Spring 2013

The role of vocabulary knowledge and novelty biases in word learning: Exploring referent selection and retention in 18- to 24- month-old children and associative models

Sarah Christine Kucker
University of Iowa

THE ROLE OF VOCABULARY KNOWLEDGE AND NOVELTY BIASES IN WORD LEARNING: EXPLORING REFERENT SELECTION AND RETENTION IN 18- TO 24-MONTH-OLD CHILDREN AND ASSOCIATIVE MODELS

by

Sarah Christine Kucker

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Associate Professor Larissa K. Samuelson

ABSTRACT

In order to learn a new word, a young child must extricate the correct object from multiple possible items in front of them, make an initial association between the specific word-form and the particular referent, robustly link the new word and referent and integrate the new word into their lexicon. Recent research suggests processes that focus attention on the most novel objects in a complex environment, as well as the child's own developing vocabulary play critical roles in this process. This thesis aims to understand the influence of novelty and prior vocabulary knowledge on referent selection and how the interaction of novelty and knowledge can lead to word learning.

A series of empirical studies first probed the use of children's endogenous novelty bias in a referent selection task, and then explored how the use of novelty was related to retention of newly mapped word-referent pairs. A second set of studies explored children's use of vocabulary knowledge in ambiguous learning situations by varying the strength of knowledge for competing items present during novel word learning. Finally, a Hebbian Normalized Recurrent Network model was used to explore the underlying associative process of referent selection and retention in novelty- or knowledge-based word learning tasks.

Counter to prior work, results here suggest that novelty can override knowledge and in fact, be a detriment to word learning. Children demonstrate a novelty bias across multiple contexts and tasks, but the dominant use of novelty does not translate to retention and does not appear to implicate the use of the child's lexicon. As novelty diminishes and vocabulary knowledge increases, some children can overcome this bias and demonstrate retention for new word-referent pairs. Moreover, the results also suggest that when disambiguation requires the use of weak prior knowledge, more cognitive processing is necessary. The increases in processing subsequently translate to retention for new word-referent pairs. The empirical and computational results together suggest potential limitations of these findings to word learning and suggest future directions exploring variability in object and word representations during learning.

Abstract Approved: ____________________________________

Thesis Supervisor

____

Title and Department

____

Date

THE ROLE OF VOCABULARY KNOWLEDGE AND NOVELTY BIASES IN WORD LEARNING: EXPLORING REFERENT SELECTION AND RETENTION IN 18- TO 24-MONTH-OLD CHILDREN AND ASSOCIATIVE MODELS

by

Sarah Christine Kucker

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Associate Professor Larissa K. Samuelson

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Sarah Christine Kucker

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Psychology at the May 2013 graduation.

Thesis Committee: ______________________________
Larissa K. Samuelson, Thesis Supervisor

____
Bob McMurray

____
John P. Spencer

____
Prahlad Gupta

____
Karla K. McGregor

To my family, who still might not know exactly what I do, but support me unconditionally nonetheless

Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

Lewis Carroll
"Jabberwocky"
Through the Looking-Glass, and What Alice Found There

ACKNOWLEDGMENTS

I first want to thank my advisor, Larissa Samuelson, for everything over the past five years. She has pushed me to do my best, instilling in me a strong sense of scientific rigor and dedication to the study of development. I have great respect for her as a strong female scientist and furthermore as an individual who gracefully balances research, teaching, and family, showing me how to do the same. I can truthfully say that without Larissa's mentorship, I would not have had the ability to be where I am physically, mentally, and academically.

I would also like to thank Bob McMurray as my pseudo advisor. Even prior to being accepted to Iowa's graduate program, he gently steered me toward the lab and mentor best suited for my interests. Bob's flexibility and focus on theoretical questions has helped me mature as a researcher. I am indebted to his instruction and patience in teaching me computational modeling and infiltrating his lab from time to time.

A thank you also goes to the other members of my committee – John Spencer, Prahlad Gupta, and Karla McGregor – for giving great advice not just on the dissertation, but throughout the years for various projects and ideas I have thrown their way.

In addition, I want to express gratitude to my close colleagues and peers at Iowa for keeping each other sane throughout graduate school. Particular thanks go to Lynn Perry, Gavin Jenkins, Amber Dewey, and Megan Galligan for their lab adventures and field trips in addition to their spontaneous scientific and philosophical debates. I want to thank Kim Halvorson for being a great source of camaraderie and kinship and the members of the graduate student journal club for constantly validating why science is fun. I also want to thank both the current and past members of the Language and

Category Development lab for help in recruitment, running, coding, and cheerleading, especially over the last year. Finally, a special thanks to my partner, Mike Chmielewski, for his unconditional support and for being a source of calm over the years. He has also been an essential part of stimulating my scientific thinking and fostering my potential in academia.

ABSTRACT

In order to learn a new word, a young child must extricate the correct object from multiple possible items in front of them, make an initial association between the specific word-form and the particular referent, robustly link the new word and referent and integrate the new word into their lexicon. Recent research suggests processes that focus attention on the most novel objects in a complex environment, as well as the child's own developing vocabulary play critical roles in this process. This thesis aims to understand the influence of novelty and prior vocabulary knowledge on referent selection and how the interaction of novelty and knowledge can lead to word learning.

A series of empirical studies first probed the use of children's endogenous novelty bias in a referent selection task, and then explored how the use of novelty was related to retention of newly mapped word-referent pairs. A second set of studies explored children's use of vocabulary knowledge in ambiguous learning situations by varying the strength of knowledge for competing items present during novel word learning. Finally, a Hebbian Normalized Recurrent Network model was used to explore the underlying associative process of referent selection and retention in novelty- or knowledge-based word learning tasks.

