Title
Preschoolers' acquisition of functional metaphors

Author
Zhu, Rebecca

Publication Date
2022

Peer reviewed|Thesis/dissertation

Preschoolers' Acquisition of Functional Metaphors

By

Rebecca Zhu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor in Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Dr. Alison Gopnik, Chair
Dr. Steven Piantadosi
Dr. John Campbell

Spring 2022

Preschoolers' acquisition of functional metaphors

Abstract

Preschoolers' Acquisition of Functional Metaphors

by

Rebecca Zhu

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Alison Gopnik, Chair

Metaphors are ubiquitous in both everyday speech (e.g., "I was lost in a *sea* of people") and creative literature (e.g, "A word is *elegy* to what it signifies"). Metaphors provide frameworks for reasoning about abstract concepts, influencing how humans attend to, remember, and process information. Metaphors are also a force for creative change across various disparate domains, for example by spurring the creation of new explanatory theories in science and new word meanings in language. While metaphors are a powerful tool for thinking and reasoning in adulthood, previous research suggests that children do not understand metaphors until quite late in development, by some accounts not until adolescence. However, using novel experimental paradigms, the current dissertation provides new data suggesting that children as young as four years of age can not only understand, but also learn from, metaphors.

Chapter 1 outlines previous research on the development of metaphor comprehension. Chapter 2 uses novel paradigms to demonstrate that preschoolers can understand metaphors based on shared abstract, functional similarities (e.g., "Clouds are sponges"; "Roofs are hats"). Chapter 3 explores whether preschoolers can select the most relevant metaphor to learn from, and finds that providing preschoolers with explanations can shift their metaphor preferences in a manner that is conducive to further learning. Chapter 4 shows that preschoolers can successfully use metaphors to make additional inferences about novel concepts. Finally, Chapter 5 discusses the implications of this work for research in linguistic and cognitive development, as well as future directions.

Overall, this series of experiments suggest that children possess an early-emerging capacity to understand and use complex non-literal language. Moreover, metaphors may be a powerful cognitive mechanism that facilitates learning, even early in ontogenesis.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am exceptionally lucky to have spent the past five years exploring children's acquisition of symbolic systems. The most fully realized line of research, on children's metaphor comprehension, became this dissertation. There are so many people to thank for helping me along this journey.

First and foremost, I am deeply grateful for my advisor, Dr. Alison Gopnik. To borrow one of her own metaphors, Alison was an exceptional gardener, rather than a carpenter, who provided me with opportunities to flourish in productive and unexpected ways. She was always there to provide a secure base and sounding board, while allowing me to take risks and create my own path forward. She encouraged me to spend a summer in Kenya, which led to the development of an entirely new line of research. Thanks to Alison, I have grown in so many ways.

I am also thankful for my dissertation committee members, Drs. John Campbell and Steven Piantadosi, for thoughtful, provocative, and fun conversations. Additionally, I am grateful to Drs. Jan Engelmann and Lia Fernald for supporting my research endeavours in Kenya. Thanks also to my post-doctoral collaborators, Drs. Helen Pitchik and Mariel Goddu. In particular, Helen's meticulousness, statistical knowledge, and desire to produce research with a clear positive impact continues to improve our work exponentially.

An amazing group of research assistants deserve special recognition. Zoe Carwin, Audrey Kaufman, Lily Zihui Zhu, Mary Nguyen, Mina Chang, Anna Cialfi, Sophie McMullen, Pamee Sapasap, Zhimeng Mona Li, Randall Hale, Esmerelda Herrera, Grace Horton, and Sam Ku – you are all wonderful. I looked forward to every one of our weekly meetings because of the level of dedication, care, kindness, patience, curiosity, and joy that you brought to these projects. This dissertation would not be possible without your hard work and enthusiasm. A special thanks to Sophie McMullen for taking charge of online testing logistics during the start of the COVID-19 pandemic, and Zoe Carwin and Pamee Sapasap for creating the artwork in this dissertation.

My research is made possible thanks to the essential work of lab managers Teresa Garcia, Alissa Gomez, Emily Rose Reagan, and Catherine Berner, as well as the continued support of local preschools, museums, and parents. My graduate career was generously supported by the Natural Sciences and Engineering Council of Canada, the Aga Khan Foundation, the Jacobs Foundation, the Center for Effective Global Action, the Berkeley Fellowship, the John L. Simpson Graduate Research Fellowship, the Barbara and Arnold Silverman Graduate Fellowship, and the SURF SMART Mentorship Program. Parts of Chapters 3 and 4 were previously published by the Cognitive Science Society, as conference proceedings.

My amazing friends – including but certainly not limited to Alena Egner, Antonia Langenhoff, Eszter Meszaros, Garrett Maron, Grace Huckins, Greyson Abid, Isabelle Boni, Joshua Confer, Joshua Price, Kara Lessin, Kay Simon, Madeleine Welsch, Micah Brush, Monica Ellwood-Lowe, Oded Ritov, and Smitha Milli – enriched my research and my life. Thanks for joining me on my adventures, and for facilitating and/or tolerating my personal antics.

Finally, I owe so much to my parents – I am grateful for your steadfast, unconditional love and support. And to Charlie, for always making me smile.

Chapter 1: Introduction

1.1 General Introduction

"A word is elegy to what it signifies" – Robert Hass, *Meditation at Lagunitas*

A metaphor is a figurative utterance that directly compares a concept from one domain to another concept in an unrelated domain. For example, in his poem 'Meditation at Lagunitas', Robert Hass compares two highly disparate concepts, namely *words* and *elegies*. By incorporating these two different concepts in a single metaphor, Hass highlights their shared similarities (Bowdle & Gentner, 2005; Holyoak & Stamenkovíc, 2018). In this case, according to Hass, words, like elegies, possess an element of loss, "because there is in this world no one thing/to which the bramble of *blackberry* corresponds" (Hass, 1979). Moreover, Hass's innovative metaphor presents the familiar, everyday concept of *words* from a new perspective (Camp, 2009), encouraging readers of his poem to take on a novel, abstract, and unconventional framework to reason about a commonplace concept. Thus, by drawing comparisons between two disparate concepts, metaphors allow humans to think about old concepts in a new light (Camp, 2009; 2015; Reimer & Camp, 2006).

Additionally, Hass's metaphor is especially apt for the current dissertation because it raises another fundamental topic in cognitive science, namely problems of referential ambiguity. By arguing that words cannot fully do justice to the concepts they represent, Hass's metaphor touches upon the arbitrary nature of word-meaning mappings and the problem of referential ambiguity in language (Genone & Lombrozo, 2012; Li et al., 2018; Machery et al., 2011). Natural languages are inherently ambiguous symbol systems, compromised mostly, if not entirely, of arbitrary mappings between words and concepts. The problem of referential ambiguity may be especially pronounced for some kinds of non-literal language or "loose talk" in which people do not actually "say what they mean". Examples of "loose talk" include but are not limited to metaphors (Camp, 2005; 2009), metonyms (Littlemore, 2015; Falkum et al., 2017; Köder & Falkum, 2020; Zhu, 2021), hyperboles (Kao et al., 2014), and ironic utterances (Demorest et al., 1983). Consequently, while metaphors are extremely common in natural languages (Lakoff & Johnson, 1980), they may also pose difficulties for language comprehension (Gentner & Clement, 1998; Reimer & Camp, 2006) and acquisition (Silberstein et al., 1982; Winner et al., 1980), perhaps due to a heightened sense of vagueness or ambiguity in meaning. This dissertation investigates the cognitive processes underlying metaphor comprehension, and in turn, how metaphors might facilitate further thinking, reasoning, and learning.

1.1.1 Metaphor comprehension in adults

Though metaphors are ubiquitous in natural languages (Lakoff & Johnson, 1980), there is still much debate in cognitive science about the exact mechanisms underlying humans' metaphor comprehension abilities. For example, psychologists and philosophers have posited at least five theories of metaphor comprehension (Camp, 2006; Holyoak & Stamenkovíc, 2018), such as juxtaposition (Davidson, 1978), feature matching (Fogelin, 1988), embodied conceptual mappings (Lakoff & Johnson, 1980), categorization (Glucksberg & Keysar, 1990), and structural alignment (Bowdle & Gentner, 2005). Each theory has strengths and weaknesses, and some theories align more consistently with empirical evidence than other theories (see Holyoak & Stamenkovíc, 2018 for review). Indeed, some philosophers have questioned whether researchers can explain metaphors at all, or if it is a hopeless endeavor: for example, the philosopher Davidson (1978) argues that metaphors do not even contain propositional content. However, these many theories of metaphor comprehension may not be mutually exclusive, or as incompatible as they seem at first glance (Camp, 2006). Indeed, while the details of the theories (e.g., the specifics of online metaphor processing) may vary, many theories explicitly or implicitly highlight similar cognitive mechanisms necessary for metaphor comprehension. For example, multiple theories highlight *relational reasoning* – namely, the ability to notice similarities across concepts – as a key cognitive mechanism underlying adult metaphor comprehension.

In addition to the literature on adult metaphor comprehension, a growing body of research investigates how metaphors might impact cognition in human adults (Thibodeau et al., 2017). Historically, metaphors are a force for creative change across multiple disparate domains, such as science (Kuhn, 1993), art (Camp, 2009; 2015), and everyday language (Holyoak & Stamenkovíc, 2018). In particular, researchers have suggested that metaphors might influence cognition by providing concrete frameworks to reason about abstract concepts, such as emotions (Lakoff & Johnson, 1980), crime (Thibodeau & Boroditsky, 2011), and climate change (Flusberg et al., 2017; Thibodeau et al., 2017). Consequently, while metaphors may sometimes pose a cognitive challenge in terms of language comprehension, they may also confer cognitive benefits. Specifically, metaphors may help adults view familiar concepts from a new perspective (Camp, 2009; Thibodeau et al., 2017) or facilitate the discovery of new information (Kuhn, 1993).

1.1.2 Metaphor comprehension in children

Children face the complex task of acquiring their native language within the first few years of life. Some aspects of language seem easy to acquire early in development: for example, previous research on semantic acquisition shows that infants understand the meaning of frequent common nouns (e.g., “apple”) at six months (Bergelson & Swingley, 2012), and more abstract function words (e.g., “the”) at 17 months (Hochmann et al., 2010). Moreover, infants are sensitive to abstract syntactic rules, for example using word order to infer subjects and objects at 21 months (Gertner et al., 2006). In general, extensive research in cognitive and linguistic development shows that children are already in the process of successfully acquiring many aspects of language within the first few years of life (Csibra & Shamsudheen, 2015; Dewar & Xu, 2009; Landau et al., 2003; Markman, 1990; Shafto et al., 2014; Vouloumanos & Waxman, 2014; Waxman & Gelman, 2009; Xu & Tenenbaum, 2007). Many aspects of natural language, though remarkably complex, are acquired easily and rapidly by children within the first few years of development.

Though some aspects of language are acquired quite early in development, other aspects of language learning seem difficult for children to fully grasp, demonstrating more protracted developmental trajectories. Given that metaphor comprehension can sometimes be difficult even for adults (Bowdle & Gentner, 2005), it is unsurprising that previous research has found that metaphor comprehension can also be difficult for young children (Demorest et al., 1983; Silberstein et al., 1982; Winner et al., 1976; 1980). For example, Winner and colleagues (1976) presented six-year-old children with novel metaphors (e.g., “The prison guard is a hard rock”;

“Her perfume was bright sunshine”) and found that the young children attempted to interpret the metaphors in a literal manner (e.g., “the guard worked in a prison with rock walls”), perhaps sometimes with an element of magic or fantasy (e.g., “the king turned the guard into a rock”; “her perfume was made out of the sun”). Indeed, Winner and colleagues (1976) found that genuine metaphoric understanding did not seem to arise until around ten to fourteen years of age. Moreover, children performed similarly in both a forced-choice and free explanation paradigm, suggesting that young children’s poor performance was due to a genuine lack of understanding rather than task performance difficulties (e.g., difficulty articulating clear explanations). Researchers have speculated on why exactly children may have particular difficulty understanding metaphors: for example, metaphors might pose special challenges because they require knowledge about unfamiliar conceptual content (Keil, 1986), pragmatic inference abilities (Winner et al., 1976), or sophisticated relational reasoning abilities (Nippold & Sullivan, 1987; Winner et al., 1980).

While many empirical studies have demonstrated that children acquire a full-fledged, adult-like understanding of metaphors quite late in development (Demorest et al., 1983; Gardner et al., 1975; Silberstein et al., 1982; Winner et al., 1976; 1980), researchers have recently begun to question this established perspective (Pouscoulous, 2011; Pouscoulous & Tomasello, 2020). For example, Pouscoulous (2011) argues that even young children already possess the necessary cognitive abilities underlying metaphor comprehension, and that children’s poor performance in previous experimental tasks can be explained by a lack of vocabulary or world knowledge, as well as and generally unfavorable task demands. Indeed, more recent research demonstrates that young children possess some of the cognitive capacities thought to underlie metaphor comprehension, such as the pragmatic capacity to understand non-literal language (Falkum et al., 2017; Köder & Falkum, 2020) and the capacity to reason about shared relations between different concepts (Anderson et al., 2018; Carstensen et al., 2019; Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017, Holyoak et al., 1984; Walker et al., 2016; Walker & Gopnik, 2017).

Ultimately, the question of when exactly children develop a full-fledged, adult-like understanding of metaphors is important not only for linguistic development, but also for broader, domain general learning. Since metaphors facilitate thinking, reasoning, and learning in human adults (Flusberg et al., 2017; Kuhn, 1993; Lakoff & Johnson, 1980; Thibodeau et al., 2017; Thibodeau & Boroditsky, 2011), it is possible that metaphors also confer similar cognitive benefits for children. For example, metaphors might be a powerful learning mechanism that furthers children’s discovery of the world around them. However, in order to learn from metaphors, children must of course first possess the capacity to understand metaphors. Consequently, this dissertation explores questions relating both to children’s *understanding* and *use* of metaphors.

1.1.3 The development of relational reasoning

One particularly important cognitive capacity underlying the linguistic capacity to understand metaphors is *relational reasoning*, a term that is also synonymous with *analogical reasoning* (Bowdle & Gentner, 2006; Gentner, 1988; Gentner & Clement, 1998; Holyoak & Stamenkovíc, 2018; Roberts & Kreuz, 1994; Wolff & Gentner, 2011). Relational reasoning is the ability to attend to similarities and differences between two objects or concepts. Consequently, it is evident that relational reasoning might be a critical component of metaphor comprehension, which involves comparing two disparate concepts.

Recently, developmental psychologists have discovered that relational reasoning abilities emerge quite early in development, in the preschool years or earlier. In particular, relatively young children already possess the capacity to represent abstract concepts such as *same* and *different*, and can flexibly apply these concepts to succeed on many different kinds of experimental tasks. For example, Hochmann and colleagues (2017) found that five-year-olds spontaneously succeed on a classic relational match-to-sample task, in which children must match either a pair of same objects (e.g., XX) or a pair of different objects (e.g., YZ) to a target card with either another same pair (e.g., AA) or another different pair (e.g., BC). Given sufficient training, even four-year-olds succeed at this classic relational match-to-sample task (Kroupin & Carey, 2022).

Moreover, other researchers using innovative paradigms have found evidence of relational reasoning even earlier than four years of age. For example, Walker and colleagues (2016) used a blicket detector paradigm in which either a *same* pair (i.e., two perceptually identical blocks) or a *different* pair (i.e., two perceptually dissimilar blocks) activated the blicket detector. In this experimental paradigm, toddlers were successful at learning a single rule predicated on either sameness or difference (i.e., *same* blocks make the blicket light up, or *different* blocks make the blicket light up), demonstrating an early-emerging capacity to reason about similarities and differences across objects.

Indeed, developmental psychologists have even found evidence of relational reasoning in infancy (Anderson et al., 2018; Hochmann et al., 2016). Using a habituation paradigm, Anderson and colleagues (2018) find evidence that 3-month-olds represent sameness and difference. In this experiment, 3-month-olds were habituated to either the *same* relation (i.e., viewing object pairs XX, YY) or the *different* relation (i.e., viewing object pairs XY, AB). Consequently, on the test trials, 3-month-olds also looked longer when presented with the opposite relation. Taken together, these findings suggest that relational reasoning abilities emerge remarkably early in development.

Though this recent research output is extremely promising, there are still unanswered questions. Specifically, much of the research demonstrating the capacity for relational reasoning in preschoolers, toddlers, and infants uses stimuli in which sameness and difference is predicated on perceptual identity (Anderson et al., 2018; Hochmann et al., 2017; Walker et al., 2016; Walker & Gopnik, 2017). For example, many experiments use two objects that are physically identical in terms of shape, color, and size to represent sameness. Even experiments that do not use physically identical pairs still frequently use a dimension of similarity predicated upon a single perceptual feature, such as size or shape (Goddu et al., 2020; Kroupin & Carey, 2022). In contrast, the kinds of similarities used in metaphors are often predicated upon more abstract features, such as shared functional similarities (e.g., “Conscience is a man’s compass”). Consequently, it is still an open question as to whether young children can represent the kinds of similarities necessary to understand metaphors, namely more abstract kinds of similarities that are not predicated on any kind of perceptual likeness. Thus, by investigating children’s ability to understand and use metaphors based on shared functional relations (e.g., “Clouds are sponges”; “Roofs are hats”), the current dissertation also contributes to the literature on children’s relational reasoning abilities – namely, by investigating whether young children possess the capacity to represent shared abstract similarities between concepts, in addition to their capacity to represent shared perceptual similarities between concepts.

1.2 Précis

The current dissertation investigates how and when children understand metaphors based on shared abstract, functional similarities, and in turn, how these metaphors might further facilitate children's early learning.

Chapter 2 explores whether preschoolers might already be capable of understanding metaphors. Previous work argues that metaphor comprehension is an ability acquired late in development, possibly not until adolescence. Researchers argue that relational reasoning - the ability to attend to similarities between objects, rather than to the features of individual objects - underlies metaphor comprehension. Thus, children's failure to understand metaphors may be attributed to their inability to notice relational structures, such as functional similarities, between concepts. However, recent work shows evidence of the ability to represent abstract relations quite early in ontogenesis, in preschoolers, toddlers, and even infants. Using novel experimental paradigms, Chapter 2 demonstrates that preschoolers are already capable of understanding functional metaphors, namely metaphors based on shared abstract, functional similarities (e.g. "Clouds are sponges"; "Roofs are hats").

Chapter 3 examines whether preschoolers can select the most relevant and maximally informative metaphor to learn from. Preschoolers' metaphor preferences are not merely aesthetic choices; rather, metaphor preferences are important for other cognitive processes, such as learning by analogy. Two concepts can be the same or different along an infinite number of dimensions, but some dimensions are more useful for learning than others; thus, in order for metaphors to facilitate learning, children must not only understand metaphors, but also be able to select the most relevant metaphor to learn from. In this series of experiments, I find that adults prefer functional metaphors based on abstract, functional similarities (e.g. "Eyes are windows") while preschoolers prefer perceptual metaphors based on surface-level, perceptual similarities (e.g. "Eyes are buttons). However, Chapter 3 also finds that providing explicit explanations of the similarities underlying the metaphor (e.g. "Eyes are windows *because you see through both of them*"; "Eyes are buttons *because both are round*") significantly shift preschoolers' metaphor preferences towards abstract functional metaphors. Thus, providing preschoolers with explanations about how objects in metaphors are similar can shift their preferences in a manner that is conducive to further learning.

