. J., & Stamenkovíc, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin, 144*, 641-671.
Kuhn, T.S. (1993) Metaphor in science. *In Metaphor and* Thought (Ortony, A., ed.), pp. 533–542, Cambridge University Press.
Pouscoulous, N., & Tomasello, M. (2019). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*.
Reynolds, R.E., & Ortony, A. (1980). Some issues in the measurement of children's comprehension of metaphorical language. *Child Development, 51*, 1110-1119.
Roberts, R.M., & Kreuz, R.J. (1994). Why do people use figurative language? *Psychological Science, 5*, 159-163.
Silberstein, L., Gardner, H., Phelps, E., & Winner, E. (1982). Autumn leaves and old photographs: The development of metaphor preferences. *Journal of Experimental Child Psychology, 34*, 135-150.
Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How linguistic metaphor scaffolds reasoning. *Trends in Cognitive Science, 21*, 852-863.
Vosniadou, S., & Ortony, A. (1983). The emergence of the literal-metaphorical-anomalous distinction in young children. *Child Development, 54*, 154-161.
Walker, C. M., Bridgers, S., & Gopnik, A. (2016). The early emergence and puzzling decline of relational reasoning: Effects of knowledge and search on inferring abstract concepts. *Cognition, 156*, 30-40.
Walker, C. M., & Gopnik, A. (2017). Discriminating relational and perceptual judgments: Evidence from human toddlers. *Cognition, 166*, 23-27.
Winner, E., Engel, M., & Gardner, H. (1980). Misunderstanding metaphor: What's the problem? *Journal of Experimental Child Psychology, 30*, 22-32.
Wolff, P., & Gentner, D. (2011). Structure-mapping in metaphor comprehension. *Cognitive Science, 35*, 1456-1488.
Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review, 126*, 841-864.