Counter to prior work, results here suggest that novelty can override knowledge and in fact, be a detriment to word learning. Children demonstrate a novelty bias across multiple contexts and tasks, but the dominant use of novelty does not translate to retention and does not appear to implicate the use of the child's lexicon. As novelty diminishes and vocabulary knowledge increases, some children can overcome this bias and demonstrate retention for new word-referent pairs. Moreover, the results also suggest

that when disambiguation requires the use of weak prior knowledge, more cognitive processing is necessary. The increases in processing subsequently translate to retention for new word-referent pairs. The empirical and computational results together suggest potential limitations of these findings to word learning and suggest future directions exploring variability in object and word representations during learning.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Learning a new word is hard, especially for a young child. With immature cognitive systems, relatively little experience in the world, and a comparatively small lexicon relative to adults, young children are faced with the daunting task of determining what single word in a string of continuous speech refers to what single object in a complex scene in front of them. For instance, in order to learn that her new silver coiled spring toy is called a "slinky", 18-month-old Samantha must segment the single word *slinky* from "Oh, honey! You found your brand new slinky! Isn't it fun?" In addition, Samantha must identify and disentangle the silver coiled spring from the visual scene around her, understanding that these features might be representative of a broader category of items. Most importantly, Samantha must make a mapping between both the word-form and the referent such that they can be recalled at a later instance. Despite this complexity, young children like Samantha are surprisingly proficient at learning. Over the next six months, Samantha will encounter many such ambiguous instances of word learning, becoming more and more adept at solving this referential dilemma. By the time she is two-years-old, she will have increased her productive vocabulary from just a handful of words to containing well over 200 words (Fenson et al., 1994).

Word learning, particularly in ambiguous naming situations, has been the focus of much research in the last few decades. Prior work has explored how children segment a new word from the speech stream (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Jusczyk, 2009), how they disambiguate a novel referent from a complex scene in front of them (Bion, Borovsky, & Fernald, 2013; Halberda, 2006), the nature of the mapping between the word and object in-the-moment (Baldwin & Markman, 1989; Kucker, Samuelson, & McMurray, submitted), and how the new word is integrated into the lexicon (Markson & Bloom, 1997). Other work has explored what factors might matter in tackling these challenges, suggesting that aspects of the child's cognitive system as well as the environment may play critical roles. In particular, the size of the child's developing productive vocabulary (Bion et al., 2013; Mervis & Bertrand, 1994) and processes that focus attention on the most novel objects in a complex environment (Horst, Samuelson, Kucker, & McMurray, 2011; Mather & Plunkett, 2012; Mervis & Bertrand, 1994) have been shown to be related to helping a child find a referent initially, increasing their ability to map the word and object together, and possibly boosting long-term retention of the new word. The present thesis focuses on the role of these factors in the developmental trajectory of word learning. I ask how novelty and vocabulary knowledge, two factors known to be important for word learning, interact in the early stages of word learning in children between 18 and 24-months-of-age and how these factors help children get from their first exposure to a new word to robust retention.

Overview

As children encounter new words and objects, they must accurately and robustly associate each of the new words with their respective referents. However, in many cases the exact referent is not clear and children must first solve the problem of referential ambiguity (Quine, 1960). Traditionally, children's ability to complete the first step in this process has been referred to as "fast-mapping" – making a quick initial link between a word and referent. In the initial demonstration of this phenomenon, Carey and Bartlett (1978) asked preschool-age children to "get the *chromium* tray, not the *blue* one, the *chromium*, one" while setting up for snack time. The children correctly chose a novel olive-green colored tray as the referent, suggesting that by 4-years-of-age, they can use prior knowledge of "blue" to determine the referent of the novel word.

In the time since Carey's initial work, numerous studies have explored the phenomenon of fast-mapping and the extent to which young children are able to select and map new word-referent pairs (Baldwin & Markman, 1989; Dollaghan, 1985; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Heibeck & Markman, 1987; Markson & Bloom, 1997; McGregor, Sheng, & Smith, 2005; Oviatt, 1982). This ability has been examined in multiple ages (Halberda, 2003; Mervis & Bertrand, 1994), in the context of multiple novel items (Wilkinson & Mazzitelli, 2003; Wilkinson, Ross, & Diamond, 2003), with atypical populations (Alt, 2011; Capone & McGregor, 2005) and with both looking and reaching measures (Bion, Borovsky, & Fernald, 2013; Gurteen, Horne, & Erjavec, 2011; Halberda, 2003). The research clearly demonstrates that children are very good at the first step in the word learning process – initially selecting a referent in ambiguous situations.

The second step in the word learning process – building on the fast-mapped link – has traditionally been referred to as "slow-mapping". When tested a week later, the majority of children from Carey's original study retained some knowledge of "chromium", having slowly reorganized their lexicons to include the new word. The idea is that over time children strengthen the initial association between the word and referent such that both the phonological form and the physical object are integrated and contrasted with prior knowledge about similar objects, words, and their current vocabulary. The result is that critical elements and features of the target category become apparent and the original word-referent associations are strengthened (Carey, 1978, 2010; Swingley, 2010).

However, we can no longer rigidly adhere to this originally simple view of "fast" and "slow" mapping. Simply selecting an object in an ambiguous referent selection task is not equivalent to making an association between a word and object; young children look longer to a novel referent without showing evidence of encoding the novel word (Bion et al., 2013). Likewise, fast-mapping is not equivalent to long-term retention (Horst & Samuelson, 2008); 24-month-old children do not demonstrate retention of fast-mapped words after a short delay (see also Kucker & Samuelson, 2012). These findings have led to a new perspective on the process by which children make an initial quick mapping and, over time, integrate that mapping into their vocabulary (Kucker, et al., submitted; McMurray, Horst, & Samuelson, 2012). In this view, there is a distinction between referent selection and slow associative learning as it relates to retention. Referent selection is based on the specifics of the stimuli presented in-the-moment combined with the knowledge children bring to the task, whereas learning is the longer process of building associations between words and referents such that a word-referent pair can be retained over time. In this way, known words were at one time, novel words that underwent the slow continuous process of building on relevant associations and representations to become familiar and known. This is the perspective adopted (and supported) in the current thesis. In what follows, I review the literature on the role of novelty and vocabulary knowledge, two prominent contributors to the process of early word learning from ambiguous situations, exploring their influence first in terms of referent selection, and then in terms of retention.