Chapter 4 demonstrates that preschoolers can use metaphors to make learn. Historically, metaphors have been a creative force for scientific and linguistic change. Psychologically, metaphors guide thinking and reasoning in adults. Thus, it is plausible that preschoolers can also use metaphors to guide their learning, for example by making additional inferences about novel concepts. Chapter 4 finds that both preschoolers and adults can use metaphors about novel artifacts (e.g., "Wugs are songbirds") to infer functional properties of the artifacts (e.g., that wugs make music), thus providing empirical evidence that children can use metaphors to facilitate further learning.

Taken together, the current dissertation demonstrates that young children can not only *understand* metaphors, but also *use* metaphors to guide their thinking and reasoning. These findings advance the field of developmental psychology by providing evidence of children's sophisticated linguistic and cognitive capacities, early in development.

Chapter 2: Preschoolers’ comprehension of functional metaphors

2.1 Introduction

Metaphors are ubiquitous 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), influencing attention, memory, and information processing (Thibodeau et al., 2017; 2019). Metaphors are also a force for creative change: they can facilitate the discovery of new scientific theories (Kuhn, 1993) and the creation of new word meanings (Bowdle & Gentner, 2005; Holyoak & Stamenkovíc, 2018).

While metaphors promote novel ways of thinking and reasoning (Thibodeau et al., 2017), they also pose unique language comprehension challenges. Researchers have posited multiple theories of metaphor comprehension that rely on different underlying cognitive mechanisms, such as relational reasoning (Gentner, 1988; Gentner & Clement, 1998; Holyoak, 2019), categorization (Glucksberg & Keysar, 1990; Glucksberg et al., 1997), or embodied conceptual mappings (Lakoff & Johnson, 1980; Thibodeau et al., 2019). *Relational reasoning* – the ability to attend to similarities based on abstract relations between objects – plays a critical role (Bowdle & Gentner, 2005; Gentner, 1988; Holyoak & Stamenkovíc, 2018; Roberts & Kreuz, 1994; Wolff & Gentner, 2011). To understand a novel metaphor, a listener must identify the relational bases for an equivalence drawn between two objects that are not conventionally associated with each other (e.g. making sense of “clouds are sponges” entails recognizing that both objects *hold water*).

Researchers attributed children’s failure to understand metaphors in part to their inability to notice abstract relational structures between concepts. In one study, participants were asked to complete sentences (e.g. “The volcano is…). Six-year-olds tended to select *perceptual* completions (e.g. “a bright firetruck”) and adults tended to select *conceptual* completions (e.g. “a very angry man”) (Silberstein et al., 1982). Thus, preschoolers tend to prefer perceptual metaphors, while adults tend to prefer more abstract metaphors (Gentner & Clement, 1988). These and other results have led many to argue that children do not understand abstract metaphors in an adult-like fashion until late in development, possibly not until adolescence (Demorest et al., 1983; Silberstein et al., 1982; Winner et al., 1976; 1980).

In contrast to this view, a few previous studies demonstrate early metaphor comprehension in preschoolers (Pouscoulous & Tomasello, 2020; Vosniadou & Ortony, 1983). Notably, however, these previous tasks involve metaphors based on *perceptual* similarities such as shape or color (e.g. “Moons are cookies”; “Eyes are buttons”) rather than *abstract* similarities such as function (e.g. “Moons are lightbulbs”; “Eyes are windows”). In general, perceptual metaphors have limited utility for thinking and reasoning about novel or abstract concepts. Given that perceptual metaphors are frequently based on coincidental, surface-level similarities, they tend not to facilitate further insight or understanding. For example, there are few useful additional inferences to be made from the observation that eyes are a similar shape to buttons, or that moons are a similar shape to cookies. Moreover, metaphors are particularly useful for reformulating complex abstract concepts in terms of concrete concepts (e.g. “love is a journey”; “conscience is a man’s compass”) (Thibodeau et al., 2017). However, many abstract concepts used in metaphors (e.g. love, conscience) simply don’t *have* perceptual features and cannot enter into perceptual metaphors. Thus, while evidence that preschoolers understand metaphors based on perceptual similarities is a promising advance, the current literature falls short of demonstrating that preschoolers can reason about abstract relations and understand novel metaphors in an adult-like fashion—especially in a way that promotes learning and reasoning.

Similarly, some research shows that young children can understand metaphors comparing spatial and temporal domains (e.g. "time *flies* by"; Özçalışkan, 2002; 2007). While these kinds of metaphors are certainly more abstract than perceptual metaphors, they compare two concepts (i.e. time and space) that are very closely related (Lakoff & Johnson, 1980). Indeed, many researchers argue that time and space are part of the same representational system, namely the general magnitude system (Lourenco & Longo, 2011). However, it remains unclear if preschoolers can also understand metaphors based on two entirely disparate, unrelated concepts. Thus, while preschoolers' comprehension of conceptually related time-space metaphors is another promising advance, this prior work falls short of demonstrating that children possess the sophisticated domain-general understanding of metaphors that adults use to think, reason, and learn about a wide variety of known and novel concepts.

Interestingly, a number of recent studies investigating the development of relational reasoning suggest that the capacity to represent abstract relations is present in preschoolers (Carstensen et al., 2019; Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017, Holyoak et al., 1984), toddlers (Walker et al., 2016; Walker & Gopnik, 2017), and even infants (Anderson et al., 2018; Hochmann et al., 2016). Moreover, preschoolers are not only able to represent basic abstract relations such as sameness or difference of identity (Carstensen et al., 2019; Christie & Gentner, 2014; Hochmann et al., 2017), but also relations based on other dimensions, such as size, number, and color (Goddu et al., 2020). This new 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.

In light of these recent findings demonstrating that young children are capable of relational reasoning, the present study aims to investigate whether young children might be able to understand *functional* metaphors – namely, metaphors based on abstract functional similarities – at an earlier developmental timepoint than researchers initially estimated. In the previous literature (e.g. Gentner, 1988; Gentner & Clement, 1988), metaphors based on abstract similarities, such as shared function or structure, were called *relational* metaphors. In the current paper, we aim for more terminological specificity, by defining exactly what kind of similarity the metaphor is based on (i.e. *functional* metaphors are based on the functional similarity between the two objects in a metaphor, and *perceptual* metaphors are based on the perceptual similarity between two objects in a metaphor).

The main feature of the current studies is that Experiments 1 and 3 use novel paradigms, which may be more sensitive to children's metaphor comprehension abilities. In particular, we introduce novel tasks that ask children to judge statements as "smart" or "silly", and contrast children's judgments of functional metaphors and nonsense statements. In some of the studies, we also ask children to provide justifications, and compare their performance on metaphors versus similes. In addition, given that causal frameworks can induce a relational mindset (Goddu et al., 2020; Walker et al., 2016; Walker & Gopnik, 2018), we also explored whether causal framing might facilitate preschoolers' metaphor comprehension.

In the current paper, we present multiple exploratory experiments demonstrating preschoolers' competence with functional metaphors. Experiment 1a asked preschoolers to rate functional metaphors (e.g. “roofs are hats”) and nonsense statements (e.g. “boats are skirts”) as “smart” or “silly” in an absolute judgment paradigm, and to provide justifications for their responses. Experiment 1b validated this novel paradigm with adults. Experiment 2 asked whether preschoolers preferred functional explanations (e.g. “roofs and hats both keep you dry”) over perceptual explanations (e.g. “roofs and hats both have pointy tops”) when interpreting the functional metaphors used in Experiment 1a. Finally, Experiment 3 asked whether preschoolers preferred functional metaphors (e.g. “roofs are hats”) to nonsense statements (e.g. “roofs are scissors”) in a dichotomous-choice paradigm, and to again provide justifications for their responses. Taken together, these results suggest that preschoolers can already understand complex metaphors based on abstract similarities, such as shared function.

2.2 Experiment 1a

In Experiment 1a, we use a novel experimental paradigm to explore preschoolers’ metaphor comprehension abilities. While previous research used dichotomous-choice paradigms (e.g. Gentner, 1988; Vosniadou & Ortony, 1983), we use an absolute judgment paradigm – asking for “smart” or “silly” ratings of functional metaphors and nonsense statements – that may be a more sensitive measure of preschoolers’ emerging abilities. We also included causal and non-causal training trials, to test for facilitative effects of causal framing on metaphor comprehension (Goddu et al., 2020; Walker et al., 2016; Walker & Gopnik, 2018).

Additionally, there are multiple reasons why preschoolers might struggle with metaphor comprehension. While one possibility is that preschoolers struggle with relational reasoning, another possibility is that preschoolers struggle with non-literal interpretations (Allen & Butler, 2020; Reynolds & Ortony, 1980; Vivaldi & Allen, 2021). Consequently, Experiment 1 involves conditions with both metaphors, which are non-literal (e.g. “Clouds are sponges”), and similes, which are literal (e.g. “Clouds are *like* sponges”). Preschoolers’ ability to understand similes would indicate an early competency with relational reasoning, whereas preschoolers’ ability to understand metaphors would indicate early competencies with both relational reasoning and non-literal language. While relational reasoning underlies both metaphor and simile comprehension, the two phenomena are also distinct: adults seem to favor metaphors over similes, reporting that metaphors are more interesting (Roberts & Kreuz, 1994) and cognitively “forceful” (Glucksberg & Keysar, 1990) than similes.

2.2.1 Methods

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 children understand similes more easily than metaphors (Reynolds & Ortony, 1980; but see also Winner et al., 1980), 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”). Some of the metaphors in the experiment were taken from previous studies (e.g. "Moons are lightbulbs"; Gentner, 1988), while others (e.g. "Pools are bowls") were newly-generated.

2.2.1.1 Participants

We tested 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; range = 4.01 – 5.88 years; 61 males, 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 in this paper lasted approximately five to ten minutes, and were conducted independently (i.e. participants did not complete additional studies during the same testing session). All sample sizes in this paper, though not formally preregistered, were set prior to testing based on counterbalancing requirements. Since our recruitment techniques drew from local convenience samples, participants were predominantly White and upper middle class across all experiments. All experiments in this paper were approved by the university's Committee for the Protection of Human Subjects. All parents of child participants provided informed consent.

2.2.1.2 Stimuli & Procedure

The experimenter presented participants with the stimuli on a laptop computer. Each child participated in one of four conditions. The *Causal Metaphor* condition involved causal training trials prior to the test trials, and used metaphors throughout. The *Causal Simile* condition involved causal training trials prior to the test trials, and used similes throughout. The *Control Simile* condition involved non-causal training trials prior to the test trials, and used similes throughout. Finally, the *Baseline Simile* condition involved only test trials using similes. During the test trials, all participants were presented with metaphors and nonsense statements, and had to differentiate between the two kinds of utterances. For the full list of metaphors, similes, and nonsense statements used in each experiment, see Supplementary Materials.

2.2.1.2.1 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. These trials were modelled after another experiment (Goddu et al., 2020) that demonstrated preschoolers' understanding of abstract relations in the context of causal transformations. For example, children who saw a wizard turn a small apple into a large apple predicted that the wizard's action on a new object would lead to a transformation that exemplified the same relation (e.g. turn a small dog into a large dog).

In the Causal Metaphor training trials, the experimenter introduced the task 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*, namely an object that shared the same function (e.g. a glove), and one that was an *object match*, namely an object from the previous trial (e.g. a hot air balloon). The experimenter pointed to and named the object, and then 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 what the novel object on the conveyor belt will turn into. Participants received feedback on each of their choices. 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. For a full list of training trials, see Table 2.1.

It is worth noting that though the statements in the Causal Metaphor condition follow a standard “X is Y” metaphor form, these statements might not be considered metaphors because the statements are literally true: one object in the statement undergoes a causal transformation and literally turns into the other object. However, whether or not preschoolers believe these statements to be literal or non-literal should not affect whether they select the functional match or the object match.

2.2.1.2.2 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.

2.2.1.2.3 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. The experimenter still gave participants feedback on their responses.

2.2.1.2.4 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.

2.2.1.2.5 Test Trials. During the Test Trials, we deliberately emphasized to participants that they are playing a new game with a new character, so that the test trial metaphors - which involve objects merely appearing onscreen, rather than undergoing causal transformations - are more likely to be interpreted as non-literal statements. 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.

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?").

There were sixteen test trials total: 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. No more than three of the same kind of trial appeared consecutively, and the last trial was always a metaphor. Each of the eight metaphors appeared as the last trial an equal number of times (e.g. within each condition, children were asked to explain how clouds are like sponges as frequently as they were asked to explain how tires are like shoes). For a full list of test trials, see Table 2.2.

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").

Training Trial Part 1	Training Trial Part 2	Functional Match	Object Match
Moons are lightbulbs	Ladders are…	Hills	Lightbulbs
Birds are hot air balloons	Sleeping bags are…	Gloves	Hot air balloons
Teeth are knives	Treebarks are…	Skins	Knives
Bonfires are ovens	Brains are…	Computers	Ovens

Table 2.1 Experiment 1 Training Trials.

Functional Metaphors	Nonsense Statements
Roofs are hats	Books are cups
Tires are shoes	Boats are skirts
Windows are eyes	Flags are phones
Clouds are sponges	Pots are paintbrushes
Pools are bowls	Giraffes are snowflakes
Fireworks are ornaments	Forks are strawberries
Grasses are rugs	Pennies are sunglasses
Chimneys are volcanoes	Dogs are scissors

Table 2.2 Experiment 1 Test Trials

2.2.2 Results & Discussion

2.2.2.1 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, $p < .001$. 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 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, $p = .04$.

2.2.2.2 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, $M = 13.67\%$, $SE = 2.91\%$, $t(127) = 4.70$, $p < .001$, $d = .42$. 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, $M = 54.39\%$, $SE = 2.43\%$, $t(127) = 1.81$, $p = .07$.

2.2.2.3 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're both flat") 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 (see Figure 2.1). 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) = .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.

Additionally, we found no age differences between the three groups of explainers. A one-way between-subjects ANOVA found no effect between explanation type (Functional, Perceptual, Irrelevant) and age, F(2,126) = 2.17, *p* = .19. Similarly, using Welch's t-test to account for unequal variance due to the different sample sizes of the explanation groups, we find no difference in the ages of children across different explanation groups. Specifically, we find that the age of children who provided functional explanations (*M* = 5.01 years, *SE* = .09, range = 4.17 – 5.88 years) was not significantly different from the age of children who provided perceptual explanations (*M* = 4.85 years, *SE* = .09, range = 4.01 – 5.87 years), t(62.88) = 1.22, *p* = .23, or the age of children who provided irrelevant explanations (*M* = 4.78 years, *SE* = .06, range = 4.02 - 5.88 years) after a Bonferroni correction for multiple comparisons, t(64.72) = 2.09, *p* = .04. Likewise, the age of children who provided perceptual explanations was also not signficantly different from the age of children who provided irrelevant explanations, t(61.42) = .65, *p* = .52. We also conducted Welch's t-tests to examine whether the ages of any of the explanation groups was significantly different from the overall age of the entire sample of 128 children. We found no significant difference between the age of the children in the entire sample and the age of the children who provided functional explanations, t(50.13) = 1.52, *p* = .14, the age of the children who provided functional explanations, t(47.60) = 0.05, *p* = .96, and the age of the children who provided irrelevant explanations, t(125.17) = .98, *p* = .33. Thus, all our age analyses showed no age differences across the three different types of explainers.

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 the quarter of the 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., 2020; 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. However, there was no effect of the causal framework training trials on the metaphor comprehension test trials.

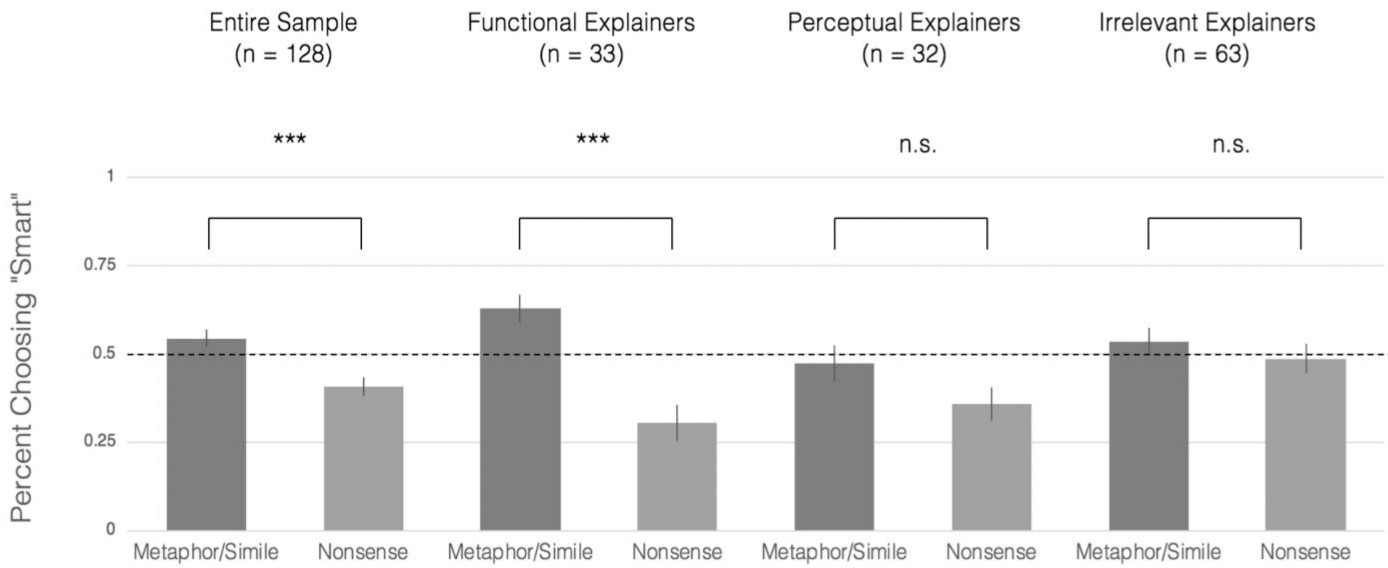

Figure 2.1 Test trial data from preschoolers. Error bars show 1 standard error.

2.3 Experiment 1b

Since our "smart" and "silly" judgment task is a novel experimental paradigm, we ran a sample of adults in order to validate the paradigm. Comprehension of novel metaphors can be challenging 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 and would rate metaphors as "smart" and nonsense statements as "silly".

2.3.1 Methods

2.3.1.1 Participants

We tested 32 participants per condition, leading to a total of 64 adult participants (M = 24.70 years; SD = 5.97 years; range = 18.62 – 41.02 years; 25 males, 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.

2.3.1.2 Stimuli & 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 1a.

2.3.2 Results & Discussion

2.3.2.1 Training Trials

There was no significant difference in training trial performance between conditions; indeed, adults performed identically in the two conditions, $t(62) = 0$, $p = 1.00$. Adults were 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, $M = 93.75\%$, $SE = 1.94\%$, $t(31) = 22.50$, $p < .001$.

2.3.2.2 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. Aggregating together adults' responses from both the Causal Metaphor and Causal Simile conditions, we found that overall Accuracy, as measured by the Composite Scores, is significantly different from chance levels, $t(63) = 24.50$, $p < .001$, $d = 4.33$. Additionally, we found that Accuracy 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 metaphors. 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, $t(62) = 2.21$, $p = .03$. 4 out of 32 adults in the Causal Metaphor condition rated all metaphor and nonsense statements as "silly", thus driving down the overall percentage of metaphors rated as "smart" in the Causal Metaphor condition. In contrast, none of the adults in the Causal Simile condition rated all metaphors and nonsense statements as "silly".

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", $M = 92.97\%$, $SE = 2.24\%$, $t(31) = 19.18$, $p < .001$. 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.

The results of Experiment 1b validate our paradigm, by showing that adults in both conditions judge metaphors as significantly "smart" and nonsense statements as significantly "silly." However, consistent with previous work demonstrating that novel metaphor comprehension is difficult even for adults (Blasko & Connine, 1993), we find that adults are not always at ceiling at this task, especially in terms of rating metaphors as "smart". Interestingly, while there was no difference between preschoolers' "smartness" ratings of metaphors and similes, adults rated similes as smarter than metaphors. This result is consistent with previous work showing that adults prefer novel comparisons, such as the stimuli used in this experiment, in literal simile form rather than non-literal metaphor form (Bowdle & Gentner, 2005).

2.4 Experiment 2

The results of Experiment 1a and 1b demonstrate that both preschoolers and adults are capable of differentiating metaphors from nonsense statements. However, an outstanding question is whether preschoolers' performance in Experiment 1a was actually driven by their understanding of functional similarities between objects in the functional metaphors. Although a quarter of preschoolers provided functional explanations to justify their choices in Experiment 1a, it is still possible that preschoolers do not have sufficient understanding of the objects' functions (e.g. preschoolers do not possess the background knowledge that clouds store water), or judge the metaphors based on other kinds of non-functional similarities (e.g. preschoolers think clouds and sponges are alike because both are fluffy, not because both hold water). Using an established paradigm (Gentner, 1988), Experiment 2 seeks to validate and strengthen the results of Experiment 1a, by demonstrating that preschoolers notice the functional similarities in the functional metaphors used in Experiment 1a. Thus, Experiment 2 explores whether preschoolers preferred *functional explanations* (i.e. explanations involving functional similarities between two concepts) over *perceptual explanations* (i.e. explanations involving perceptual similarities between two concepts) when interpreting functional metaphors.

2.4.1 Methods

2.4.1.1 Participants

We tested 24 participants per condition in two conditions, leading to a total of 48 children who participated in the study (M = 5.02 years; SD = .62 years; range = 4.01 – 5.93 years; 28 males, 20 females). Researchers tested an additional participant, whose data were excluded because they failed the attention check. Children were recruited and tested in a quiet preschool or museum setting.

2.4.1.2 Stimuli & Procedure

As in Experiment 1a, the experimenter presented participants with the stimuli on a laptop computer. Each child was presented with four training trials and eight test trials, and participated in one of two conditions: the Causal Metaphor condition or the Control Metaphor condition. The training trials differed between the two conditions, but the test trials were identical between the two conditions. Experiment 2 used the same functional metaphors as Experiment 1, in both the training and test phase.

2.4.1.2.1 Training Trials. The Causal Metaphor training trials in Experiment 2 were identical to the Causal Metaphor training trials in Experiment 1a. The Control Metaphor training trials in

Experiment 2 were almost identical to the Control Simile training trials in Experiment 1, except that all statements were presented as metaphors instead of similes.

2.4.1.2.2 Test Trials. Participants in both the Causal Metaphor condition and Control Metaphor condition received identical test trials. The experimenter introduced the test trials by saying, “You did such a good job at that game! Now we’re going to play a new game! In this game we’re going to play with Annie’s friend Meg. Meg is going to ask questions. One person will give her an answer to her question. Then, another person will give her a different answer to her question. Your job is to point at the person who gives Meg the better answer. Let’s play!”

On each trial, Meg posed a question (e.g. “How are clouds sponges?”) as the two objects in the metaphor (e.g. a cloud and a sponge) appeared on the screen. Two people then appeared at the bottom of the screen. First, one person appeared on the left and provided an explanation (e.g. “Clouds are sponges because both give water!”). Then, another person appeared on the right and provide an explanation (e.g. “Clouds are sponges because both are fluffy!”). The experimenter prompted the participant to choose an explanation by asking, “Whose answer is better?” Once the participant answered by pointing at one of the two people or providing a verbal response (e.g. “The one who said fluffy”), the experimenter began the next trial. No feedback was provided.

There were eight test trials total, with each trial involving one of the eight functional metaphors from the test trials in Experiment 1. We counterbalanced whether the functional explanation appeared on the left or the right. For a full list of stimuli, see Table 2.3.

Since we did not ask participants to provide their own explanations in Experiment 2, we added an attention check at the end of the study. In the attention check trial, Meg asked, “What is this animal called?” while a picture of a dog appeared on the screen. The person on the left provided the correct description (i.e. “The animal is a dog!”) and the person on the right provided an incorrect description (i.e. “The animal is a fish!”). Children needed to select the correct description in order to pass the attention check.

Functional Metaphors	Functional Explanation	Perceptual Explanation
Roofs are hats	Both keep you dry	Both have pointy tops
Tires are shoes	Both help you go places	Both are made of rubber
Windows are eyes	You see out of both of them	Both are shiny
Clouds are sponges	Both give water	Both are fluffy
Pools are bowls	Both hold liquids	Both have rims
Fireworks are ornaments	Both decorate things	Both are round
Grasses are rugs	You walk on both of them	Both have long tips
Chimneys are volcanoes	Both have smoke coming out	Both are brown and red

Table 2.3 Experiment 2 Test Trials.

2.4.2 Results & Discussion

2.4.2.1 Training Trials

There was no significant difference in training trial performance between the Causal and Control conditions, t(46) = .99, p = .33. In fact, preschoolers were significantly more likely to select the functional match over the object match in both the Causal Metaphor condition, M = 73.96%, SE = 5.32%, t(23) = 4.51, p < .001, and the Control Metaphor condition, M = 66.67%, SE = 5.14%, t(23) = 3.24, p = .004. Preschoolers' ability to select the functional match over the object match in both conditions suggests that preschoolers already have some competence with relational reasoning.

2.4.2.2 Test Trials

Similar to the training trial results, there was also no significant difference in test trial performance between the Causal Metaphor condition and the Control Metaphor condition, t(46) = .54, p = .59. Consequently, we aggregated data across conditions and analyzed them together. We find that, when interpreting functional metaphors, preschoolers were significantly more likely to select functional explanations than perceptual explanations as the better interpretation of the functional metaphor, M = 69.79%, SE = 2.88%, t(47) = 6.87, p < .001, d = .99 (see Figure 2.2). Additionally, when examining individual participant responses, we find that none of the 48 preschoolers in the sample consistently preferred perceptual explanations (i.e. by selecting functional explanations on zero, one, or two of eight test trials). Rather, the individual participant responses ranged from a minimum of chance performance (i.e. selecting functional explanations on at least three out of eight test trials) to a maximum of consistent, unanimous preference for functional explanations (i.e. selecting functional explanations on eight out of eight test trials). Overall, the results of Experiment 2 suggest not only that preschoolers are capable of understanding the functional similarities between two objects in a functional metaphor, but also that preschoolers interpret functional metaphors based on functional similarities rather than perceptual similarities.

2.5 Experiment 3

The results from Experiment 1a show that some preschoolers differentiate between functional metaphors and nonsense statements; moreover, the results from Experiment 2 suggest that this differentiation occurs because preschoolers are capable of recognizing the functional similarities between objects in a functional metaphor. However, while the preschoolers in Experiment 1a were able to rate functional metaphors as "smarter" than nonsense statements, the overall sample of 128 preschoolers did not rate functional metaphors as "smart" above chance levels. Thus, in Experiment 3, we use another paradigm in order to provide converging evidence to support Experiment 1a's claim that preschoolers are capable of differentiating between functional metaphors and nonsense statements. Specifically, in Experiment 3, we use a dichotomous choice task that directly contrasts functional metaphors against nonsense statements. In previous research (Vosniadou & Ortony, 1983), preschoolers presented dichotomous choice tasks that directly contrasted objects based on perceptual similarities against objects with no discernible similarity (e.g. "Is a sun like an orange or a chair?") were able to select the perceptual match (e.g. the orange) over the nonsense match (e.g. the chair). While this previous research suggests an emerging competence with metaphors based on surface-level similarities such as color or shape, the present experiment is the first to explore this kind of dichotomous choice paradigm with metaphors based on more abstract, conceptual similarities, such as functional metaphors.

2.5.1 Methods

2.5.1.1 Participants

We tested 24 participants in this study (*M* = 5.33 years; SD = .55 years; range = 4.11 – 5.95 years; 11 males, 13 females). One additional participant was tested but excluded due to fussiness. Children were recruited via email from a local Bay Area child database and tested online over Zoom. All children viewed the stimuli using a computer or tablet. The experimenter asked parents to help standardize the experimental set-up by entering full-screen mode, hiding their own videos, and moving the experimenter's video to the bottom-center of the screen.

2.5.1.2 Stimuli & Procedure

The experimenter presented participants with the stimuli on a laptop computer. The experimenter introduced the study by saying, "We're going to play with my friend Meg. Meg is going to ask questions. One person will give her an answer to her question. Then, another person will give her a different answer to her question. Your job is to point at the person who gives Meg the better answer. Let's play!"

On each trial, Meg appeared on the screen and posed a question (e.g. "Can you tell me something about windows?"). Two people then appeared at the bottom of the screen. First, one person appeared on the left and provided a statement (e.g. a functional metaphor such as "Windows are eyes!"), as the two objects in the statement (e.g. a window and an eye) appeared on screen in a speech bubble. Then, another person appeared on the right and provided a statement (e.g. a nonsense statement such as "Windows are skirts!"), as the two objects in the statement (e.g. a window and a skirt) appeared on screen in a speech bubble. The experimenter prompted the participant to choose an explanation by asking, "Whose answer is better?" Once the participant provided a verbal response (e.g. "The person who said that windows are eyes", "Eyes!"), the experimenter began the next trial. No feedback was provided. On the last trial, after the participant made a selection between the functional metaphor or the nonsense statement, the experimenter asked for an open-ended explanation about the similarity between the two components of whichever statement the participant chose (e.g. "How are roofs like hats? How are these two things alike?").

There were eight test trials total, with each trial involving one of the eight functional metaphors from the test trials in Experiment 1. We counterbalanced whether the functional metaphor appeared on the left or the right. For a full list of stimuli, see Table 2.4.

Functional Metaphors	Nonsense Statements
Roofs are hats	Roofs are scissors
Tires are shoes	Tires are paintbrushes
Windows are eyes	Windows are skirts
Clouds are sponges	Clouds are phones
Pools are bowls	Pools are strawberries
Fireworks are ornaments	Fireworks are cups
Grasses are rugs	Grasses are snowflakes
Chimneys are volcanoes	Chimneys are sunglasses

Table 2.4 Experiment 3 Trials

2.5.2 Results & Discussion

2.5.2.1 Test Trials

We first analyzed the data from the entire sample of preschoolers and found that overall, preschoolers are significantly above chance at selecting the functional metaphor over the nonsense statement, $M = 76.56\%$, $SE = 4.34\%$, $t(23) = 6.12$, $p < .001$, $d = 1.25$. This result suggests that preschoolers are not only capable of differentiating between functional metaphors and nonsense statements, but also prefer functional metaphors to nonsense statements.

2.5.2.2 Explanations

As in Experiment 1a, we examined the explanations that preschoolers provided for how the two objects they selected on the last trial were alike (e.g. "How is a roof like a hat?"). There were 24 explanations total, as each child provided an explanation on the final trial. We used the same three explanation categories from Experiment 1a (functional, perceptual, and irrelevant) to code the explanations in Experiment 3. 50% of preschoolers (12 out of 24) provided functional explanations that appealed to the function or structure of the two objects (e.g. "because you can see through them", "because hat is on top of the head and roof is on top of the house"). Only 17% of preschoolers (4 out of 24) provided perceptual explanations that appealed to surface-level similarities between the two objects (e.g. "because they're kind of shaped like rectangles", "because they both look a little bit pointy"). Thirty-three percent of preschoolers (8 out of 24) provided irrelevant explanations (e.g. "I like strawberries and I like going to the pool", "I don't know why"). Two coders coded all explanations. Intercoder reliability was 83%, converging on the same category for 20 out of 24 explanations. The categorization of the remaining four explanations was resolved through discussion.

We analyzed data from the children who provided functional explanations, perceptual explanations, and irrelevant explanations separately, examining whether children from each explanation group were able to select the functional metaphors over the nonsense statements at significantly above-chance levels (see Figure 2.2). We find that the children who provided functional explanations ($n = 12$) were significantly more likely to select functional metaphors over nonsense statements, $M = 85.42\%$, $SE = 3.38\%$, $t(11) = 10.47$, $p < .001$. In contrast, children who provided perceptual explanations ($n = 4$) did not select functional metaphors over nonsense statements at above-chance levels, $M = 78.13\%$, $SE = 9.38\%$, $t(3) = 3.00$, $p = .06$. Similarly, children who provided irrelevant explanations (*n* = 8) also did not select functional metaphors over nonsense statements at above-chance levels, *M* = 62.5%, *SE* = 9.74%, t(7) = 1.28, *p* = .24.

Finally, we examined whether there were age differences across the three different explanation groups. A one-way between-subjects ANOVA found no effect between explanation type (Functional, Perceptual, Irrelevant) and age, F(2,21) = 1.72, *p* = .20. Similarly, using Welch's t-test to account for unequal variance due to the different sample sizes of the explanation groups, we find no difference in the ages of children across different explanation groups. Specifically, we find that the age of children who provided functional explanations (*M* = 5.49 years, *SE* = .12, range = 4.68 – 5.95 years) was not significantly different from the age of children who provided perceptual explanations (*M* = 5.11 years, *SE* = .40, range = 4.11 – 5.84 years), t(3.56) = 1.02, *p* = .37, or the age of children who provided irrelevant explanations (*M* = 5.14 years, *SE* = .20, range = 4.11 - 5.73 years), t(12.09) = 1.71, *p* = .11. Likewise, the age of children who provided perceptual explanations was also not signficantly different from the age of children who provided irrelevant explanations, t(4.54) = .06, *p* = .95. We also conducted Welch's t-tests to examine whether the ages of any of the explanation groups was significantly different from the overall age of the entire sample of 24 children. We found no significant difference between the age of the children in the entire sample and the age of the children who provided functional explanations, t(28.30) = 1.23, *p* = .23, the age of the children who provided functional explanations, t(3.50) = 0.53, *p* = .63, and the age of the children who provided irrelevant explanations, t(11.89) = .85, *p* = .41. In summary, all of the age analyses found no age difference across the three types of explainers.

Thus, similar to the results in Experiment 1a, the overall significant result in Experiment 3 was largely driven by a subset of children who explicitly provided functional explanations to justify their choices. In Experiment 1a, which presented functional metaphors and nonsense statements across trials, 25% of preschoolers in the sample were able to notice functional similarities and provide functional explanations. In Experiment 3, when presented a dichotomous choice task that explicitly contrasted functional metaphors against nonsense statements within trials, even more preschoolers (50% of the sample) were able to notice the functional similarities and provide functional explanations.

The performance of the perceptual explainers in Experiment 3 was marginally significant, such that the perceptual explainers were marginally more likely to select the functional metaphors over the nonsense statements. However, it is worth noting that the proportion of perceptual explainers in Experiment 3 was extremely small: only 4 out of 24 children provided perceptual explanations. Moreover, perceptual explainers must still use some form of relational reasoning to guide their choices in this task. Even if perceptual explainers do not notice the similarity between the two objects in the functional metaphors, they may notice the lack of similarity between the two objects in the nonsense statements, and use a mutual exclusivity strategy (e.g. Halberda, 2003) to guide their choices. Specifically, preschoolers in the perceptual explanation category may reason that objects in nonsense statements, such as "tires are paintbrushes", have absolutely no similarity, and thus that the functional metaphors, such as "tires are shoes", must be the better choice, even if they do not notice the functional similarity between the two objects in the functional metaphor. Thus, even the small proportion of perceptual explainers in Experiment 3 demonstrate some capacity for relational reasoning, and overall, Experiment 3 shows that preschoolers are capable of differentiating between functional metaphors and nonsense statements.

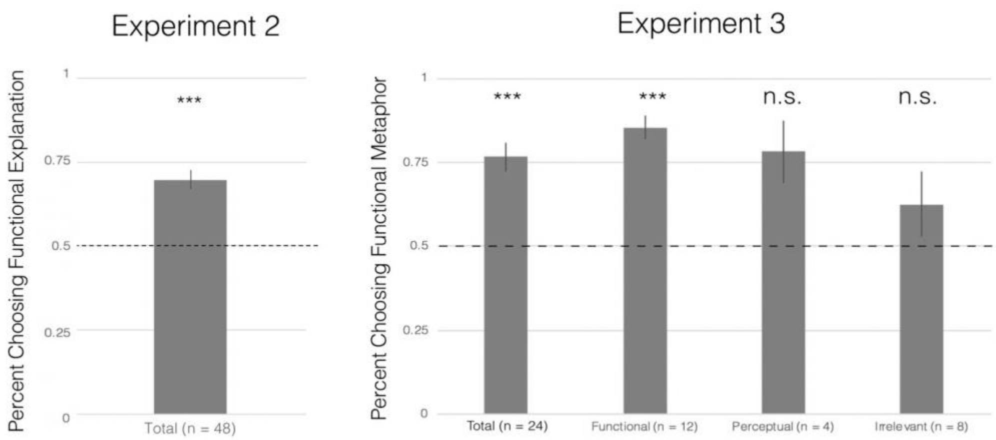

Figure 2.2 Test trial data from Experiment 2 and 3. Error bars show 1 standard error.

2.6 General Discussion

This paper introduces novel paradigms that investigate preschoolers' capacity to reason about abstract relations, such as shared function, and to understand metaphors in an adult-like fashion. Overall, findings suggest that preschoolers distinguish between functional metaphors and nonsense statements by noticing the functional similarities between objects in functional metaphors. In Experiment 1a, preschoolers rated functional metaphors (e.g. "tires are shoes") as "smarter" than nonsense statements (e.g. "boats are skirts") in a "smart" or "silly" absolute judgment paradigm. In Experiment 2, preschoolers preferred functional explanations (e.g. "both give water") over perceptual explanations (e.g. "both are fluffy") when interpreting the functional metaphors (e.g. "clouds are sponges") used in Experiment 1. In Experiment 3, preschoolers preferred functional metaphors (e.g. "roofs are hats") over nonsense statements (e.g. "roofs are scissors") in a dichotomous-choice preference paradigm. Additionally, preschoolers not only selected functional explanations to interpret functional metaphors in Experiment 2; over a quarter of preschoolers in Experiment 1a and half of preschoolers in Experiment 3 also explicitly articulated the functional similarities between two objects (e.g. "the hat shades you and the top of the roof does too"; "you can drive with wheels and walk with feet"), and the performance of these subsets of children drove the success of the entire sample in both studies. Preschoolers' sophisticated functional explanations are consistent with previous work showing that preschoolers are capable of reasoning about abstract relations (Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017) and the functions of objects (Diesendruck et al., 2003; Haward et al., 2018). Taken together, these three experiments demonstrate that children possess the capacity to understand metaphors based on abstract similarities at a much earlier developmental timepoint than previously assumed (Demorest et al., 1983; Silberstein et al., 1982; Winner et al., 1980).