Note that this review focuses narrowly on two factors that are known to matter for referent selection and later retention: novelty and vocabulary knowledge. The literature suggests that there are many other influences on referent selection and word learning. For example, children can use social cues such eye gaze and joint attention to find referents in ambiguous situations (Baldwin, 1991; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Diesendruck & Markson, 2001). Similarly, the principle of contrast suggests that children assume that a speaker is most likely using a novel name to refer to a novel object as the speaker would have otherwise used the commonly-known familiar name or additional details to supersede the child's bias to avoid synonyms (Clark, 1990). Children can also use syntactic cues (Bloom & Markson, 1998; Naigles & Swenson, 2007), or a bias to attend to the whole object over parts (Markman, 1991) to get to the correct referent in an ambiguous environment. Alternatively, new knowledge can be gained through multiple repetitions or exposures (Carey & Bartlett, 1978; Goodman, McDonough, & Brown, 1998; Horst & Samuelson, 2008; Mervis & Bertrand, 1994; Smith & Yu, 2008; Suanda & Namy, 2012; Yu & Smith, 2007). The focus of the current thesis on novelty and prior knowledge is in part because children's use of pragmatic or syntactic cues appear to be secondary to the child's own vocabulary knowledge and novelty detection (for further discussion on this issue, see McMurray, Horst, & Samuelson, 2012). The hope is that a detailed examination of the processes by which novelty and knowledge influence referent selection and retention will set the stage for integrating other cues to meaning into a more complete picture of word learning. I begin by operationally defining novelty and knowledge and then reviewing the literature on the role of each in referent selection and retention. It is important to note, however, that novelty and knowledge cannot be rigidly defined. An individual's perception of an item as novel is dependent upon his or her relative familiarity with the item or similar items and the relative novelty of competing items and the specific context in which the item is encountered. Further, given the variability in individual lexicons, and the complexity of semantic connections, it is impossible to quantify how well known an item is. Instead, novelty and knowledge can be viewed as opposing ends on a familiarity continuum where words or objects begin as completely novel and become more and more known over time and exposure. Thus, the level of novelty and strength of knowledge for a particular word or object is minimally viewed as relative to an individual's current lexicon, encoded representations, and other words or objects present.

Defining novelty and knowledge

Both novelty and knowledge have many different instantiations in empirical work. The current thesis defines novelty and knowledge as relative according to a larger continuum of familiarity. Novelty, for instance, is defined as a lack of familiarity. In developmental research, novelty is often behaviorally identified by either a motor movement toward a new stimulus or an increased total looking time at an unfamiliar object. As an item becomes more and more familiar and the representation in memory is strengthened, time to orient toward that item and total time spent looking toward it diminish systematically (Bornstein & Mash, 2010; Fantz, 1964; Uehara, 2000). When two items are paired, children often orient to the item that is relatively more novel. In these forced-choice tasks, children are typically biased to choose the most novel item present, with the relative percentage of such choices decreasing as familiarity with the stimuli increases (Civan, Teller, & Palmer, 2005; Snyder, Blank, & Marsolek, 2008).

Thus, on a familiarity continuum, novelty is a lack of any prior familiarity and critically, a connection with the lexicon is not required.

Novel items are often considered to be salient as well, meaning that they attract a young child's attention (Kagan, 2009). In the present study, I differentiate novelty and salience as two distinct mechanisms that drive attention. The novelty of an item is defined as the extent to which it differs from prior long-term representations (i.e. the item does not fit with what the child already knows). Perceptual salience is the degree to which an object attracts attention independent of whether there is prior knowledge of the item or category. For instance, a parent shaking an old teddy bear up and down could attract a young child's attention, even though the teddy bear may be well known. On the other hand, an unfamiliar stuffed octopus would also attract the child's attention even if it was not in motion or otherwise highlighted. This difference between novelty and salience is consistent with conclusions from the neural literature that different pathways and neural mechanisms exist for attention to novel items as opposed to orientation to salient stimuli (Ackles, 2008; Tarbi, Sun, Holcomb, & Daffner, 2011). However, it is certainly the case that novel items can be salient and salient items are also novel. In the present proposal, I keep salience consistently low by eliminating any eye-catching features (such as extra motion or lights) and using consistent features and standard toy colors across items. Novel items are differentiated primarily by their level of familiarity as defined as a lack of prior exposure to that item.

As novelty for a particular item decreases, knowledge increases. There is a long debate about the specific form knowledge takes from fixed representations to more dynamic patterns of features. Knowledge here is defined as a consolidated representation of some aspect of the stimulus in conjunction with its word-form. In the case of interest that includes a lexical encoding of the word form, the referent, and a mapping between the two such that an individual can correctly identify the word or object when presented (Bion et al, 2013). Stronger knowledge is represented by a better representation of the word and referent individually compounded with a stronger link between the two, meaning it is also less susceptible to interference or decay (Munro, Baker, McGregor, Docking, & Arciuli, 2012) and can be more easily retained over time.

Recent work has also demonstrated that the higher number of associations new words have to the current lexicon, the better those new words can be learned (Hills, Maouene, Riordan, & Smith, 2010), the more variability during training, the more robust vocabulary acquisition is (Perry, Samuelson, Malloy, & Schiffer, 2010), and the more typical the exemplars, the stronger behavior in a word learning task is (Southgate & Meints, 2000). Taken together, this work suggests that knowledge can be quantified as the strength of representation in memory for that word, including both the breadth of the category and the typicality of the exemplar. Thus, though knowledge might be characterized as increased familiarity without a label in some cases, here I limit the term knowledge to refer to the strength and familiarity of a word-referent association. Specific word knowledge is thus the strength of a given word and overall vocabulary is the total number of words the child can produce.