In contrast to previous research demonstrating a facilitative effect of causal framing on early relational reasoning abilities (Goddu et al., 2020; Walker et al., 2016; Walker & Gopnik, 2018), we did not find such an effect on preschoolers' performance in metaphor comprehension tasks in Experiments 1a and 2. This lack of effect may be due to the fact that preschoolers were already surprisingly successful at multiple metaphor comprehension tasks without any kind of training.

Specifically, in Experiment 1a, preschoolers in the baseline condition without training trials performed equally well in the metaphor comprehension test trials as preschoolers in the three other conditions with training trials. Moreover, in Experiment 3, preschoolers successfully differentiated between functional metaphors and nonsense statements without any kind of training or warm-up. Thus, preschoolers were able to spontaneously apply relational reasoning skills in multiple metaphor comprehension tasks, and consequently did not require any kind of training - causal or otherwise – to elicit a “relational mindset” (e.g. Goldwater & Jamrozik, 2019; Simms & Richland, 2019). Indeed, preschoolers’ ability to spontaneously apply relational reasoning skills to multiple metaphor comprehension tasks once again suggest that preschoolers’ competence with abstract metaphors and relational reasoning may have been underestimated previously. With our novel paradigms, we were able to demonstrate preschoolers’ early ability to understand functional metaphors.

In addition to demonstrating preschoolers’ competence with metaphors, our paradigms find that preschoolers do not have more difficulty interpreting non-literal language, such as metaphors, than literal language, such as similes. Specifically, we find no difference between preschoolers’ performance when presented with metaphors (e.g. “Clouds are sponges”) or similes (e.g. “Clouds are *like* sponges”) in Experiment 1a. Indeed, the preschoolers in Experiment 1a were actually more flexible and accepting of non-literal language than adult in Experiment 1b, who rated metaphors as significantly “sillier” than similes. While adults show a preference for novel linguistic comparisons in simile rather than metaphor form (Bowdle & Gentner, 2006), children have not developed this preference, and perform equally well with metaphors and similes. Moreover, preschoolers performed successfully in Experiments 2 and 3, which were conducted solely with metaphors as opposed to similes. Consequently, the results of the three current experiments demonstrate that preschoolers did not have difficulty with the non-literal aspect of metaphors.

These positive findings pave the way for new and exciting future research directions. For example, while the current research focuses on the success of the overall sample of preschoolers, the results of Experiment 1a and 3 show that the success of the overall group is driven by a subset of preschoolers who can explicitly provide functional explanations. One interesting future direction might be to explore why some preschoolers, but not others, are capable of understanding functional metaphors. For example, perhaps the preschoolers providing explicit explanations have better relational reasoning skills, better executive function abilities (e.g. Ballestrino et al., 2016), or more conceptual knowledge about the items in the metaphors (Keil, 1986). While relational reasoning plays a crucial role in metaphor comprehension (Bowdle & Gentner, 2005; Gentner, 1988; Gentner & Clement, 1988; Holyoak, 2019), follow-up studies exploring individual differences might shed light on how other cognitive mechanisms, such as conceptual knowledge and executive function, also contribute to metaphor comprehension.

Additionally, while the present findings demonstrate that preschoolers can differentiate functional metaphors from nonsense statements, future research could explore whether preschoolers understand metaphors presented in other, more naturalistic settings. Experimental studies on children’s metaphor comprehension frequently juxtapose multiple metaphors against each other (e.g. Silberstein et al., 1982; Vosniadou & Ortony, 1983), but more naturalistic contexts such as parent-child conversations or written poetry might focus on a single metaphor at a time. Thus, while Experiment 1a and 3 use novel paradigms involving both functional metaphors and nonsense statements, future work might investigate whether preschoolers also understand metaphors under different circumstances, without the direct juxtaposition of nonsense statements.

Another exciting potential research question is whether preschoolers are capable of using metaphor and relational reasoning in the service of other complex learning processes, such as thinking and reasoning about abstract concepts in the contexts of scientific discovery (Kuhn, 1993) and conceptual change (Xu, 2019). Researchers have argued that linguistic metaphors provide useful conceptual frameworks, allowing new, insightful ways of reasoning about old concepts (Thibodeau et al., 2017), as well as facilitating the acquisition of novel concepts and word meanings (Bowdle & Gentner, 2005; Holyoak & Stamenkovic, 2018). Thus, metaphors could potentially be powerful tools for children, helping them acquire more information about the world, but no research to date has demonstrated that children can learn from metaphors.

Moreover, additional work with non-Western populations is required to determine the generalizability of the current findings. Given the evidence of cross-cultural variation in the development of relational reasoning abilities (Carstensen et al., 2019), it is possible that children in other cultures may also understand metaphors sooner or later in development than U.S. children. A limitation of our current work is its reliance on WEIRD (i.e. Western, educated, industrialized, rich, and Democratic; Heinrich et al., 2010) convenience samples; thus, future work should investigate the possibility of early cross-cultural diversity in children's metaphor comprehension abilities.

Overall, the current research shows that 4- to 5-year-olds are already capable of understanding functional metaphors based on abstract similarities between two disparate concepts. Preschoolers' success with functional metaphors provides exciting groundwork for future research on children's early comprehension of non-literal language, and how children think, reason, and learn more broadly.

Chapter 3: Providing explanations shifts preschoolers' metaphor preferences

3.1 Introduction

In order for metaphors to facilitate learning, children must not only understand metaphors, but also be able to appreciate the relative informativeness of the commonalities that metaphors highlight. Two concepts can be the same or different along an infinite number of dimensions, but some dimensions are more informative and useful for learning than others. In particular, relations that are more abstract are often better for learning. For example, it is useful to know that dogs and cats are the same in that they are both animals, but less useful to know that they are the same in that both existed in medieval France. Both similarities are true, but the former similarity – "both animals" – facilitates category learning, while the latter similarity – "both existing in medieval France" – is mere trivia. The capacity to select the most informative analogy is crucial for successful learning and problem-solving (Richland & McDonough, 2010). Yet, selecting an informative or relevant analogy is also a difficult task, given the infinite set of analogies to choose from (i.e., because two concepts can be the same or different across an infinite number of dimensions). Thus, the cognitive process of learning from analogy requires that children be able not only to *represent* similarities, but also to *select* the most informative analogies to pursue from infinitely large set of possible options.

Moreover, given that children are active learners who do not merely receive passive information from the environment, but rather actively and rationally seek information to learn from (Xu, 2019), it is possible that children also benefit most when they can independently recognize the relevant analogies for themselves, rather than passively receiving the information from others. Indeed, research on children's acquisition of natural number – a prominent case study of childhood conceptual change that involves noticing critical analogies (Carey, 2009; Marchand & Barner, 2018) – demonstrates that parental instruction generally does *not* help children learn natural numbers (Ramscar et al., 2011). Rather, children acquire an understanding of natural number in a protracted, piecemeal fashion (Carey, 2009; Wagner et al., 2015; Wynn, 1992), possibly arriving at the relevant analogy slowly and independently. Thus, preschoolers' ability to actively notice relevant metaphors and analogies may be crucial to their learning process. Overall, in order to learn from metaphors and analogies, young children must not only understand metaphors, but also notice and select the most informative metaphors to learn from.

Thus, an outstanding question is whether preschoolers can appreciate which metaphors allow for the most learning (e.g. by licensing additional inferences) and to prefer those metaphors over others. Previous research suggests that the answer is no: while adults prefer functional metaphors that highlight abstract features (Gentner & Clement, 1988), preschoolers prefer perceptual metaphors that highlight arbitrary, surface-level features (Silberstein et al., 1982). Thus, previous research suggests that while adults may prefer functional metaphors like "Tires are shoes", which may license further inferences (i.e. tires and shoes both facilitate transportation, and thus both must be made of similarly durable materials like rubber), preschoolers may prefer perceptual metaphors like "Tires are donuts", which are not useful for additional learning (i.e. the fact that tires and donuts both have holes is a mere coincidence). Consequently, even though some preschoolers are able to understand functional metaphors, they may not recognize that the comparisons drawn by functional metaphors provide more relevant information for learning than those drawn by perceptual metaphors. If children are simply unable to appreciate the greater relative informativeness of functional over perceptual metaphors (i.e., if they are insensitive to the benefits of abstract, functional commonalities over superficial, perceptual commonalities for learning), then the utility of metaphor as an early childhood learning mechanism is severely limited. If, however, children to correctly distinguish between the relative informativeness of functional versus perceptual metaphors in some contexts, then this would suggest that children have the skills to learn from metaphors far earlier than previously believed. It would also provide additional evidence to the growing body of research suggesting early competence in abstract relational thought (Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017; Walker et al., 2016), as well as suggest interventions that could allow children to learn from metaphors in the powerful ways that adults do.

The current study asks whether preschoolers' preferences might *shift from perceptual metaphors to functional metaphors* when they encounter metaphors in a pedagogical context (Experiment 1), when they are provided with explanations for the conceptual similarities that underlie each metaphor (Experiment 2), and when they are provided with explicit instructions to convey either functional or perceptual information (Experiment 3). We test two experimental manipulations that might facilitate preschoolers' performance. First, previous research suggests that preschoolers are sensitive to pedagogical contexts: for example, they flexibly select what information to teach others (Bridgers et al., 2019) and who to learn from (Gweon et al., 2018; Gweon & Asaba, 2018; Koenig & Harris, 2005; Sobel & Kushnir, 2013) depending on the circumstances. Thus, Experiment 1 asks whether adults and preschoolers shift their metaphor preferences in a context in which they are instructed to convey information to a naïve agent. Second, since preschoolers sometimes struggle to notice relations (Kroupin & Carey, 2021), Experiment 2 asks whether preschoolers shift their metaphor preferences when provided with explanations that highlight how two concepts in a metaphor are alike. Finally, Experiment 3 asks whether preschoolers can indeed differentiate between functional and perceptual metaphors, when given explicit instructions to communicate either functional or perceptual information.

3.2 Experiment 1

In Experiment 1, we investigated whether preschoolers might be capable of shifting their preferences away from perceptual metaphors (e.g., "Eyes are buttons"), and towards functional metaphors (e.g., "Eyes are windows"), given a pedagogical context (i.e., a context in which they were asked to teach someone else). We tested both adults and preschoolers in either the Pedagogical condition, in which participants helped teach a naïve agent, or the Baseline condition, in which no context was given to guide participants' choices. We hypothesized that both adults and preschoolers would shift their metaphor preferences given a pedagogical context, such that both adults and preschoolers would be more likely to prefer functional metaphors over perceptual metaphors in the Pedagogical condition relative to the Baseline condition.

3.2.1 Methods

3.2.1.1 Participants

We adhered to a stopping rule of 24 participants per condition, leading to a total of 48 adult participants (*M* = 25.93 years; SD = 6.57 years; range = 18.83 – 48.61 years; 16 males) and 48 4- and 5-year-olds participants (*M* = 4.93 years; SD = .55 years; range = 4.02 – 5.91 years; 24 males). Researchers tested an additional child, whose data were excluded due to failure on the attention check. Adults were recruited and tested in-person, on a university campus. Most children (45 out of 48) were recruited and tested in-person, in a preschool or museum. Due to COVID-19, three children were recruited from a local database and tested online over Zoom. All experiments reported in this paper were approved by the university's Committee for the Protection of Human Subjects. All adult participants and parents of child participants provided informed consent. The preschooler component of Experiment 1 is preregistered at https://osf.io/uptsf/.

3.2.1.2 Stimuli & Procedure

The experimenter presented participants with stories on a computer. Participants were assigned to either the Pedagogical or Baseline condition.

3.2.1.2.1 Pedagogical Condition. In the Pedagogical condition, the experimenter showed a picture of an alien and said, "We're going to play with my friend Zorpa. I've got something special to tell you about Zorpa. Zorpa is actually an alien from Planet Meelee! So, she doesn't know anything about the objects on Earth. We need your help teaching Zorpa about the objects here on Earth. In this game, Zorpa is going to ask two teachers about any object. One teacher will give her an answer. Then, another teacher will give her a different answer. Your job is to figure out which teacher Zorpa should learn from."

On each trial, Zorpa stated what concept she wanted to learn about (e.g., "I want to learn about eyes!"). A teacher appeared on the left side of the screen and provided a metaphor (e.g., a functional metaphor, such as, "This teacher says, 'Eyes are windows!'"). As the metaphor was uttered, the two objects in the metaphor (e.g., eye and window) appeared on the screen. Then, a second teacher appeared on the right side of the screen and provided another metaphor (e.g., a perceptual metaphor, such as "This teacher says, 'Eyes are buttons!'"). Once again, as the metaphor was uttered, the two objects in the metaphor (e.g., eye and button) appeared on the screen. The experimenter then asked, "Which teacher should Zorpa learn from?" Once the participant answered by providing a verbal response (e.g., "buttons") or by pointing at one of the teachers, the experimenter began the next trial. No feedback was provided. For an example of the trial structure, see Figure 3.1.

Importantly, the perceptual metaphors were all based on arbitrary, surface-level similarities that did not indicate any kind of deeper, more abstract similarities, such as shared function or category membership. For example, the fact that tires and donuts both have holes is a mere coincidence, and did not license further inferences that these two objects might possess other, more meaningful similarities. In contrast, the functional metaphors might license further inferences about the objects in question. For example, tires and shoes both help people go places, and consequently are both made from durable materials, like rubber. For a full list of stimuli, see Table 3.1.

Finally, participants completed an attention check at the end of the study. In the attention check, the experimenter asked, “What is this animal called?” while a picture of a dog appeared on the screen. The person on the left provided the correct description (i.e., “The animal is a dog!”) and the person on the right provided an incorrect description (i.e., “The animal is a fish!”). Participants needed to select the correct description in order to pass the attention check.

Each participant received eight metaphor preference trials. Each trial’s structure followed the design described above, in which the participant had to select between a functional metaphor or a perceptual metaphor. The order of the eight trials was randomized and the left-right placement of the functional metaphors was counterbalanced.

3.2.1.2.2 Baseline Condition. In the Baseline condition, preschoolers participated in a very similar dichotomous-choice metaphor preference paradigm, but without any pedagogical framing. The experimenter introduced the task by saying, “We’re going to play with my friend Meg. Meg is going to ask questions! One person will give her an answer to her question. Then, another person will give her a different answer to her question. Your job is to point at the person who gives Meg the better answer. Let’s play!”

The Baseline trials were similar to the Pedagogical trials, with three exceptions. First, while the Pedagogical condition emphasized learning (e.g., “Zorpa says, “I want to *learn* about eyes”), the Baseline condition did not (e.g., “Meg says, “Can you tell me something about eyes?”). Second, while the Pedagogical condition emphasized that the respondents were teachers (i.e., “This *teacher* says...”), the Baseline condition did not (i.e., “This *person* says...”). Third, instead of selecting the teacher who should be learned from, participants were simply asked “Whose answer was better?”

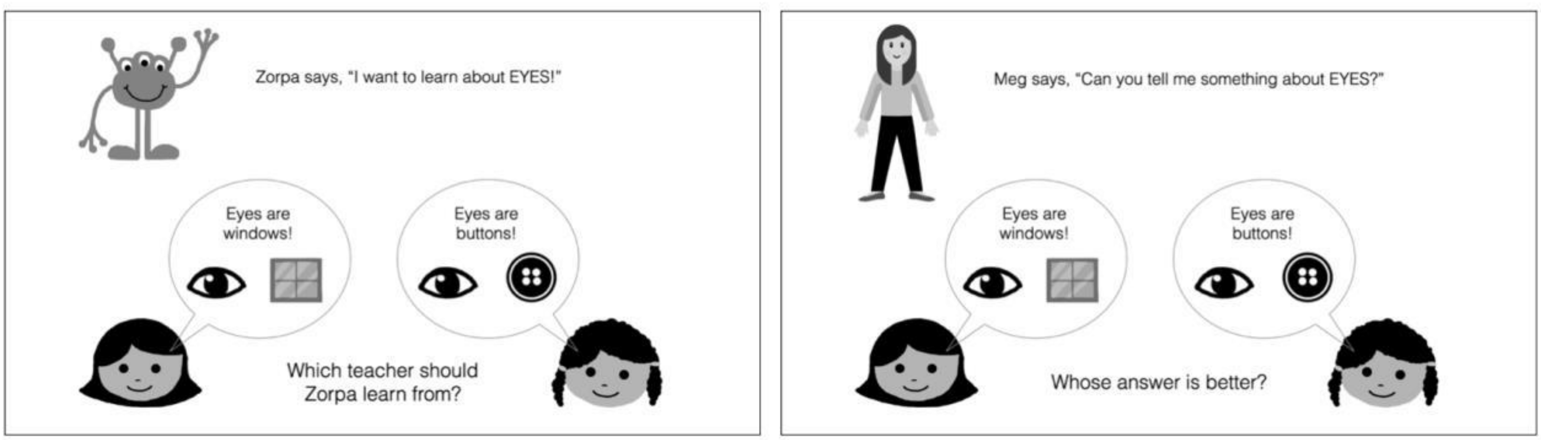

Figure 3.1 An example of an Experiment 1 trial, presented in either the Pedagogical condition (left) or Baseline condition (right).

Functional Metaphors	Perceptual Metaphors	Functional Explanations (Exp 2)	Perceptual Explanations (Exp 2)
Eyes are windows	Eyes are buttons	You see through both of them	Both are round
Moons are lightbulbs	Moons are cookies	Both give light	Both are circles
Clouds are sponges	Clouds are ice creams	Both hold water	Both are fluffy
Ladders are hills	Ladders are traintracks	You climb both of them	Both are long
Roofs are hats	Roofs are clotheshangers	Both cover you	Both are triangles
Tires are shoes	Tires are donuts	Both have holes	Both help you go places
Grasses are rugs	Grasses are hairs	You walk on both of them	Both are pointy
Suns are candles	Suns are oranges	They both light up	They're both the same color

Table 3.1 Columns 1 and 2 list the functional and perceptual metaphors used in Experiments 1-3. Columns 3 and 4 list the explanations used in Experiment 2's Explanation and Explanation & Pedagogy conditions.

3.2.2 Results & Discussion

A between-subjects ANOVA with condition (Pedagogical, Baseline) and age (adult, preschooler) as independent variables yielded a main effect of Age, F(1,92) = 111.40, $p < .001$, and a main effect of Condition, F(1,92) = 4.56, $p = .04$. Adults were significantly more likely to select functional metaphors over perceptual metaphors in both the Baseline condition, $M = 70.31\%$, $SE = 4.18\%$, t(23) = 4.28, $p < .001$, and the Pedagogical condition, $M = 85.42\%$, $SE = 3.89\%$, t(23) = 9.12, $p < .001$. Moreover, there was a significant effect of condition, such that adults in the Pedagogical condition were more likely to select functional metaphors than adults in the Baseline condition, t(46) = 2.65, $p = .01$. In contrast, preschoolers were significantly more likely to select perceptual metaphors over functional metaphors in both the Baseline condition, $M = 31.25\%$, $SE = 4.05\%$, t(23) = 4.63, $p < .001$, and the Pedagogical condition, $M = 34.38\%$, $SE = 4.89\%$, t(23) = 3.19, $p = .004$. There was no difference in preschoolers' performance between the Baseline and Pedagogical conditions, t(46) = .49, $p = .63$. All the significant statistics reported above remained significant after correcting for multiple comparisons (Benjamini & Hochberg, 1995).