Behaviorally, Fernald and colleagues (1998) proposed that the knowledge or strength of individual words can be assessed by the speed at which children recognize and orientate appropriately in response to a known word, suggesting that knowledge (i.e. the strength of connections between a word and referent) can change over development.

This increases as exposures to the category rise such that the number of repetitions of the word-referent pair can indicate strength of knowledge in most cases. The field has also adopted standard measures of children's general vocabulary knowledge through parent report (e.g. Fenson et al., 1994). In the end, the raw number of words reported in the MCDI has been shown to be a good indicator of a child's overall vocabulary knowledge and their progress in language development generally (Mayor & Plunkett, 2011). Thus, in the current thesis, strength of specific vocabulary knowledge will be controlled by the amount and variability of exposure to newly-learned items and evidence for weaker consolidation by the number of errors on comprehension. Overall vocabulary knowledge will be assessed with a parent report of productive vocabulary according to the MCDI.

Referent selection

The role of novelty

Since Carey and Bartlett's (1978) original demonstration of children's notable ability to solve the problem of referential ambiguity, much work has examined how children succeed at such tasks. The dominate view proposes that various lexical constraints or principles limit the ambiguity in a given scene, thus facilitating easy and accurate selection and subsequent mapping between a new word and referent. One such word learning constraint, the Novel-Name Nameless Category principle (N3C, Mervis & Bertrand, 1994) suggests that children solve the problem of referential ambiguity by focusing on both the novel word and the unknown item as the most novel of all the items present.

This novelty-driven constraint is grounded in work showing that children's attention to a stimulus can be driven by a relative lack of familiarity without regard to linguistic or vocabulary knowledge. An endogenous novelty bias, or an early orientation toward novel items, has been demonstrated in newborn infants, children and adults. Classic work in this domain demonstrates that children prefer to attend to and orient toward novel stimuli (Ackles, 2008; Civan et al., 2005; Fantz, 1964; Hunter & Ames, 1988; Kagan, 2009; Mundy, 1984). As the stimulus becomes less novel, orientation toward it diminished and when both items in a two-alternative forced-choice paradigm are familiar and known, children look between both equally (Bornstein & Mash, 2010; Hunter & Ames, 1988). When a novel label is introduced, children increase their looking to the novel object *only following* the onset of the label (Halberda, 2003; Mather & Plunkett, 2010a), suggesting that referent selection is influenced by the introduction of novel labels, and is also dependent on children's ability to detect the relative novelty of the objects present.

Merriman and Schuster (1991) systematically explored the influence of novelty on referent selection by asking how altering the level of familiarity or novelty with the target item influences disambiguation. They gave 24-month-old children varying levels of familiarity with the novel item prior to a word learning task, effectively diminishing strong novelty biases in some items and leaving other items completely novel. It was only when the novel item was completely unknown and very novel (not pre-exposed or familiarized) that two and four-year-old children reliably selected the novel referent in response to a novel word. A lack of novelty, as defined by increased familiarity in this case, diminished the selection of new word-object pairs in both two- and four-year-old children.

Perhaps the most convincing evidence of the predominant role of novelty in a referent selection task, however, is a recent study demonstrating that when given multiple novel items and a single novel word, 24-month-old children find and select the most novel item as the referent (Horst, Samuelson, Kucker, & McMurray, 2011). In this case, children were not given any known-word items and thus, could not use prior word knowledge to deduce which items should go with the novel name. Instead, the only differentiating feature present was the lack of familiarity and high novelty of the target item. This suggests that children map novel names to the most novel object, not just the unnamed object. Similar work with 22-month-old children in a looking paradigm demonstrated heightened attention to the most novel item present over repetitions of the stimuli (Mather & Plunkett, 2012), showing that novel word-object associations change over exposures.

Taken together, these studies suggest that children bring their endogenous biases toward novelty to a word-learning task and that this novelty orientation helps children at least in the initial stage of referent selection on the track to learning. Though initial referent selection appears to rely heavily on novelty, the full process of word learning is much more complex and novelty alone might do little to advance that dynamic trajectory (see also Mather, 2013). Because retention in prior work was not tested, it is also unclear how novelty is integrated with the process of learning and whether the accurate selection of a novel referent based on novelty will translate to demonstrating retention at a later point in time. One goal of the current thesis is thus to explore the role of novelty beyond referent selection.

The role of knowledge

Another constraint theorized to aid children's disambiguation of complex naming situations is the principle of mutual exclusivity (Heibeck & Markman, 1987; Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003). This constraint rests on an ability to find a referent based on prior knowledge of names for the other stimuli present, rejecting multiple labels for any given item. Competition between what a child already knows and what they do not yet know is an inherent part of this process. In experiments examining mutual exclusivity, children as young as 15-months-of-age have been shown to reject a second label for a known object and instead treat new labels as the name for a novel object or novel part of an object (Markman & Wachtel, 1988).

Markman and colleagues (2003), attempted to focus on the role of prior knowledge in referent selection by exploring children's performance when target items were absent. In this study, 18- to 20-month-old children saw a known familiar object alongside an opaque bucket. When asked to find a novel object by name (e.g. "Can you find the dax?"), instead of selecting the known item, the majority of children tended to search in the bucket, under the table, or look at the experimenter for a better referent than the present known familiar item. These results suggest that children draw on their existing specific vocabulary knowledge to reject a second label for known items and correctly select the novel item as the referent for the novel word. In other words, having prior word knowledge for objects present when a new label is given leads children to look for other possible referents.

Merriman and Schuster (1991) explored the role of prior knowledge and object representations further by presenting 2- and 4-year-old children either typical or atypical exemplars for the known items. They tested referent selection with either prototypical exemplars (e.g. such as a 4-legged black and white cow) that could be more easily aligned with prior knowledge, or atypical exemplars (e.g. an 8-legged blue cow with purple spots) that would be harder for children to align to their existing knowledge. They found that children of both ages only correctly selected the novel target as the referent of a novel word when the known competitor item was a highly prototypical exemplar. When the known item was atypical, both groups failed to disambiguate the known from the novel, selecting atypical exemplars of known categories and completely novel items equally often as the referent of a novel word. This demonstrates that the ability to draw on specific prior representations and category knowledge for a known lexical entry plays a critical role in the ability to use mutual exclusivity to select a novel referent in response to a novel label.

Taken together, this work suggests that children compare what they already know to what they do not yet know in order to determine which the most likely referent for a novel word is. The competitive process has thought to underlie word learning in various situations, particularly in ambiguous word learning situations (McMurray et al., 2012; Swingley & Aslin, 2007; Yoshida & Hanania, 2011; Zosh, Brinster, & Halberda, 2013) and have parallels with general vocabulary knowledge.

Other work supports the idea that developmental changes in overall vocabulary knowledge subsequently influence referent selection. Total productive vocabulary, or overall word knowledge, has been shown to be an important part of young children's ability to find the meaning of a word in the moment. Fernald and colleagues (1998) systematically demonstrated that the speed and accuracy of selecting a specific referent for a given known item is dependent on general cognitive developmental progress. In this study, 15- to 24-month-old children were presented with two known items and a known label. Time to orient and overall looking time to each item were measured. Unsurprisingly, children generally were faster to orient toward and looked longer to the object that matched the prompt. These variables were also directly correlated with age, suggesting that selection of a correct referent and later consistency in choosing that referent in response to a label depends on developmental changes in cognitive abilities.

When Mervis and Bertrand (1994) examined differences in novel word selection in children between 16- and 20-months of age, they found that only children whose productive vocabulary was approximately 50 nouns or higher were able to successfully disambiguate known items from unknown ones and correctly select the novel referent for a novel word. Children with fewer than 50 nouns were not able to select the correct referent initially, but were tracked longitudinally until they surpassed the 50-noun threshold and subsequently demonstrated referent selection. Other work has used similar looking paradigms to examine novel word selection, finding increases in abilities from 14- to 17-months of age (Halberda, 2003; Markman & Wachtel, 1988; Mather & Plunkett, 2010). Bion et al. (2013) found a direct linear relationship between age and disambiguation abilities and, crucially, a correlation between productive vocabulary and referent selection in a looking paradigm. Work with late talkers has also shown that lexicon size correlates with word-learning and referent selection abilities more than with age (Lederberg & Spencer, 2008; McGregor et al., 2005). Taken together, this work on prior vocabulary knowledge suggests that the more a child knows about a known item, the stronger the bias to find another referent for a novel name. And, the more a child develops their vocabulary, the more efficient they become at referent selection.

Overall, it is clear that novelty and knowledge play critical roles in referent selection. Learning a word, however, is more than simply selecting an item as a referent. Rather, word learning involves forming an association between the word and referent and encoding the pair in memory such that it can be recalled at a later time. It has been suggested that constraints based on novelty and prior knowledge such as N3C and mutual exclusivity do not play a critical role in learning, but only in the selection of a novel target item (Horst & Samuelson, 2008; Wilkinson, 2005, 2007). This is an open question, however. Thus, the next section reviews what is known about the role of novelty and prior knowledge on longer retention of new word-referent pairs.

Learning and retention

In their seminal work, Carey and Bartlett (1978) proposed that the initial speedy "fast-mapping" period is followed by a slower, prolonged period of "slow-mapping" in which the initial link is built on. During this process, the target word is integrated with existing knowledge and the new word-referent pair becomes fully integrated into the lexicon. Evidence of this additional learning process has often been quantified as retention of new word-referent pairs over time and exposures (Bion et al., 2013; Capone & McGregor, 2005; Carey, 1978, 2010; Horst & Samuelson, 2008; Jaswal & Markman, 2001, 2003; Mervis & Bertrand, 1994; Swingley, 2010). Successful retention has been demonstrated in studies with the aid of additional lexical cues (Bloom & Markson, 1998; Carey & Bartlett, 1978; Heibeck & Markman, 1987; Markson & Bloom, 1997), or after multiple repetitions or review of the word-referent link (Goodman et al., 1998; Horst &

Samuelson, 2008; Mervis & Bertrand, 1994; Woodward, Markman, & Fitzsimmons, 1994), but critically Horst and Samuelson (2008) have demonstrated no retention after a single ambiguous referent selection trial without additional cues. The suggestion in all these cases is that word learning results from strengthening of an initial quick mapping by the repeated exposure or rehearsal of the word and referent. This process could be supported by a multitude of factors, but very little work has explored the long-term influence of novelty after referent selection. Rather, the brunt of research has focused on the role of prior vocabulary knowledge on long-term retention.

The role of novelty

Despite the long history of work exploring the role of novelty on referent selection, little work has extended to looking at how item novelty during referent selection influences longer-term retention. There is some evidence that phonological novelty of the word-form may aid learning in 16-month-old infants (Mather & Plunkett, 2010b) whereas other work suggests a *lack* of novelty for the to-be-named object in slightly older 24-month-old children is, in fact, the best factor for retention (Kucker & Samuelson, 2012; Mayor & Plunkett, 2010). Mayor and Plunkett (2010) also support the idea of *less* novelty helping retention. They used a computational model to demonstrate that previous robust knowledge of the category boundaries for a referent improved both the ability to make referential links to labels as well as overall rate of acquisition for novel words. Other work suggests that semantic knowledge or familiarity with target items may help children retain newly learned word-referent pairs (Capone & McGregor, 2005; McGregor, Rohlfing, Bean, & Marschner, 2008), implying that more category familiarity, not just less novelty for the specific object is important for retention.