Experiment 1 showed that adults', but not preschoolers', performance on the metaphor preference task benefited from pedagogical context (see Figure 3.2). Specifically, while adults already preferred functional metaphors in a contextless baseline condition, they preferred functional metaphors *even more* in a pedagogical context. In contrast, preschoolers' performance on the metaphor preference task did not change across experimental contexts: preschoolers preferred perceptual metaphors over functional metaphors in both contexts.

These results demonstrate that adults are already able to select appropriate metaphors to learn from (i.e., functional metaphors) even without context, and this selection ability increases when a pedagogical context is introduced. However, preschoolers prefer shallow surface-level metaphors (i.e., perceptual metaphors) with or without a pedagogical context. These results suggest that preschoolers are not sensitive to the relative informativeness of functional over perceptual metaphors, even in a pedagogical context.

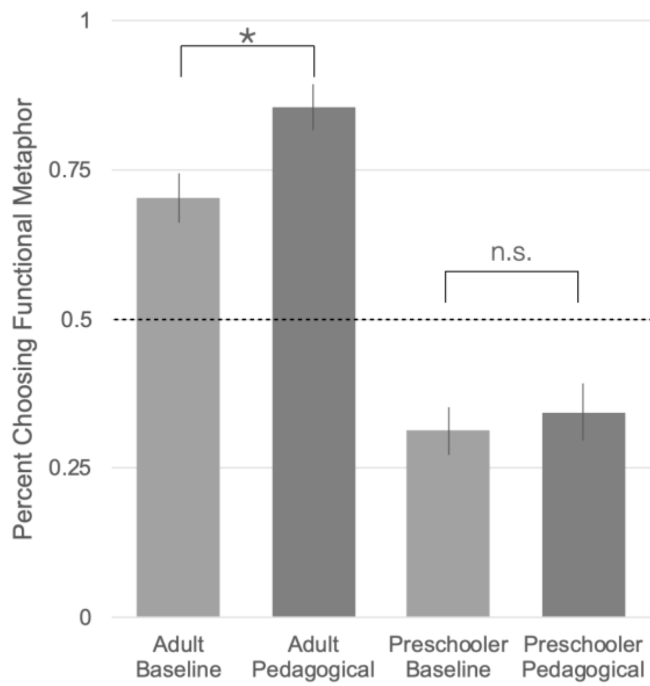

Figure 3.2 Experiment 1 results. Error bars show 1 standard error.

3.3 Experiment 2

Experiment 1 showed that pedagogical context improved adults', but not preschoolers', performance on a metaphor preference task. Although preschoolers are often sensitive to pedagogical contexts (Bridgers et al., 2019; Gweon et al., 2018; Gweon & Asaba, 2018; Koenig & Harris, 2005; Sobel & Kushnir, 2013), Experiment 1's results suggest that they are *not* sensitive to the relative informativeness of functional metaphors over perceptual metaphors when teaching others.

However, it is not clear whether children's failure to select the more informative metaphors for the alien character in Experiment 1 derived from a difficulty with reasoning about which information would be useful to that character, or whether it stemmed from a general inability to appreciate the usefulness of functional over perceptual metaphors. In order to directly test the latter possibility, Experiment 2 investigated whether children were able to select functional metaphors over perceptual metaphors when given explanations for how the two concepts in the metaphors were alike. Earlier work shows that explanations can lead preschoolers to make broader and deeper generalizations, and attend to abstract relations (Walker et al., 2014; 2017). Although young children are capable of representing relations between objects (Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017;) and thus capable of understanding metaphors (Pouscoulous & Tomasello, 2020), preschoolers also sometimes fail to spontaneously notice relations between objects (Kroupin & Carey, 2021). Thus, providing explanations of how two concepts in a metaphor are similar might help preschoolers notice and fully consider the relevant conceptual relations underlying metaphors, and thus facilitate their performance on a metaphor preference task. In Experiment 2, we replicate preschoolers' baseline performance, and investigate whether their performance shifts when 1) provided with explanations, and 2) provided with both explanations *and* pedagogical context.

3.3.1 Methods

3.3.1.1 Participants

We adhered to a stopping rule of 24 participants per condition, leading to a total of 72 4- and 5- year-old participants (*M* = 5.04 years; SD = .60 years; range = 4.01 – 5.99 years; 44 females). Researchers tested six additional children, whose data were excluded due to failure on the attention check (four children), experimenter error (one child), and external interference (one child). All children were recruited from a local participant database and tested online over Zoom. Experiment 2's preregistration can be found at https://osf.io/uptsf/.

3.3.1.2 Stimuli and Procedure

The experimenter presented stories, which participants viewed on either a computer or large tablet. Participants were assigned to the Baseline, Explanation, or Explanation and Pedagogy condition.

3.3.1.2.1 Baseline Condition. Experiment 2's Baseline condition was similar to Experiment 1's Baseline condition, except that data was collected online rather than in person.

3.3.1.2.2 Explanation Condition. Experiment 2's Explanation condition was similar to the Baseline condition in Experiment 1, except that data was collected online and the experimenter provided explanations for how the two concepts in the metaphors were alike (e.g., "This person says, 'Tires are donuts *because both have holes*'"; "This person says, 'Tires are shoes *because both help you go places*'").

3.3.1.2.3 Explanation and Pedagogy Condition. The Explanation and Pedagogy condition was similar to Experiment 1's Pedagogy condition, except that data was collected online and the experimenter provided explanations for how the two concepts in the metaphors were alike (e.g., "This teacher says, 'Eyes are windows *because you see through both of them*"; "This teacher says, 'Eyes are buttons *because both are round"*).

3.3.2 Results & Discussion

In line with previous research and the results of Experiment 1, preschoolers in Experiment 2 were significantly more likely to select perceptual metaphors over functional metaphors in the Baseline condition, *M* = 34.38%, *SE* = 4.53%, t(23) = 3.45, *p* = .002. In contrast, preschoolers did not prefer perceptual metaphors over functional metaphors in the Explanation condition, *M* = 52.60%, *SE* = 4.82%, t(23) = .54, *p* = .59, and the Explanation and Pedagogy condition, *M* = 48.96%, *SE* = 4.32%, t(23) = .24, *p* = .81. Compared to performance in the Baseline condition, preschoolers were significantly more likely to select functional metaphors in the Explanation condition, t(46) = 2.76, *p* = .008, and the Explanation and Pedagogy condition, t(46) = 2.33, *p* = .02. There was no difference in performance between the Explanation condition and the Explanation and Pedagogy condition, t(23) = .56, $p = .58$. All the significant statistics reported above remained significant after correcting for multiple comparisons (Benjamini & Hochberg, 1995).

Experiment 2's results suggest that explanations help preschoolers recognize the informativeness of functional metaphors (see Figure 3.3). When they were provided with explanations for the comparisons between concepts drawn by the metaphors, children's preferences shifted away from perceptual metaphors, and towards functional metaphors. Notably, Experiment 2 demonstrated that preschoolers shifted their preferences towards functional metaphors in not one, but two, conditions involving explanations.

Replicating the results of Experiment 1, there was no effect of pedagogy, as preschoolers performed similarly in the two explanation conditions, with or without pedagogical context. Experiment 2 also replicated Experiment 1's finding that preschoolers prefer perceptual metaphors (e.g., "Tires are donuts") to functional metaphors (e.g., "Tires are shoes") in a contextless baseline condition. Overall, the results of Experiment 2 show that preschoolers' metaphor preferences are not fixed. Given that providing explanations for the comparisons between concepts in the metaphors shifted preschoolers' metaphor preferences away from perceptual metaphors and towards functional metaphors, we can conclude that they are indeed sensitive to the informativeness of functional metaphors. This suggests that, given sufficient scaffolding, even young children are able to appreciate the abstract similarities that make metaphors conducive for learning.

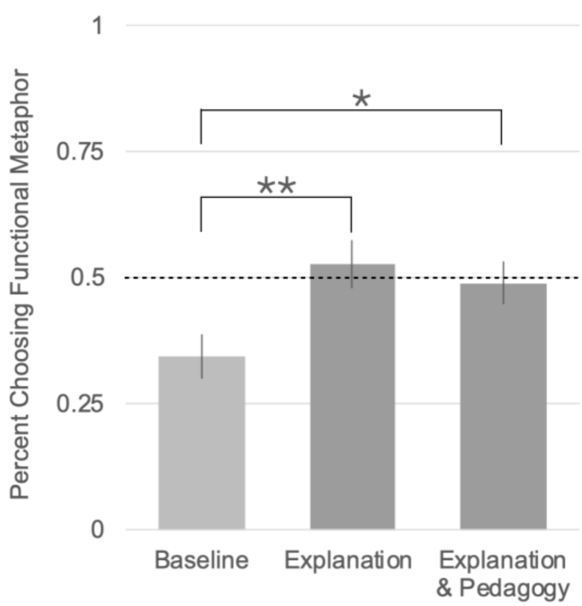

Figure 3.3 Experiment 2 results. Error bars show 1 standard error.

3.4 Experiment 3

In Experiment 3, we seek additional evidence that children do indeed appreciate the difference between functional and perceptual metaphors (Pouscoulous & Tomasello, 2020), even in the absence of explanations that explicitly highlight the similarities between the two concepts in the metaphors. Thus, in Experiment 3 there were two pedagogical conditions *without* explanations, but with more explicit instructions regarding Zorpa's learning objectives.

Specifically, in the Perceptual condition, preschoolers were told to select the perceptual metaphors, because Zorpa wanted to know “what objects look like.” In contrast, in the Functional condition, preschoolers were told to select the functional metaphors, because Zorpa wanted to know “what objects are used for.”

3.4.1 Methods

3.4.1.1 Participants

We adhered to a stopping rule of 24 participants per condition, leading to a total of 48 4- and 5- year-old participants (M = 4.94 years; SD = .53 years; range = 4.00 – 5.98 years; 19 females). Researchers tested three additional children, whose data were excluded due to failure on the attention check. All children were recruited from and tested at local museums and preschools, in a quiet setting.

3.4.1.2 Stimuli & Procedure

The experimenter presented stories, which participants viewed on a laptop computer. Participants were assigned to either the Perceptual or Functional condition.

2.4.1.2.1 Perceptual Condition. Experiment 3’s Perceptual condition was similar to Experiment 1’s conditions, except participants were provided with explicit instructions to select the perceptual metaphor. Specifically, when introducing the task, the experimenter added the following instructions: “Zorpa wants to know what objects *look like*. Okay? Let’s teach Zorpa what objects *look like*. Remember, Zorpa wants to know what objects *look like*.” Then, the experimenter presented the functional and perceptual metaphors, without explanations (e.g., “This person says, ‘Tires are donuts”; “This person says, ‘Tires are shoes”). At the end of each trial, the experimenter reminded participants to select the perceptual metaphor, by saying, “Remember, Zorpa wants to know what objects *look like*. So whose answer is better?”

2.4.1.2.2 Functional Condition. Experiment 3’s Functional condition was similar to Experiment 1’s conditions, except participants were provided with explicit instructions to select the functional metaphor. In the introduction, the experimenter added the following instructions: “Zorpa wants to know what objects *are used for*. Okay? Let’s teach Zorpa what objects *are used for*. Remember, Zorpa wants to know what objects *are used for*.” Additionally, at the end of each trial, the experimenter reminded participants to select the functional metaphor, by saying, “Remember, Zorpa wants to know what objects *are used for*. So whose answer is better?”

3.4.2 Results & Discussion

Experiment 3’s results show that preschoolers are capable of differentiating between perceptual and functional metaphors (see Figure 3.4). Specifically, preschoolers’ responses were significantly different between the Perceptual and Functional conditions, $t(46) = 3.11$, $p = .003$. In the Perceptual condition, preschoolers selected perceptual over functional metaphors above chance levels, *M* = 29.17%, *SE* = 3.96%, t(23) = 5.26, *p* < .001. In the Functional condition, preschoolers selected at chance levels between perceptual and functional metaphors, *M* = 47.92%, *SE* = 4.56%, t(23) = .46, *p* = .65. Additionally, there was no difference between preschoolers' performance on Experiment 3's Functional condition and Experiment 2's Explanation condition, t(46) = .71, *p* = .48; or Experiment 2's Explanation and Pedagogy condition, t(46) = .17, *p* = .87.

Overall, Experiment 3 shows that when given explicit instructions to teach perceptual or functional information – namely, what objects "look like" or "are used for" respectively – preschoolers differentiate between functional and perceptual metaphors. Moreover, there was no difference between Experiment 2's facilitative conditions (i.e., the Explanation condition, and the Explanation and Pedagogy condition) and Experiment 3's functional condition. This result suggests that providing explicit explanations that highlight the similarities between two concepts within a metaphor not only shifts preschoolers' metaphor preferences, but also shifts preschoolers' preferences as much as providing framing that directly instructs them to choose the functional metaphor. Consequently, this reinforces the idea that explanation is a powerful mechanism driving information search (Walker et al., 2014; 2017). Indeed, in the present studies, providing explanations seems to be as powerful as explicitly relaying the information itself.

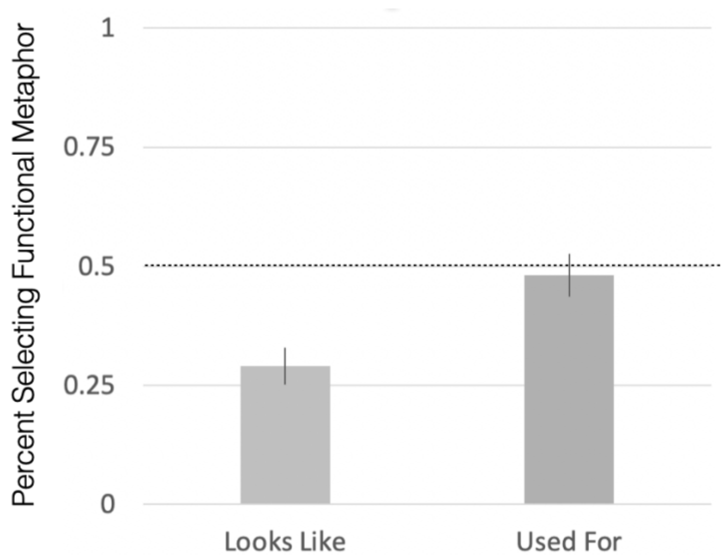

Figure 3.4 Experiment 3 results. Error bars show 1 standard error.

3.5 General Discussion

The present findings suggest that preschool-aged children recognized the usefulness of functional metaphors (e.g., "Eyes are windows") when explanations for the comparisons drawn by those metaphors were made explicit. This suggests that, with the right kind of scaffolding, young children are indeed able to appreciate the types of abstract comparisons that are conducive to learning. These findings also corroborate a growing body of work demonstrating that even preschool-aged children are able to understand and reason about abstract relations (Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017; Pouscoulous & Tomasello, 2020; Walker et al., 2016; Walker & Gopnik, 2017). Critically, the results of the present experiments demonstrate that preschoolers *are able* to appreciate these abstract metaphors when the underlying commonalities are made salient. This is a new and different conclusion than those drawn in many earlier studies, which have interpreted children's preference for perceptual metaphors as evidence that children are *unable* to appreciate, reason with, and learn from the abstract relations expressed in functional metaphors.

Moreover, the present findings demonstrate that providing explanations might help preschoolers fully consider the abstract similarities present in functional metaphors. Experiment 1 demonstrated that, in a contextless baseline condition, adults prefer functional metaphors and preschoolers prefer perceptual metaphors. This result is consistent with previous findings (Gentner & Clement, 1988; Silberstein et al., 1982). Introducing a pedagogical context significantly shifted metaphor preferences in adults, but not preschoolers. Experiment 2, however, showed that preschoolers' metaphor preferences shifted away from perceptual metaphors, and towards functional metaphors, when they were provided with explanations for the ways in which two concepts in a metaphor were similar (e.g., "Suns are oranges *because they're both the same color";* "Suns are candles *because they both light up*"). Thus, Experiment 2 showed that preschoolers' metaphor preferences are not fixed, but rather can shift when the underlying comparisons are made explicit. Experiment 3 showed that preschoolers differentiate between functional and perceptual metaphors when given explicit instructions to convey perceptual or functional information (i.e., "what objects look like" versus "what objects are used for"). These finding suggests that children can not only understand metaphors, but also appreciate the informativeness of different kinds of metaphors. Moreover, providing explanations for how two concepts in a metaphor are alike (Experiment 2) shifted preschoolers' preferences as much as explicit instruction to select the functional metaphor (Experiment 3), suggesting that explanation may be an especially helpful mechanism for scaffolding children's emerging analogical reasoning abilities.

While the current studies show that children *can* shift their metaphor preferences when provided with explanations, further research should investigate *why* explanations cause this shift in children's preferences. Chapter 2 of this dissertation demonstrated that young children are capable of understanding metaphors based on abstract, functional similarities. Thus, preschoolers' preference for perceptual metaphors over functional metaphors cannot be explained by a lack of representational ability (i.e., an inability to represent the abstract similarities between two concepts in a functional metaphor). Rather, providing explanations of how two concepts in a metaphor are alike might change the inductive biases that preschoolers bring to the experimental task (Kroupin & Carey, 2021).

Future research will be required to determine why exactly explanations are helpful. There are at least five possible accounts that might explain why preschoolers prefer perceptual metaphors in the baseline version of a metaphor preference task, but shift their preferences towards functional metaphors when provided with explanations. One possibility is that without explanations, preschoolers fail to *spontaneously* notice how two concepts in a functional metaphor are similar. When presented with two metaphors (e.g., "Clouds are sponges"; "Clouds are ice creams"), preschoolers might immediately notice the surface-level similarities within the perceptual metaphor (e.g., how clouds and ice creams are alike) and not pause to consider whether there are also similarities within the functional metaphor (e.g., how clouds and sponges are also alike). Since the perceptual commonalities are more readily available, children may be more likely to spontaneously identify perceptual commonalities than abstract commonalities.

A second possibility is that preschoolers *do* notice similarities within functional metaphors, but not the correct kinds of similarities. For example, preschoolers might be interpreting functional metaphors in perceptual terms (e.g., thinking that clouds and sponges are alike because both are fluffy, not because both hold water). If this second possibility is true, explanations facilitate preschoolers' metaphor preferences because the explanations highlight the correct *kind* of similarity (i.e., same function) required to interpret the functional metaphors. A third possibility is that providing explanations eases executive function demands on metaphor comprehension and relational reasoning (e.g., Ballestrino et al., 2016). Considering multiple speakers, concepts, and similarities between concepts within a single trial might tax preschoolers' attention and working memory; explicitly stating similarities between concepts might ease these difficulties. A fourth possibility is that preschoolers' shift in preferences is caused by the process of explanation itself. Under this possibility, explanations might change the kinds of features or similarities that preschoolers notice or prefer (Walker et al., 2014; 2017). A fifth possibility is that explanations invite comparison, which in turn may facilitate the ability to notice underlying abstract similarities (Edwards et al., 2019). Thus, while the current work shows that preschoolers can flexibly shift their metaphor preferences, future work should explore the mechanisms underlying this shift.

While explanations facilitated preschoolers' shift in metaphor preferences, the apparent absence of the effect of pedagogical framing may seem surprising. Experiment 1 shows that adults already prefer functional metaphors, and that introducing a pedagogical context significantly boosts this preference. In contrast, preschoolers' preference for functional versus perceptual metaphors in Experiment 1 did not shift when presented with a pedagogical context. This lack of pedagogical effect in Experiment 1 may initially seem surprising given previous work on preschoolers' sensitivity to pedagogical contexts (Bridgers et al., 2019; Gweon et al., 2018; Gweon & Asaba, 2018; Koenig & Harris, 2005; Sobel & Kushnir, 2013). However, it is worth noting that the current experimental task is more subtle, and thus likely more difficult, than previous pedagogical tasks. Specifically, in previous pedagogical paradigms, there is often a single correct answer or desirable outcome: for example, toys either will, or will not, cause a machine to light up (Bridgers et al., 2019). In contrast, when communicating information through metaphors and analogies, there is no obviously *wrong* answer from the infinite possible set of metaphors; however, some answers are better and more informative than others. Indeed, even scientists sometimes struggle to differentiate between various possibly informative metaphors, and are misled through the pursuit of an unfruitful metaphor (Sullivan-Clarke, 2019). Thus, the lack of pedagogical facilitation for preschoolers in Experiment 1 may be due to the fact that the present task is subtler and more difficult than the kinds of pedagogical tasks that preschoolers typically engage in.

The present studies demonstrate that preschoolers flexibly shift their preferences for perceptual versus functional metaphors across experimental contexts, and providing explanations facilitates this shift as much as explicitly instructing preschoolers to select functional metaphors. However, a limitation of the current research is that there were no contexts in which preschoolers consistently selected functional metaphors *over* perceptual metaphors. Rather, providing explanations helped preschoolers shift away from significantly preferring perceptual metaphors, towards preferring perceptual and functional metaphors equally. Thus, while providing explanations significantly improved preschoolers' performance on the metaphor preference task, preschoolers still did not perform as well as adults, who consistently preferred functional metaphors.

The present results align with previous research demonstrating that young children's analogical reasoning abilities are highly sensitive to experimental paradigms: for example, children fail at classic relational match-to-sample paradigms until approximately four or five years of age (Christie & Gentner, 2014; Hochmann et al., 2017), but pass causal reasoning tasks involving analogy in toddlerhood (Walker & Gopnik, 2014; Walker et al., 2016; Walker & Gopnik, 2017). Similarly, while preschoolers clearly differentiate between functional metaphors and nonsense statements in a dichotomous-choice paradigm (see Chapter 2), preschoolers in the present studies had more difficulty differentiating between functional metaphors and perceptual metaphors. Indeed, the format of these experimental tasks may have been especially challenging, since the visual presentation of the metaphors might make the perceptual similarities especially salient (e.g., the pictorial depictions make it especially easy to notice that moons and cookies share the same shape). In contrast, humans typically using only language, rather than language and pictures together, to communicate metaphors. Consequently, future research could investigate whether a purely linguistic, non-visual mode of metaphor presentation might facilitate even greater preferential shifts in preschoolers. Thus, since the current research shows that preschoolers' preferences are to some extent flexible, future work might investigate whether there are additional contexts that might facilitate even greater shifts in preschoolers' metaphor preferences, such that preschoolers consistently prefer functional metaphors.

Moreover, while adult studies of metaphor comprehension will sometimes collect additional information (e.g., ratings of salience, immediacy, and importance) to better understand how adults interpret the metaphoric stimuli (e.g., Gentner & Clement, 1988), developmental studies rarely, if ever, collect additional "norming" information or control variables about metaphors. This discrepancy between adult and child research is likely due to the fact that it is much more difficult to obtain an additional sample of preschooler than adult participants to provide these kinds of ratings. However, future metaphor work would benefit from more prior knowledge of how exactly children understand and interpret the individual items in metaphors, on dimensions such as attractiveness (e.g., perhaps some items, such as donuts and ice cream, are especially appealing to children) or familiarity (e.g., perhaps children more frequently encounter some items than others).

Overall, the current work contributes to a growing body of literature on young children's understanding of non-literal language (Falkum et al., 2017; Pouscoulous & Tomasello, 2020; Zhu, 2021), by suggesting that children may be able to not only understand, but also learn from, metaphors. Specifically, the current research shows that preschoolers are sensitive to the informativeness of functional metaphors, suggesting that they possess a critical initial requirement for understanding metaphors in a manner that is conducive for learning. The ability to select appropriate metaphors is important: in order to successfully learn from a metaphor, a child must not only understand metaphors, but also recognize which metaphors are useful for learning and additional inferential reasoning (Richland & McDonough, 2013). By demonstrating that providing explanations can change children's metaphor preferences, the current studies pave the way for future research on ways to use metaphor as a powerful learning mechanism early in human development.

Chapter 4: Preschoolers and adults make inferences from novel metaphors

4.1 Introduction

In human adults, metaphors can facilitate communication and provide effective frameworks for reasoning about abstract concepts, thus influencing attention, memory, and information processing (Camp, 2009; Thibodeau et al., 2017). Moreover, metaphors are a force for creative change across many disparate domains: for example, metaphors facilitate the development of new insights about old concepts in art and poetry (Camp, 2009; Kulvicki 2020), new word meanings in language (Bowdle & Gentner, 2005; Camp, 2006; Holyoak & Stamenkovíc, 2018), and new discoveries and theories in science (Kuhn, 1993).

While researchers have investigated the influence of metaphors on human adult cognition, less is known about whether and how metaphors might impact thinking and reasoning in young children. Some previous research has suggested that young children have difficulties understanding metaphors (Winner et al., 1980), possibly due to an inability to reason about abstract relations (Silberstein et al., 1982) or a pragmatic inability to understand non-literal language (Winner, 1997). Under this view, children may only understand metaphors in an adult-like fashion quite late in development, possibly not until adolescence (Demorest et al., 1983; Silberstein et al., 1982; Winner, 1997). However, other researchers have argued that metaphor comprehension might actually emerge much earlier in ontogenesis (Pouscoulous & Tomasello, 2020). Indeed, recent work showed that children develop sophisticated relational reasoning abilities in their preschool years (Christie & Gentner, 2014; Goddu, Lombrozo, & Gopnik, 2020; Hochmann et al., 2017) or even earlier (Anderson et al., 2018; Walker et al., 2017; Walker & Gopnik, 2018). Moreover, additional research demonstrated that preschoolers can understand other kinds of non-literal language, such as metonyms (Falkum et al., 2017; Köder & Falkum, 2020; Zhu, 2021). Consistent with these findings that preschoolers can reason about abstract relations and understand non-literal language, more recent work suggested that preschoolers are also able to understand metaphors. For example, Pouscoulous and Tomasello (2020) showed that children as young as three years of age understand metaphors based on perceptual similarities (e.g. "The bottle with the big belly" to refer to a round bottle over a slender bottle). Similarly, Chapter 2 of this dissertation demonstrated that four- and five-year-olds understand abstract metaphors based on functional similarities between concepts (e.g., "clouds are sponges"; "roofs are hats"). Specifically, young children differentiated between functional metaphors (e.g., "roofs are hats") and nonsense statements (e.g., "dogs are scissors"), and a subset of children were even able to explicitly state the functional similarities between concepts in the metaphors (e.g., "roofs and hats both cover you").

While this research suggests that children can understand metaphors, less is known about whether metaphors might facilitate further thinking and reasoning in children, as they do in adults (Thibodeau et al., 2017). Given that metaphors can facilitate the discovery of new information (Kuhn, 1993), one possibility is that children may be able to use metaphors to make novel inferences. Consequently, metaphors may be a powerful learning mechanism, not only in adulthood, but also in early childhood. Though developmental psychologists have extensively studied children's learning mechanisms, there is little research on children's capacity to learn from metaphors.

Thus, the current paper is the first to investigate whether young children can use metaphors to make novel inferences, and thus guide their acquisition of new knowledge. In two experiments, we investigated whether preschoolers can *learn* from metaphors. In Experiment 1, we presented four-year-olds with vignettes about novel artifacts and compared these novel artifacts to natural or social kinds, using both *positive metaphors*, which assert that two disparate concepts are similar (e.g., "Daxes are clouds"), and *negative metaphors*, which use negation to assert that two disparate concepts are dissimilar (e.g., "Daxes are not suns"). Moreover, Experiment 1 also validated this novel paradigm with adult participants. In Experiment 2, we conceptually replicated the preschooler results from Experiment 1, using only positive metaphors.

4.2 Experiment 1

In Experiment 1, we investigated whether preschoolers and adults are capable of learning from metaphors. Specifically, we asked whether preschoolers and adults can make additional inferences about the functions of novel artifacts, after hearing about the novel artifacts in metaphoric utterances (e.g. "Daxes are clouds"). In order to ensure that participants interpreted the utterances as non-literal metaphors comparing two conceptually distinct items (e.g., "Juliet is the sun") rather than literal category statements (e.g., "Juliet is a girl"), we explicitly specified that all the novel items were artifact kinds (i.e., toys) and compared these novel items to natural or social kinds (e.g., animals, occupations). Additionally, in Experiment 1, we presented participants with both positive and negative metaphors about each novel item (e.g., "Daxes are clouds. Daxes are not suns."). This ensured that participants' correct responses were driven by a sensitivity to the contents of the metaphor and the overall sentence structure, and not by simpler, lower-level associative mechanisms (e.g., hearing "cloud" might encourage participants to select the cloud-related response, without attending to the actual metaphor).

4.2.1 Methods

4.2.1.1 Participants

We adhered to a stopping rule of 32 participants in each age group, leading to a total of 32 4-year-old participants (M = 4.59 years, SD = .26 years, range = 4.04 – 4.99 years, 14 females and 18 males) and 32 adult participants (M = 21.19 years, SD = 1.68 years, range = 18.20 – 25.64 years, 26 females and 6 males). Researchers tested three additional children whose data were excluded due to experimenter error. Children were recruited from a local database and adults were recruited from a university campus. All participants were tested online, over Zoom. All experiments reported in this paper were approved by the university's Committee for the Protection of Human Subjects. All adult participants and parents of child participants provided informed consent. The preschooler component of the experiment is preregistered at https://osf.io/uptsf/.

4.2.1.2 Stimuli & Procedure

The experimenter presented participants with stories, which participants viewed using either a computer or tablet. The experimenter introduced the paradigm to the participant by showing them a clipart picture of a girl and saying, "This is my friend Sophie. Sophie makes a lot of toys in her toy factory. She's going to tell you about her toy, and then your job is to guess what Sophie's toy can do! Ready to play?"

On each trial, the experimenter introduced a novel toy, using both a positive and a negative metaphor (e.g., "Sophie says, 'This toy is a dax. Daxes are clouds. Daxes are not suns.'"). As the experimenter presented this information verbally, clipart pictures (e.g., a novel toy, a cloud, and a sun) appeared on the screen. To help participants remember which metaphor was the positive comparison and which metaphor was the negative comparison, the pictures of the two comparison items were accompanied with either a small green checkmark to indicate a positive metaphor (e.g., a checkmark placed beside the cloud served as a reminder that daxes *are* clouds) or a small red "x" symbol to indicate a negative metaphor (e.g., an "x" symbol placed beside the sun served as a reminder that daxes are *not* suns). Then, the experimenter asked about the toy's function (e.g., "What do you think daxes can do?"). A person appeared on the left side of the screen and provided an answer consistent with one of the metaphors (e.g., "This person says, 'I think daxes can let out water'", an inference consistent with the cloud metaphor) Then, another person appeared on the right side of the screen and provided an answer consistent with the other metaphor (e.g., "This person says, 'I think daxes can light up'", an inference consistent with the sun metaphor). The experimenter then asked the participant to choose between the two choices (i.e., "Whose answer do you think is better?"). Once the participant answered by providing a response (e.g. "let out water"), the experimenter began the next trial. No feedback was provided. For an example of a trial, see Figure 4.1. On the final trial, the experimenter also asked participants for an open-ended explanation to justify their response on that trial (e.g., if the participant selected "let out water" on the final trial, the experimenter followed up by asking, "Why do you think daxes let out water?").

Each participant received eight trials. For a complete list of metaphors and corresponding inferences, see Table 4.1. Each trial's structure followed the design described above, in which a participant must infer the function of the novel toy based on the metaphor they heard. The order of the eight trials was randomized. Within participants, we counterbalanced the left-right placement of the correct answer. Across participants, we also counterbalanced which metaphors were positive and negative, such that half of the participants heard that daxes were clouds and not suns, and the other half of the participants heard that daxes were suns and not clouds.

Moreover, across participants, we counterbalanced whether the positive or negative metaphors were mentioned first, such that half of the participants heard the positive metaphor before the negative metaphor (e.g., "Daxes are sun. Daxes are not clouds."), and the other half of the participants heard the negative metaphor before the positive metaphor (e.g., "Daxes are not suns. Daxes are clouds.").

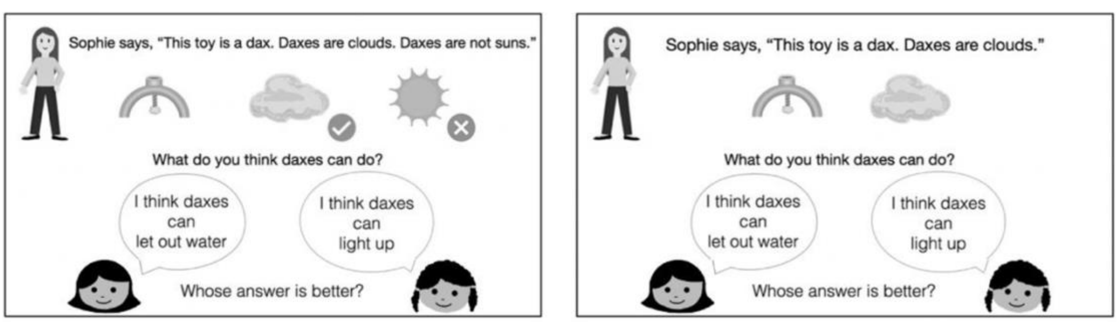

Figure 4.1 Example of a test trial, presented in either Experiment 1 with positive and negative metaphors (left) or Experiment 2 with only positive metaphors (right).

Novel Toy	Metaphor A	Metaphor B	Inference A (Corresponding to Metaphor A)	Inference B (Corresponding to Metaphor B)
Daxes	Daxes are clouds	Daxes are suns	Daxes can let out water	Daxes can light up
Lubbos	Lubbos are snails	Lubbos are bees	Lubbos can move slowly	Lubbos can buzz loudly
Wugs	Wugs are songbirds	Wugs are cheetahs	Wugs can make music	Wugs can move quickly
Feps	Feps are ballerinas	Feps are soldiers	Feps can twirl around	Feps can shoot pebbles
Biboos	Biboos are seagulls	Biboos are kangaroos	Biboos can fly	Biboos can bounce
Blickets	Blickets are eyes	Blickets are teeth	Blickets can help you see things	Blickets can help you chop things
Meelees	Meelees are stars	Meelees are ponds	Meelees can sparkle	Meelees can hold water
Pims	Pims are ducks	Pims are fireflies	Pims can float in the water	Pims can glow in the dark

Table 4.1 Experiment 1 Metaphors and Inferences

4.2.2 Results & Discussion

In the following analyses, the dependent variable was the proportion of correct (i.e., metaphor-consistent) responses. We found that adults overwhelming selected the correct response, *M* = 99.61%, *SE* = .40%, t(31) = 127, $p < .001$. In a preregistered analysis, we found that four-year-olds also selected the correct response significantly above chance levels, *M* = 85.94%, *SE* = 3.40%, t(31) = 10.56, $p < .001$ (see Figure 4.2).

In additional exploratory analyses, we examined participants' average performance on each of the eight test trials. Adults selected the correct response significantly above chance levels across all eight trials ($p < .001$ on all trials). On individual trials, adults' responses ranged from 97% correct (on the ballerina/soldier trial) to 100% correct (on all other trials). Likewise, preschoolers also selected the correct response significantly above chance levels across all eight trials ($p < .002$ on all trials). Preschoolers' responses ranged from 75% correct (on the snail/bee and eye/teeth trials) to 97% correct (on the duck/firefly trial). Adults' and preschoolers' responses on individual trials were still significantly above chance levels after correcting for multiple comparisons (Benjamini & Hochberg, 1995).

In addition to examining individual trials, we also examined the performance of individual participants. Specifically, we demonstrate that a significant proportion of adults and children in the sample perform above chance levels by responding correctly on 100% (8/8) trials (binomial test, $p < .01$). 97% of adults (31 out of 32 participants) responded correctly on all eight trials (a number significantly higher than one would expect by chance, binomial test, $p < .001$). Similarly, 69% of preschoolers (22 out of 32 participants) also responded correctly on all eight trials (again, a number significantly higher than one would expect by chance, binomial test, $p < .001$).

In further exploratory analyses, we examined the explanations that participants provided to justify their responses on the final trial of the experiment. Each participant provided a single explanation on the final trial, leading to a total of 64 explanations (i.e., 32 adult explanations and 32 child explanations). Explanations were coded blind to participants' performance on the test trials. Explanations were sorted into four categories: Explicit Metaphor, Implicit Metaphor, Toy, and Irrelevant. Explicit Metaphor explanations appealed explicitly to the natural/social kind in the positive metaphor (e.g., "Because blickets are eyes and eyes are used to see things"; "Because it's a seagull"). Implicit Metaphor explanations appealed to the features of the natural/social kind involved in the positive metaphor, but did not explicitly name the natural/social kind itself (e.g., "Because they have wings"; "To catch their prey"). Toy explanations appealed to features of the novel toys, rather than mentioning the comparison items (e.g., "They have a little bucket at the end"; "Because they have batteries"). Irrelevant explanations were nonsensical or non-responses (e.g., "Because it sounds like the right answer"; "I don't know"). Two coders coded all explanations. Intercoder reliability was 95%, converging on the same category for 61 out of 64 explanations.

All adults provided explanations that appealed to the metaphor. Specifically, 94% of adults (30 out of 32 adults) provided Explicit Metaphor explanations and 6% of adults (2 out of 32 adults) provided Implicit Metaphor explanations. Similarly, 66% of preschoolers (21 out of 32 preschoolers) also provided explanations that appealed to the metaphor. Specifically, 47% of preschoolers (15 out of 32 preschoolers) provided Explicit Metaphor explanations, 19% of preschoolers (6 out of 32 preschoolers) provided Implicit Metaphor explanations, 6% of preschoolers (2 out of 32 preschoolers) provided Toy explanations, and 28% of preschoolers (9 out of 32 preschoolers) provided Irrelevant explanations. Thus, all adults and the majority of preschoolers appealed to the metaphors in their explanations, either explicitly or implicitly.