Critically however, in all these cases, novel words are defined as either having no prior reference in the world (Mather & Plunkett, 2010b) or objects without any prior word association (Kucker & Samuelson, 2012; Mayor & Plunkett, 2010). In all cases, words and items fall closer to novel than familiar and none have prior word-referent links, although it is not clear whether variable levels of novelty may differentially influence retention.

The role of knowledge

The previous work examining the role of prior vocabulary knowledge in retention of new words is more expansive than the work on novelty, but is still limited. Work on specific word knowledge has focused on knowledge for competing items during referent selection, but has not extended to retention. Minimal work has demonstrated that competition of known items can influence retention, but this work has focused exclusively on the number of competitors and not variability in knowledge for the specific word-referent pairs (Horst, Scott, & Pollard, 2010; Zosh et al., 2013). On the other hand, work on overall vocabulary knowledge has explored its role in the process of both referent selection and learning, finding positive correlations between overall vocabulary knowledge and retention.

Bion and colleagues (2013) directly examined correlations between overall vocabulary size and retention after referent selection. They conducted a referent selection task with 18-, 24-, and 30-month-old children, and found a linear increase in performance at each age level. Only the oldest group was able to find the novel referent, associate the words and objects and retain that mapping over time. The 24-month-old children performed above chance on referent selection, but did not demonstrate evidence of retention and the youngest group with the smallest overall vocabulary failed at all stages. Their results also offer another important developmental perspective; over all ages, disambiguation performance was strongly correlated with retention, which they suggest was due to children who are best able to disambiguate having more linguistic and cognitive resources at their disposal to encode new word-referent mappings over time.

Other previous work supports these recent findings by Bion and colleagues (2013). For instance, Mather and Plunkett (2009) gave 19.5 to 22.5-month-old children a series of ambiguous learning trials with both a known and novel object present along with either a known or novel label. Though both ages performed poorly on initial referent selection trials, the older children with higher vocabularies demonstrated increased abilities to accurately map a word and object as trials repeated. Repetition via repeat testing in such a format has been shown to be helpful to boost retention (Munro et al, 2012; Vlach & Sandhofer, 2012). In addition, a model by McMurray and colleagues (2012; see also Horst, McMurray, & Samuelson, 2006) lays out an associative perspective in which long-term learning is almost entirely due to repeated mappings that refine correct word-referent pairs and diminish spurious associations. Importantly, this model suggests that repetition of other irrelevant word-referent pairs can help hone in on mapping the target word to the target referent and eliminate competition foil items. In this way, long-term learning for new words appears to rely directly on overall prior knowledge and knowing what a new word *cannot* refer to.

Taken together, the literature describes a complex, dynamic picture of how children get from their initial exposure to a new word and referent to fully integrating the new word-referent pair into their lexicon. Although this review confirms that endogenous novelty biases and vocabulary knowledge influence word learning at different points, we do not have a clear picture of how they might come together in a single child to explain emerging vocabulary development. Furthermore, we do not know whether changes in novelty biases or vocabulary over development can influence long-term retention of new word-object mappings.

The current project - integrating novelty and knowledge

Following the initial referent selection episode, a child must obtain, minimally, an encoding of the word or referent. If they fail to encode any information about the word or referent, then when re-presented with multiple novel items and a novel word (such as during a test for retention), everything will again be just as novel. Without a prior representation to draw on, the child will re-engage in referent selection, in theory starting from scratch. On the other hand, if a child originally encoded some very minimal information about the word or referent during the initial exposure period, then a re-presentation of the stimuli should enable the child to build on what was encoded previously, be it information about the word-form only, the referent only, or some minimal encoding of the association. This raises an important issue – how do the specifics of the referent selection task or the knowledge and biases the child brings with them influence what children are able to take away from the initial exposure. In other words, how children solve the problem of referential ambiguity and attend to information during the first exposure might influence the information gained and thus, affect progress toward demonstrating retention.

For example, two-year-old Annabelle is presented with a series of referent selection trials that each consist of two known items (e.g. shoe and cup) and one novel item (e.g. slinky) and either a familiar or novel name (e.g. shoe or slinky). When asked to "find the slinky", Annabelle correctly reaches for the novel silver coiled spring instead of the known shoe or cup, thus suggesting that she has solved the problem of referential ambiguity and can disambiguate the novel object from the array. One possibility is that Annabelle's selection was based on the novelty of the slinky with little attention paid to the other items. It is also possible however, that Annabelle selected the slinky by a more competitive process of elimination; she knew the words "shoe" and "cup" referred to the other two items and therefore deduced that the new word must go to the other unnamed item. These possibilities likely have different consequences for what is encoded. In the first case, attention directed primarily to the novel object could result in encoding of the object but little learning about the word form or its negative relation to any of the objects. In the second case, if Annabelle attended to the novel word form and then other objects, recalling the names of the other objects to eliminate them, it would likely result in encoding of the novel word and possibly also encoding the fact that the other items are not linked to the new word form.

In both situations, Annabelle's behavior on referent selection would be identical. In addition, because neither case likely resulted in robust encoding of the novel word form, a representation of the objects, and a robust link between them, Annabelle is unlikely to demonstrate retention of the novel word at a later time in either case. The question for the current project, then, is whether systematic changes in the relative novelty of target stimulus or the strength of the child's prior knowledge of the familiar items, could result in better encoding following referent selection and thereby increase the likelihood of retention. Furthermore, given the rapid pace of vocabulary acquisition during the toddler years, I am interested in how these processes shift over development as vocabulary size increases. Because these questions concern the precise nature of the information children encode about the stimuli and word forms, and the integration of these representations into the child's prior vocabulary over time, I will also examine them in the context of a computational model which can help to elucidate the process underlying word learning across development.