In the following analyses, we also examined preschoolers' task performance based on whether they appealed to the metaphor in their explanation (i.e., Explicit Metaphor and Implicit Metaphor explanations) or not (i.e., Toy and Irrelevant explanations). We found that preschoolers who appealed to the relevant metaphors in their explanations (i.e., by providing Explicit Metaphor or Implicit Metaphor explanations) performed significantly above chance levels, $M = 91.67\%$, $SE = 3.27\%$, $t(20) = 12.76$, $p < .001$. Preschoolers who did not appeal to metaphors in their explanations (i.e., by providing Toy or Irrelevant explanations) still performed significantly above chance levels, $M = 75\%$, $SE = 6.74\%$, $t(10) = 3.71$, $p = .004$. Both groups' task performance remained significantly above chance levels after correcting for multiple comparisons (Benjamini & Hochberg, 1995). Crucially, although both groups performed above chance levels, the preschoolers who appealed to metaphors performed significantly better than the preschoolers who did not appeal to metaphors, $t(30) = 2.52$, $p = .02$.

Overall, Experiment 1 showed that both adults and preschoolers can make inferences from novel metaphors. Specifically, after hearing a metaphor about a novel artifact, adults and preschoolers successfully inferred the function of a novel artifact. Moreover, 100% of adults justified their responses by appealing to the novel metaphor, either explicitly or implicitly. Similarly, the majority of preschoolers also justified their responses by appealing to the novel metaphor explicitly or implicitly. Though both preschoolers who appealed to metaphors and preschoolers who did not appeal to the metaphors performed quite well on the task, the former group provided significantly more correct responses than the latter group. Adults' and preschoolers' verbal explanations further demonstrated that participants were using metaphors, rather than lower-level associative strategies, to guide their responses. Overall, the results of Experiment 1 thus provide initial evidence that both adults and young children can not only understand, but also *use* metaphors, specifically for higher-order thinking and reasoning. Consequently, metaphors may be a useful learning mechanism in early childhood.

4.3 Experiment 2

Experiment 1 demonstrated that both preschoolers and adults can make additional inferences from novel metaphors. Moreover, by juxtaposing positive and negative metaphors (e.g., "Daxes are clouds. Daxes are not suns."), Experiment 1 showed that preschoolers and adults made these inferences by carefully attending to the metaphoric utterances, rather than by relying on simple, lower-level associations. However, in more naturalistic contexts, metaphors are not generally contrasted against each other; rather, a single metaphor is often presented alone (e.g., Shakespeare wrote that "Juliet is the sun", not that "Juliet is the sun but not the earth"). Consequently, Experiment 2 seeks to conceptually replicate the developmental findings in Experiment 1, using a more naturalistic paradigm involving only positive metaphors (e.g., "Daxes are clouds").

4.3.1 Methods

4.3.1.1 Participants

Similar to Experiment 1, we adhered to a preregistered stopping rule of 32 participants per age group, leading to a total of 32 4-year-old participants (*M* = 4.43 years; SD = .32 years; range = 4.02 – 4.98 years; 16 females and 16 males) and 32 adult participants (M = 20.99 years, SD = 1.03 years, range = 18.50 – 23.70 years, 19 females, 12 males, 1 non-binary). Researchers tested two additional children, whose data were excluded due to experimenter error (one child) and external interference (one child). All children were recruited from a local participant database and tested online over Zoom. Experiment 2's preregistration can be found at https://osf.io/uptsf/.

4.3.1.2 Stimuli & Procedure

Experiment 2's stimuli procedure was identical to Experiment 1's stimuli and procedure, except participants received only a positive metaphor (e.g., "Daxes are clouds.") rather than both positive and negative metaphors (e.g., "Daxes are clouds. Daxes are not sponges."). Across participants, we counterbalanced which metaphor was presented, such that half the participants heard one positive metaphor (e.g., "Daxes are clouds.") and the other half of participants heard another positive metaphor (e.g., "Daxes are sponges.") Only the images corresponding to the positive metaphors (e.g., a picture of a cloud for "Daxes are clouds") appeared onscreen (see Figure 4.1).

4.3.2 Results & Discussion

We found that adults overwhelming selected the correct response, *M* = 93.75%, *SE* = 2.10%, t(31) = 20.83, $p < .001$. In a preregistered analysis, we found that four-year-olds selected the correct response significantly above chance levels, *M* = 78.13%, *SE* = 3.22%, t(31) = 8.72, $p < .001$.

In additional exploratory analyses, we examined participants' average performance on each of the eight test trials. Adults selected the correct response significantly above chance levels across all eight trials ($p < .001$ on all trials). On individual trials, adults' responses ranged from 88% correct (on the ballerina/soldier trial) to 100% correct (on the songbird/cheetah trial). Preschoolers consistently selected the correct response significantly above chance levels across all eight trials ($p < .03$ on all trials). Preschoolers' responses ranged from 69% correct (on the cloud/sun and songbird/cheetah trials) to 91% correct (on the eye/teeth trial). Adults' and preschoolers' responses on individual trials remained significant after correcting for multiple comparisons (Benjamini & Hochberg, 1995).

In addition to examining individual trials, we also examined the performance of individual participants. A significant proportion of adults and children in the sample perform above chance levels by responding correctly on 100% (8/8) trials (binomial test, $p < .01$). 72% of adults (23 out of 32 participants) responded correctly on all eight trials (a number significantly higher than one would expect by chance, binomial test, $p < .001$). 22% of children (7 out of 32 participants) responded correctly on all eight trials (a number significantly higher than one would expect by chance, binomial test, $p < .001$).

In further exploratory analyses, we examined the explanations that participants provided to justify their responses. Explanations in Experiment 2 were coded using the same four explanation categories from Explanation 1 (i.e., Explicit Metaphor, Implicit Metaphor, Toy, and Irrelevant). Two coders coded all explanations. Intercoder reliability was 95%, converging on the same category for 61 out of 64 explanations.

Similar to the results of Experiment 1, Experiment 2 showed that the majority of participants provided explanations that appealed to the metaphor, either explicitly or implicitly. 88% of adults (28 out of 32 adults) provided explanations that appealed to the metaphor, whereas only 12% of adults (4 out of 32 adults) provided explanations that did not appeal to the metaphor. Specifically, 82% of adults (26 out of 32 adults) provided Explicit Metaphor explanations, 6% of adults (2 out of 32 adults) provided Implicit Metaphor explanations, 9% of adults (3 out of 32 adults) provided Toy explanations, and 3% of adults (1 out of 32 adults) provided Irrelevant explanations. Similarly, the majority of preschoolers appealed to metaphors in their explanations. 44% of preschoolers (14 out of 32 preschoolers) provided Explicit Metaphor explanations, 28% of preschoolers (9 out of 32 preschoolers) provided Implicit Metaphor explanations, 9% of preschoolers (3 out of 32 preschoolers) provided Toy explanations, and 19% of preschoolers (6 out of 32 preschoolers) provided Irrelevant explanations. Thus, the majority of participants in both age groups appealed to the relevant metaphors, when justifying their responses on the task.

Moreover, we also examined participants' task performance based on whether they appealed to the metaphor in their explanation (i.e., Explicit Metaphor and Implicit Metaphor explanations) or not (i.e., Toy and Irrelevant explanations). Adults who appealed to the relevant metaphors (i.e., by providing Explicit Metaphor or Implicit Metaphor explanations) performed significantly above chance levels, $M = 94.20\%$, $SE = 2.27\%$, $t(27) = 19.46$, $p < .001$. Similarly, the few adults who did not appeal to metaphors in their explanations (i.e., by providing Toy or Irrelevant explanations) still performed significantly above chance levels, $M = 90.63\%$, $SE = 5.98\%$, $t(3) = 6.79$, $p = .007$. There was no difference in performance between adults who appealed to metaphors in their explanations and adults who did not, $t(30) = .56$, $p = .58$. Indeed, all adults were quite successful at the task.

Preschoolers who appealed to the relevant metaphors in their explanations (i.e., by providing Explicit Metaphor or Implicit Metaphor explanations) performed significantly above chance levels, $M = 82.61\%$, $SE = 3.22\%$, $t(22) = 10.14$, $p < .001$. Similarly, preschoolers who did not appeal to metaphors in their explanations (i.e., by providing Toy or Irrelevant explanations) still performed significantly above chance levels, $M = 66.67\%$, $SE = 6.91\%$, $t(8) = 2.41$, $p = .04$. Both groups' task performance remained significantly above chance levels after correcting for multiple comparisons (Benjamini & Hochberg, 1995). Similar to the preschooler results of Experiment 1, although both preschooler groups in Experiment 2 performed above chance levels, the preschoolers who appealed to metaphors provided significantly more correct responses than the preschoolers who did not appeal to metaphors, $t(30) = 2.39$, $p = .02$. Overall, Experiment 2 used a slightly modified experimental paradigm to conceptually replicate preschoolers' success from Experiment 1, thus providing further evidence that young children can learn from metaphors.

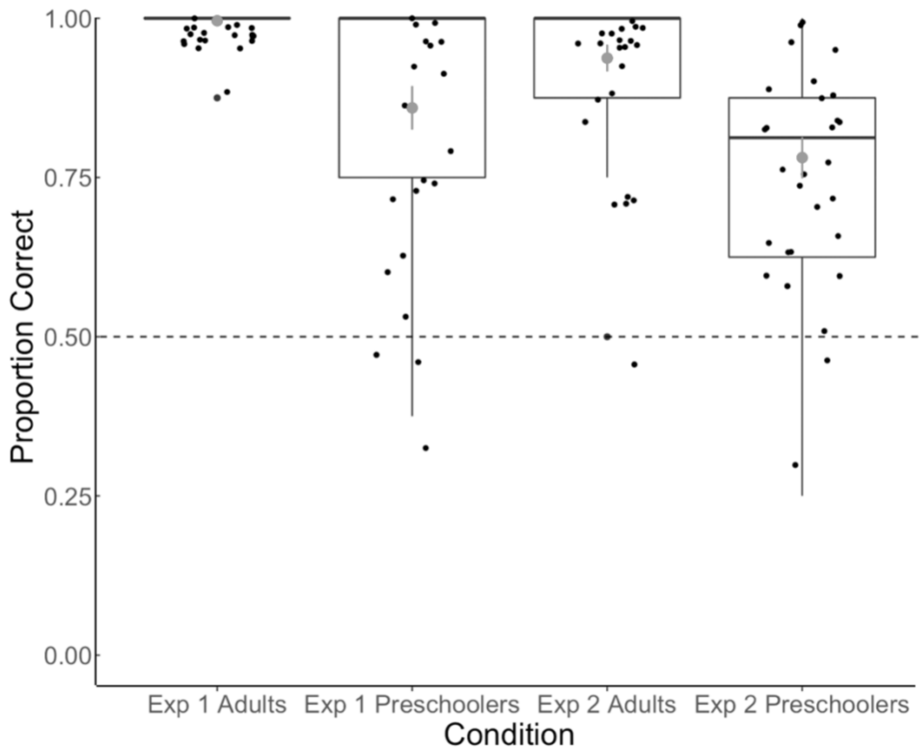

Figure 4.2 Experiment 1 and 2 results.

Overall, Experiment 2 conceptually replicated the results of Experiment 1, by confirming that preschoolers can form inferences from novel metaphors. While Experiment 1 included both positive and negative metaphors (e.g., "Daxes are clouds. Daxes are not suns."), Experiment 2 used a more naturalistic paradigm including only positive metaphors (e.g., "Daxes are clouds."). In Experiment 2's slightly modified paradigm, preschoolers were still able to form the appropriate additional inferences corresponding to the novel metaphors they heard. Moreover, similar to Experiment 1, the majority of preschoolers in Experiment 2 justified their responses by appealing to the novel metaphor explicitly or implicitly. Preschoolers who appealed to metaphors in their explanations also performed significantly better than preschoolers who did not appeal to metaphors in their explanations, though both groups performed above chance levels. Overall, Experiment 2 provides additional evidence that preschoolers can use metaphors to facilitate their thinking, reasoning, and learning.

4.4 General Discussion

This work shows that young children can not only *understand* metaphors, but also *use* metaphors in the service of further thinking and reasoning. Specifically, preschoolers and adults can use metaphors to make additional inferences, and thus learn, about novel concepts. Experiment 1 showed that preschoolers succeed at making inferences on a metaphor task that uses both positive and negative metaphors (e.g. "Daxes are clouds. Daxes are not suns."). The inclusion of both positive and negative metaphors in Experiment 1 suggests that children were indeed using the metaphors, rather than lower-level associative strategies, to guide their responses. Moreover, adults also performed at ceiling on Experiment 1, thus validating this new metaphor inference paradigm. Experiment 2 then built on these initial results by conceptually replicating preschoolers' success in Experiment 1, but using a more naturalistic metaphor task involving only positive metaphors (e.g., "Daxes are clouds"). Overall, preschoolers successfully used metaphors to guide their inferential reasoning and learning, in both a positive-and-negative-metaphor paradigm (Experiment 1) and a positive-metaphor-only paradigm (Experiment 2). Moreover, in both experiments, the majority of preschoolers appealed to the metaphors when providing an explanation for their responses, and the preschoolers who appealed to metaphors performed better than preschoolers who did not appeal to metaphors. These findings suggest that preschoolers can use metaphors to facilitate thinking and reasoning.

The present experiments contribute to a recent, growing body of literature suggesting that young children may possess a relatively sophisticated ability to understand non-literal language (Falkum et al., 2017; Pouscoulous & Tomasello, 2020; Zhu, 2021). Moreover, by demonstrating that preschoolers can make additional inferences from novel metaphors, the present research suggests that metaphors may be a powerful learning mechanism that could allow children to acquire new information. Just as metaphors facilitate novel scientific discoveries in the history of science (Kuhn, 1993) and higher-order cognitive processes in human adults (Thibodeau et al., 2017), metaphors may also contribute to young children's learning. Interestingly, because metaphors frequently provide a new perspective (Camp, 2006; 2009) without necessarily providing new information, metaphors may be a powerful case of "learning by thinking", which allows for the acquisition of new knowledge with little or no additional data (Lombrozo, 2018; Xu, 2019). Overall, the present work contributes to multiple areas of cognitive development research, such as language acquisition and early learning processes, by providing exciting initial evidence of preschoolers' ability to understand and learn from metaphors.

Chapter 5: Conclusions

5.1 Conclusions and implications of the empirical work

Humans possess a remarkable capacity to understand and use symbolic systems. Language is a ubiquitous, human-unique symbolic system that, once acquired, facilitates higher-order thinking and reasoning.

This dissertation investigates children's acquisition and use of a relatively complex kind of non-literal language, namely abstract functional metaphors. Chapter 2 showed that preschoolers as young as four years of age understand metaphors based on abstract, functional similarities (e.g., "Clouds are sponges"). Chapter 3 showed that, while preschoolers tend to prefer relatively less informative metaphors based on perceptual similarities (e.g., "Tires are donuts") to relatively more informative metaphors based on abstract functional similarities (e.g., "Tires are shoes"), providing explanations helps preschoolers shift their preferences towards more informative metaphors. Finally, Chapter 4 showed that preschoolers, like adults, can make additional inferences from novel metaphors. Taken together, these empirical studies provide new evidence that children can both understand and use metaphors from a relatively young age, contrasting previous literature that argued that metaphor acquisition is a slow and laborious process.

5.1.1 Metaphor comprehension

Chapter 2 revisits the topic of children's metaphor comprehension. While previous research suggests that children do not possess a full-fledged understanding of metaphors until quite late in development, possibly not until adolescence (Demorest et al., 1983; Silberstein et al., 1982; Winner et al., 1976; 1980), more recent research shows that the cognitive mechanisms underlying metaphor comprehension – specifically, analogical reasoning – emerges in the first few years of life (Anderson et al., 2018; Carstensen et al., 2019; Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017, Holyoak et al., 1984; Walker et al., 2016; Walker & Gopnik, 2017). Consequently, Chapter 2 uses new and more sensitive experimental paradigms to reinvestigate when children begin to understand metaphors.

Chapter 2 finds that preschoolers as young as four years of age are already capable of understanding metaphors based on abstract, functional similarities (e.g., "Clouds are sponges"; "Chimneys are volcanoes"). In Experiment 1, preschoolers and adults were presented with an absolute judgment task, in which they rated various statements (i.e., functional metaphors and nonsense statements) as "smart" or "silly". Both preschoolers and adults rated functional metaphors (e.g., "Eyes are windows") as significantly "smarter" than nonsense statements (e.g., "Giraffes are snowflakes"). Moreover, within the child sample, approximately a quarter of preschoolers provided functional explanations to justify their responses. The performance of this subset of preschoolers, who provided functional explanations, drove the success of the overall sample. In Experiment 2, preschoolers further demonstrated their understanding of functional metaphors through a dichotomous-choice task. Specifically, when interpreting functional metaphors, preschoolers selected functional explanations over perceptual explanations. Finally, preschoolers in Experiment 3 participated in a final dichotomous-choice task which directly juxtaposed functional metaphors (e.g., "Roofs are hats") against nonsense statements (e.g., "Roofs are scissors"). Preschoolers preferred the functional metaphors over the nonsense statements. Like Experiment 1, Experiment 3 also solicited free explanations from children, and found that a subset of children – indeed, half the sample – provided explanations explicitly articulating the functional similarities between concepts in a metaphor. Once again, the subset of children who provided functional explanations performed well above chance levels, and drove the success of the overall child sample.

In summary, the current findings demonstrate that by at least four years of age, children have an early-emerging capacity to understand metaphors. The current results revise previous assertions that children are slow to develop an adult-like understanding of metaphors (Demorest et al., 1983; Silberstein et al., 1982; Winner et al., 1976; 1980). Moreover, the current results update the literature on metaphor comprehension to align more consistently with findings that preschoolers are capable of analogical reasoning (Anderson et al., 2018; Carstensen et al., 2019; Christie & Gentner, 2014; Goddu et al., 2020; Hochmann et al., 2017, Holyoak et al., 1984; Walker et al., 2016; Walker & Gopnik, 2017), one of the central cognitive mechanisms underlying metaphor comprehension (Bowdle & Gentner, 2005; Holyoak & Stamenkovíc, 2018; Holyoak, 2019; Thibodeau et al., 2017).

5.1.2 Metaphor preference

In order to learn from metaphors, children must not only be able to understand metaphors, but also appreciate their relative informativeness (Richland & McDonough, 2010). Although functional metaphors based on abstract commonalities (e.g. "Eyes are windows") allow for more learning than perceptual metaphors based on superficial commonalities (e.g. "Eyes are buttons"), previous research shows that preschoolers prefer perceptual metaphors over functional metaphors. Might additional context or scaffolding shift children's preferences towards metaphors that are more informative and conducive to learning?

Chapter 3 explores whether providing additional context can shift metaphor preferences in preschoolers and adults. Experiment 1 conceptually replicates previous work showing that, at baseline, adults prefer functional metaphors while preschoolers prefer perceptual metaphors. Moreover, Experiment 1 finds that *pedagogical context* increases preferences for functional metaphors in adults, but not preschoolers. Experiment 2 finds that *providing explanations* for conceptual similarities in a metaphor increases preschoolers' preferences for functional metaphors. Experiment 3 finds that preschoolers differentiate between functional and perceptual metaphors when explicitly asked to communicate functional or perceptual information (i.e., what things are "used for" versus what things "look like").

These findings suggest that, although young children can understand metaphors, they may not always be able to identify the most informative metaphor to learn from. However, providing explanations shifts preschoolers' metaphor preferences, allowing even young children to appreciate the informativeness of functional metaphors.

5.1.3 Learning from metaphors

Previous research demonstrates that human adults use metaphors to guide their everyday thinking, reasoning, and communication (Lakoff & Johnson, 1980; Thibodeau et al., 2017). However, it is unknown whether metaphors may also play a facilitatory role in young children's cognition. Consequently, while Chapter 2 demonstrated that preschoolers *understand* metaphors, Chapter 4 investigates whether preschoolers might *learn* from metaphors.

Experiment 1 demonstrates that both four-year-olds and adults who hear novel positive and negative metaphors can form additional inferences about the functional characteristics of novel concepts, based on the metaphors provided. For example, a preschooler who hears the metaphoric utterance "Daxes are clouds. Daxes are not suns." will generally infer that daxes let out water rather than light up, consistent with the features of clouds. Experiment 2 conceptually replicates this result in both adults and four-year-olds, using a modified paradigm with only positive metaphors (e.g., "Daxes are clouds"). These results are particularly striking because children consistently perform well above chance levels across all trials, demonstrating an early-emerging ability to form additional inferences from novel metaphors.

Taken together, these findings suggest that children can not only understand, but also learn from, metaphors. Consequently, metaphors may be a powerful learning mechanism in both adulthood and early childhood.

5.2 Remaining questions and future directions

This dissertation contributes to the literature on children's linguistic and conceptual development by demonstrating that children as young as four years of age can understand and learn from metaphors based on complex, abstract similarities, such as shared functions. Consequently, these findings pave the way for more exciting future research on the acquisition and comprehension of non-literal language.

For example, one future direction may investigate young children's comprehension of metaphors involving more abstract concepts (e.g., *friendship, idea, justice, love*). Researchers have suggested that metaphors facilitate thinking and reasoning about concepts from a novel perspective (Camp, 2006; 2009; 2015); moreover, metaphors' facilitatory effects may be most powerful for abstract concepts that are relatively ill-defined or complex (Thibodeau et al., 2017). The experiments in the current dissertation presented visual stimuli (i.e., clipart drawings) alongside the verbal metaphors, and consequently used metaphors involving only concrete concepts with clear perceptual properties (i.e., natural, artifact, and social kinds). Future research could extend the current results by investigating how and when young children understand metaphors involving more abstract concepts (e.g., "Love is a journey"; "Ideas are seeds"; "Time is a thief"). It is possible that, using new and more sensitive paradigms (e.g., Chapter 2's "smart" or "silly" absolute judgment paradigm), researchers may find that young children can also understand metaphors involving highly abstract concepts.

Another future direction for language acquisition research is investigating young children's understanding of other kinds of non-literal language and "loose talk". While recent research suggests that preschool-age children already possess a sophisticated understanding of metaphor (Pouscoulous & Tomasello, 2020) and metonymy (Falkum et al., 2017; Köder & Falkum, 2020; Zhu, 2021), less research has been conducted on the developmental trajectories of other kinds of non-literal language, such as hyperbole, understatement, and irony (Demorest et al, 1983). Given the promising recent results showing early comprehension of metaphor and metonymy, it is possible that children also possess sophisticated capacities to understand other kinds of non-literal language early in development.

In addition to future directions in language acquisition, there are still many open questions in cognitive science on the structure of metaphors. For example, one deceptively simple open question is: what makes a "good" metaphor? Previous research has investigated adults' judgments of metaphors' "aptness" (e.g., Bowdle & Gentner, 2005; Gentner & Clement, 1998), but it is unclear what exactly "aptness" is. What are the variables – perhaps informativeness, abstraction, creativity, or conceptual alignment between the two concepts in a metaphor (Gentner, 1988; Wolff & Gentner, 2011) – that underlie adults' ratings of aptness? Moreover, Chapter 3 of the current dissertation shows that both adults and children have clear metaphor preferences, and that adults tend to prefer more abstract and informative metaphors (e.g., "Tires are shoes") whereas children tend to prefer less informative, perceptually-based metaphors (e.g., "Tires are donuts"). Consequently, in addition to open questions about the cognitive mechanisms underlying judgments of metaphor "aptness" or "goodness", future research might explore how these judgments develop or change over ontogenesis.

A final future direction might investigate differences in metaphor comprehension or use across domains. Metaphors are pervasive across multiple highly disparate domains, such as science (Kuhn, 1993) literature (Camp, 2009; 2015), and everyday language (Lakoff & Johnson, 1980), but little is known about how metaphors might vary across these disparate domains, if at all. Indeed, cognitive science research, both in the current dissertation and beyond, tends to examine metaphors in isolation, without much, if any, pragmatic context (Bowdle & Gentner, 2005; Gentner & Clement, 1998). Do different domains use different kinds of metaphors? Are the cognitive benefits of metaphors different across domains? A purely speculative idea is that poets may use metaphors to create evocative moods or atmospheres, whereas scientists may use metaphors to license further inferences and discoveries. Consequently, metaphors found in literature may allow for more vagueness and ambiguity, whereas metaphors found in science might require more precision. Overall, more research exploring metaphors in context, especially across domains, is warranted.

5.3 Concluding remarks

This dissertation demonstrates that preschoolers are already capable of understanding and using abstract metaphors based on shared functional similarities. In particular, both adults and children as young as four years of age differentiate between functional metaphors and nonsense statements, and successfully use functional metaphors to form additional inferences about novel concepts.

The present research contributes to a growing body of empirical evidence demonstrating that young children already possess sophisticated non-literal language comprehension abilities. Moreover, this research also provides a suggestion as to why humans use non-literal language at all. Specifically, metaphors may not be merely a whimsical and ambiguous form of "loose talk"; rather, metaphors may be a powerful mechanism for early childhood learning and discovery.

Appendix: Additional Analyses

We also analyzed the main developmental results from all three experiments using non-parametric tests. We use Wilcoxon rank sum tests to compare performance across two groups, and Wilcoxon signed rank tests to compare group performance to chance levels. All results from the non-parametric tests are consistent with the results from the parametric tests reported in the main body of the dissertation, with the exception of a single result from Chapter 2, Experiment 1 (discussed in detail below).

Chapter 2

In Experiment 1, preschoolers differentiate between metaphors and nonsense statements, $r = 0.14$, $p < .001$. Moreover, preschoolers rate metaphors as "smart" above chance levels, $r = 0.09$, $p = .005$, and nonsense statements as "silly" above chance levels, $r = 0.19$, $p < .001$. Consequently, a parametric t-test shows that preschoolers do not rate metaphors as "smart" above chance levels, whereas a non-parametric Wilcoxon test shows that preschoolers do indeed rate metaphors as "smart" above chance levels. Thus, the results of these non-parametric tests are consistent with – and even strengthen – the initial conclusion that preschoolers can understand metaphors based on abstract functional similarities.

In Experiment 2, preschoolers are significantly more likely to select functional explanations over perceptual explanations when interpreting functional metaphors, $r = 0.40$, $p < .001$.

In Experiment 3, preschoolers demonstrate a significant preference for functional metaphors over nonsense statements, $r = 0.53$, $p < .001$.

Chapter 3

In Experiment 1, preschoolers preferred perceptual metaphors over functional metaphors in both the Baseline condition, $r = 0.38$, $p < .001$, and the Pedagogical condition, $r = 0.31$, $p < .001$. Moreover, there was no difference between the Baseline condition and the Pedagogical condition, $r = 0.03$, $p = .52$.

In Experiment 2, we replicated the finding that preschoolers prefer perceptual metaphors over functional metaphors in the Baseline condition, $r = 0.31$, $p < .001$. In contrast, preschoolers chose between perceptual metaphors and functional metaphors at chance levels in the Explanation condition, $r = 0.06$, $p = 0.39$, and the Pedagogical Explanation condition, $r = 0.02$, $p = .77$. Preschoolers' performance significantly differed between the Baseline condition and the Explanation condition, $r = 0.19$, $p < 0.001$, as well as between the Baseline condition and the Explanation and Pedagogy condition, $r = 0.15$, $p = .004$. However, there was no difference between preschoolers' performance in the Explanation condition and the Pedagogical Explanation condition, $r = 0.04$, $p = 0.42$.

In Experiment 3, preschoolers significantly preferred perceptual metaphors in the Perceptual condition, $r = 0.42$, $p < .001$, but selected at chance between perceptual metaphors and functional metaphors in the Functional condition, $r = 0.04$, $p = 0.56$. There was a significant difference in preschoolers' performance across the two conditions, $r = 0.19$, $p < .001$.

Chapter 4

In Experiment 1, preschoolers were significantly more likely to select the metaphor-consistent response, $r = 0.72$, $p < .001$.

In Experiment 2, preschoolers were also significantly more likely to select the metaphor-consistent response, $r = 0.56$, $p < .001$.

References

Allen, M.L., & Butler, H. (2020). Can drawings facilitate symbolic understanding of figurative language in children? *British Journal of Developmental Psychology, 38*, 345-360.

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.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57,* 289-300.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*, 3253-3258.

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.

Bonawitz, E., Shafto, P. Gweon, H., Goodman, N.D., Spelke, E., & Schulz, L. (2011). *Cognition, 120*, 322-330.

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

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2020). Young children consider the expected utility of others' learning when deciding what to teach. *Nature Human Behavior, 4*, 144-152.

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.

Camp, E. (2015). Metaphors in literature. In N. Carroll & J. Gibson (Eds.), *The Routledge Companion to Philosophy of Literature*. New York: Routledge.

Carey, S.E. (2009). The origin of concepts. New York, NY: Oxford Univesity Press.

Carstensen, A., Zhang, J., Heyman, G.D., Fu, G., Lee, K., & Walker, C. (2019). Context shapes early diversity in abstract thought. *Proceedings of the National Academy of Sciences, 116,* 13891-13896.

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

Csibra, G., & Shamsudheen, R. (2015). Nonverbal generics: Human infants interpret objects as symbols of object kinds. *Annual Review of Psychology, 66*, 689-710.

Davidson, D. (1978). What Metaphors Mean. In *On Metaphor*. (Sacks, S., ed.), pp. 29–45, Chicago: University of Chicago Press.

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.

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? *Psychological Science, 20,* 252-257.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14,* 164-168.

Edwards, B.J., Williams, J.J., Gentner, D., & Lombrozo, T. (2019). Explanation recruits comparison in a category-learning task. *Cognition, 185*, 21-38.
Falkum, I.L., Recasens, M., & Clark, E. (2017). “The moustache sits down first”: on the acquisition of metonymy. *Journal of Child Language, 44*, 87-119.
Flusberg, S.J., Matlock, T., & Thibodeau, P.H. (2017). Metaphors for the war (or race) against climate change. *Environmental Communication, 11*, 769-783.
Fogelin, R. (1988). *Figuratively Speaking.* New Haven: Yale University Press.
Gardner, H., Kircher, M., Winner, E., & Perkins, D. (1975). Children’s metaphoric productions and preferences. *Journal of Child Language, 2*, 125-141.
Genone, J., & Lombrozo, T. (2012). Concept possession, experimental semantics, and hybrid theories of reference. *Philosophical Psychology, 25*, 717-742.
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.
Gentner, D., & Simms, N. (2011). Language and analogy in conceptual change. *Behavioral and Brain Sciences, 34*, 128-129.
Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science, 17*, 684-691.
Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*, 3–18.
Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language, 36*, 50–67.
Goddu, M., Lombrozo, T., & Gopnik, A. (2020). Transformations and transfer: Preschool children understand abstract relations and reason analogically in a causal task. *Child Development.*
Goldwater, M., & Jamrozik, A. (2019). Can a relational mindset boost analogical retrieval? *Cognitive Research: Principles and Implications, 4*, 47.
Gweon, H., & Asaba, M. (2017). Order matters: Children’s evaluation of underinformative teachers depends on context. *Child Development, 89*, 278-292.
Gweon, H., Shafto, P., & Schulz, L. (2018). Development of children’s sensitivity to overinformativeness in learning and teaching. *Developmental Psychology, 54*, 2113-2125.
Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87*, B23-B34.
Hass, R. (1979). Meditation at Lagunitas. In *Praise,* Ecco Press: New York, NY.
Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition, 176*, 255-268.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences, 33*, 61–83.
Hochmann, J.R., Endress, A.D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition, 115*, 444-457.
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. (2019). *The spider's thread: Metaphor in mind, brain, and poetry*. MIT Press: Cambridge, MA.
Holyoak, K. J., & Stamenkovíc, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin, 144,* 641-671.
Holyoak, K.J., Junn, E.N., & Billman, D.O. (1984). Development of analogical problem-solving skill. *Child Development, 55*, 2042-2055.
Özçaliskan, S. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language, 32*, 291–318.
Özçaliskan, S. (2007). Metaphors we move by: Children's developing understanding of metaphorical motion in typologically distinct languages. *Metaphor and Symbol, 22*, 147-168.
Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences, 111,* 12002-12007.
Keil, F.C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development, 1*, 73-96.
Köder, F., & Falkum, I.S. (2020). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics, 156*, 191-205.
Koenig, M., & Harris, P.L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76,* 1261-1277.
Kroupin, I., & Carey, S. (2021). Population differences in performance on Relational Match to Sample (RMTS) sometimes reflect differences in inductive biases alone. *Current Opinions in Behavioral Science, 37*, 75-83.
Kroupin, I.,G., & Carey, S.E. (2022). You cannot find what you are not looking for: Population differences in relational reasoning are sometimes differences in inductive biases alone. *Cognition, 222*, 105007.
Kuhn, T.S. (1993) Metaphor in science. In *Metaphor and Thought* (Ortony, A., ed.), pp. 533–542, Cambridge University Press: Cambridge, UK.
Kulvicki, J. (2020). *Modeling the meanings of pictures.* New York: Oxford University Press.
Lakoff, J., & Johnson, M. (1980). Metaphors we live by. University of Chicago Press: Chicago, IL.
Landau, B., Samuelson, L., Jones, S. S., Smith, L. B., & Gershkoff-Stowe, L. (2003). Object name learning provides on-the-job training for attention. *Psychological Science*, *13,* 13-19.
Li, J., Liu, L., Chalmers, E., & Snedeker, J. (2018). What is in a name? The development of cross-cultural differences in referential intuitions. *Cognition, 171,* 108-111.
Littlemore, J. (2015). *Metonymy: Hidden shortcuts in language, thought, and communication.* Cambridge, UK: Cambridge University Press.
Lombrozo, T. (2018). Learning by thinking in science and in everyday life. In P. Godfrey-Smith and A. Levy (Eds.), *The scientific imagination.* New York, NY: Oxford University Press.
Lourenco, S.F., & Longo, M.R. (2011). Origins and development of generalized magnitude representation. In *Space, Time, and Number in the Brain: Searching for the Foundations of Mathematical Knowledge* (Dehaene, S., & Brannon, E., ed.), pp. 225–244, Elsevier Academic Press: San Diego, CA.
Machery, E., Mallon, R., Nichols, S., & Stich, S. (2011). Semantics, Cross-Cultural Style. *Cognition, 92*, 1-12.

Marchand, E., and Barner, D. (2018). Analogical mapping in numerical development. In Berch, D. B., Geary, D. C., & Koepke, K. M. (Eds.). *Language and Culture in Mathematical Cognition* (Vol. 4). Cambridge, MA: Academic Press.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14,* 57-77.
Nippold, M. A., & Sullivan, M.P. (1987). Verbal and perceptual analogical reasoning and proportional metaphor comprehension in young children. *Journal of Speech and Hearing Research, 30,* 367-376.
Opfer, J.E., & Doumas, L.A. (2008). Analogy and conceptual change in childhood. Behavioral and Brain Sciences, 31, 723.
Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics, 25*, 51-79.
Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics, 156,* 160-167
Ramscar, M, Dye, M., Popick, H.M., & O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE, 6,* e22501.
Reimer, M., & Camp, E. (2006). Metaphor. In E. Lepore & B. Smith (Eds.), *Handbook of Philosophy of Language*. Oxford: Oxford University Press.
Reynolds, R.E., & Ortony, A. (1980). Some issues in the measurement of children's comprehension of metaphorical language. *Child Development, 51*, 1110-1119.
Richland, L.E., & McDonough, I.M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology, 35*, 28-43.
Roberts, R.M., & Kreuz, R.J. (1994). Why do people use figurative language? *Psychological Science, 5*, 159-163.
Shafto, C., Havasi, C., & Snedeker, J. (2014). On the plasticity of semantic generalizations: Children and adults modify their verb lexicalization biases in response to changing input. *Developmental Psychology, 50*, 794-808.
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.
Simms, N.K., & Richland, L.E. (2019). Generating relations elicits a relational mindset in children. *Cognitive Science, 43*, e12795.
Sobel, D.M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review, 120,* 779-797.
Sullivan-Clarke, A. (2019). Misled by metaphor: The problem of ingrained analogy. *Perspectives on Science, 27*, 153-170.
Thibodeau, P.H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE, 6,* e16782.
Thibodeau, P.H., Frantz, C.M., & Berretta, M. (2017). The earth is our home: Systemic metaphors to redefine our relationship with nature. *Climatic Change, 142,* 247-300.
Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How linguistic metaphor scaffolds reasoning. *Trends in Cognitive Science, 21,* 852-863.
Thibodeau, P.H., Matlock, T., & Flusberg, S.J. (2019). The role of metaphor in communication and thought. *Language and Linguistic Compass*, e12327.
Vivaldi, R.A., & Allen, M.L. (2021). Beyond literal depiction: Children's flexible understanding of pictures. *Journal of Experimental Child Psychology, 210*, 105208.

Vosniadou, S. (1987). Contextual and linguistic factors in children's comprehension of non-literal language. *Metaphor and Symbolic Activity, 2,* 1-11.
Vosniadou, S., & Ortony, A. (1983). The emergence of the literal-metaphorical-anomalous distinction in young children. *Child Development, 54*, 154-161.
Vouloumanos, A., & Waxman, S. R. (2014). Listen up! Speech is for thinking during infancy. *Trends in Cognitive Sciences, 18,* 642-646.
Wagner, K., Kimura, K., Cheung, P., & Barner, D. (2015). Why is number word learning hard? Evidence from bilingual learners. *Cognitive Psychology, 83*, 1-21.
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.
Walker, C.M., Lombrozo, T., Legare, C.H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition, 133*, 343-357.
Walker, C.M., Lombrozo, T., Williams, J.J., Rafferty, A.N., & Gopnik, A. (2016). Explaining constrains causal learning in childhood. *Child Development, 88*, 229-246.
Waxman, S., & Gelman, S. (2009). Early word learning entails reference, not merely associations. *Trends in Cognitive Science, 13,* 258-263.
Winner, E., Engel, M., & Gardner, H. (1980). Misunderstanding metaphor: What's the problem? *Journal of Experimental Child Psychology, 30*, 22-32.
Winner, E., Rosenstiel, A.K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology, 12,* 289–297.
Wynn, K. (1992). Children's acquisition of the natural number words and the counting system. *Cognitive Psychology, 24*, 220-251.
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.
Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245-272.
Zhu, R. (2021). Preschoolers' acquisition of producer-product metonymy. *Cognitive Development, 59,* 101075.