Chapter 2 introduces a Hebbian Normalized Recurrent Network (HRN) model of referent selection and retention that can simulate children's performance on a basic referent selection task at different points over vocabulary acquisition. This model shows how performance on the task arises out of the honing of word-referent associations over vocabulary acquisition and makes predictions about the relation of referent selection to overall productive vocabulary. Chapter 3 then tests a prediction of the model in 18-month-old children, an age suspected to tap the lower limit of what children are able to do in an ambiguous naming task. The results compel the implementation of a novelty bias to the model, which combined with productive vocabulary, provides a nice fit to shifts in children's referent selection behavior over development. Chapters 4 and 5 seek to examine the independent influence of stimulus novelty and prior vocabulary knowledge on referent and retention during word learning in 18- to 24-month-old children. Simulations of the empirical results in the Hebbian model demonstrate how word-referent associations can shift children's behavior during novelty or knowledge-based tasks at different points in development. The general discussion in Chapter 6 integrates the findings of all three empirical studies and all associated simulations to ask how knowledge and novelty might best be manipulated to result in retention of new word-object mappings. I conclude by discussing limitations of the current studies and implications for future work.

CHAPTER TWO

A COMPUTATIONAL APPROACH TO REFERENT SELECTION AND WORD LEARNING

Theoretical background

To explore the underlying cognitive processes of word learning in the present thesis, I use a Hebbian Normalized Recurrent Network (HRN) associative model. The HRN has a simple architecture that includes visual and auditory input that link through a decision/lexical layer. Individual objects and word forms can be presented to the model and in real-time, the model can use dynamic competition to differentiate the objects and choose a referent for a given word. Associations between concurrently active words and objects strengthen over longer time periods as activation cycles from the input layers to the decision layer. Word learning occurs when the same word-object links to the lexical layer repeatedly co-occur. In this way, the model allows both a fine-grained perspective on the process of word learning over just a handful of cycles and a longer-term perspective on vocabulary development as it settles on associations over multiple repetitions. Prior work on word learning has used similar associative networks, specifically relying on Hebbian learning to build associations between words and referents. I first review three recent associative networks that primarily use Hebbian learning to illustrate the use of computational approaches for elucidating processes critical to referent selection and word learning. Then, I present the architecture for the current model and simulate children's performance on the standard referent selection task.

Regier (2005) presents a bidirectional associative exemplar-based model that captures referent selection as an initial step in word learning. In his LEX model, input into a bank of visual nodes and auditory nodes is then distributed to a visual or auditory hidden layer. Weights in the respective hidden layer are Gaussian such that activation of one node causes nearby similar items to become slightly active, and further away dissimilar items to remain inactive. As the model is given various inputs and trained with a variety of word-object pairs, it learns to allocate attention to specific significant feature dimensions such as shape or function based on an error signal from the object input when the respective auditory input is given. This feedback results in a reorganization of the nodes such that nodes with similar features are compressed and more likely to be associated and dissimilar nodes are stretched apart. Hebbian connections link the visual and auditory hidden layers. These strengthen over repetitions and through organization of the hidden exemplar layers. Thus, word learning is defined as the association of specific nodes in the hidden layers that connect a visual item or cluster of items to an auditory item. By using only ostensive naming, Regier does not discriminate between referent selection and retention however, grouping all word learning into a more general process of associating auditory and visual nodes. He does incorporate a broader category of visual items, a factor not addressed in the present thesis, but critical to the more general question of what a new word represents (Landau, Smith, & Jones, 1988; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002).

Mayor and Plunkett (2010) also used a Hebbian model to simulate word learning. They explored the relation between word-object associations, prior categorization experience, and joint attention. In their model, two self-organizing maps individually receive visual and auditory input. As input is given, the layers use cross-modal Hebbian learning to activate similar neighboring nodes within each respective layer, thereby also capturing a well-known effect of taxonomic and categorical responding in word learning (Markman, 1991; Smith et al., 2002). A joint attentional event prompts a link to form between an item in the visual layer (and its corresponding similar neighbors) and a word-form in the auditory layer. This link is then subject to Hebbian learning cycles. Hence, word learning begins with self-organization in each layer prior to making word-object associations. A joint attentional event defines when the initial word-referent mapping takes place. The strength of that association then increases over trials as per the Hebb rule, suggesting a prolonged period of learning. In this way, Mayor and Plunkett differentiate initial selection of a word-referent pair during referent selection from the long-term learning trajectory of strengthening that mapping and thereby capture a basic perspective that word learning can operate on different timescales. However, they do not elaborate on how prior knowledge is acquired nor do they examine how overall vocabulary knowledge and growth over development influence word learning.

Finally, McMurray, Horst, and Samuelson (2012) proposed a Hebbian network model that learns words over multiple presentations in a constraint satisfaction context. The model consists of a bank of auditory nodes and a bank of visual nodes connected via a third intermediate bank of lexical nodes. Here, individual input nodes are activated to correspond to the multiple objects and single word present on an ambiguous word learning trial. Activation spreads from these input layers to the hidden lexical layer and back down to the input layers. Connections between word-referent pairs activated at the same time strengthen whereas connections to inactive nodes are inhibited and weaken. In this model, referent selection is thus defined as the selection of a single visual unit after multiple cycles of activation. The initial mapping of a word and object formed in the referent selection task is defined in this model, then, as the fragile first step of learning on this single word-object pair. In contrast, long-term learning occurs over multiple trials as the initial pair is further strengthened and spurious connections are pruned away. Retention at a particular point is shown only when specific word-referent links have been repeatedly strengthened. This model, built to work with ambiguous learning instances, can capture referent selection and retention with the same underlying structure as well as measuring overall vocabulary acquisition. Furthermore, it can distinguish between well-known or familiar words, and novel words via differences in the number and strength of connections between words, objects, and the lexical layer. This model, however, does not include any concept of perceptual biases, such as novelty-driven behavior without relation to the lexicon and its behavior on any trial is the product of the inputs interacting with the accumulated associations that are instantiated in the weights. Learning a new word or strengthening old word-referent pairs both operate based on the strength of the current associations in the model and relative competition between items or words present at any given point in time. In this way, according to the definitions of knowledge and novelty in the current thesis, the model is grounded more in prior knowledge than in novelty-driven perceptual processes.

Notably, in all these Hebbian learning models, simple associative mechanisms are the basis for learning and "knowledge" is defined as the strength of the association between a word and referent. All three models also emphasize the critical role of preliminary organization or prior knowledge in mapping new words and referents.

Regier's (2005) model includes pre-set organizational tools for modifying input layers based on features of the items it has learned. Mayor and Plunkett's (2010) model requires categorical knowledge before it can map specific words and referents, and McMurray et al.'s (2012) model bootstraps prior word-referent associations to form a new word in the moment and this can in turn shape associations. However, Regier (2005) combines the time at which organization of units takes place with the time at which new learning instances occur such that referent selection must assume a very quick, automatic transition into learning. Mayor and Plunkett's (2010) model begins with exposure to each of the objects and words in order to form self-organizing maps that can later be responsible for driving retention from few mapping instances. Because of the pre-familiarization, the model can then demonstrate learning of categorical relations from a single instance, the so-called 'fast-mapping' phenomenon touted in older work, but in this way, the model also imposes a strong effect of prior non-lexical category knowledge in learning novel words. This makes it difficult to examine the influence of vocabulary knowledge on referent selection and retention of novel word-object mappings. Only in the final model by McMurray and colleagues (2012) is knowledge emergent from the architecture of the model and dynamic competition instead of being partially pre-determined by category maps or feature comparisons. It also dissects the development of word learning into multiple timescales that extend beyond a single moment of word-object association and can capture changes in overall productive vocabulary in parallel with performance on the task. Thus, this HRN model is the focus for the current thesis. It is a simple associative network that differentiates referent selection, mapping, and retention, thereby also highlighting the cascade of word learning over development. The independence of activation values and connection strengths (weights) in the model allow the possibility of instantiating changes in stimulus novelty (activation) separately from changes in vocabulary knowledge (connection strength), as we will see in the subsequent chapters. Thus, the HRN by McMurray and colleagues (2012) offers a good tool for examining both the immediate and long-term effects of stimulus knowledge and vocabulary development on referent selection and word learning.

Architecture of the model

Like previous models, the model used here consists of two input layers (an auditory/word layer and a visual/object layer) connected via a hidden lexical layer. Each input layer consists of 40 localized nodes that each represent a distinct word or object. The lexical/decision layer consists of 500 units; much greater than the minimal necessary for connecting words and referents, but also large enough to allow for better learning (McMurray & Spivey, 2000). The network begins fully interconnected such that every node in the visual layer is connected through the decision later to every other node in the auditory later and vice versa (Figure A1).

Input is presented to the visual and auditory layers as activation of the respective nodes. For instance, in a typical referent selection task, a single label node (such as *shoe*) might be activated along with two to three visual nodes (e.g. shoe, dog, slinky), each set to a value of 1.0. The total activation within a layer is then normalized so that activations sum to 1.0 and allowed to spread to the lexical layer. Activation for a given node *x* in the lexical layer is the sum of all the auditory input (a_z) multiplied by the connection weights (w_{xz} which connects auditory node z to lexical node x) and the similarly weighted visual input ($u_{xy}v_z$) (see Equation 1).

Equation 1

$$\Delta l_x = \left(\sum_{z\in A} w_{xz}\, a_z + \sum_{z\in V} u_{xz} v_z\right)$$

The change in activation of a lexical unit is based on the sum of the weighted activation of the activation of each auditory unit multiplied by its respective connection weight to the lexical layer and the respective weighted activation of each visual unit with its connection weight to the lexical layer. This is multiplied by a value to control the weighting of prior activation with new activation, the temperature (τ_{ff}) (Equation 2).

Equation 2

$$l_x^{(t+1)} = l_x^{(t)} + \tau_{ff} \cdot \Delta l_x^{(t)}$$

The activation of a given lexical unit at a single point in time in the product of the current activation, the feed-forward temperature, and the change in activation. Over multiple cycles, quadratic normalization of activations eventually results in a binary outcome where the more strongly activated units increase and inhibit more weakly active units (Equation 3).

Equation 3

$$l_x^{(t+1)} = \frac{\left(l_x^{(t+1)}\right)^2}{\sum_{z\in L} \left(l_z^{(t+1)}\right)^2}$$

Here, each lexical unit is squared and the resulting values are summed to result in squared normalization of all lexical units. Activation then feeds back to the input layers (Equation

4) which update according to their previous activation and the weighted activation of the lexical units (Equation 5).

Equation 4

$$\Delta a_y = a_y \cdot \sum_{z \in L} l_z w_{yz}$$

Equation 5

$$a_y^{(t+1)} = a_y^{(t)} + \tau_{fb} \cdot \Delta a_y^{(t)}$$

Because the sum of the current lexical to auditory weights is multiplied by the current activation of the auditory weights, the resulting changes in activation only accounts for currently active units. Thus, non-active auditory units do not get activated by top-down processes. The updated activation of the input layers is fed back up to the lexical layer and the cycle continues until the RMS change in the lexical layer between cycles is below a very small threshold. The resulting visual node with the highest activation is then taken to be the model's choice in response to the original auditory prompt.

To capture a referent selection trial in this task, the model is presented with a label as well as an array of possible items (a cup, a dog, and a novel item), indicated by the activation of the input nodes (see Figure A2a). Activation from the visual and auditory layers feed into the lexical layer which passes activation back down to both sets of input layers. To learn, matching connections, such as the word *cup* and the object cup mutually strengthen while those that do not match (such as the presence of a dog without the auditory label *dog*) are inhibited (Figure A2b). If, however, a visual node is not present and neither is its auditory node (e.g. neither the word "*shoe*" nor the visual unit for shoe are active), there is no change in connection weights between the input nodes and the lexical nodes. Thus, if the known word *cup* is presented along with a cup, a dog and a novel item, then the connections between the label *cup* and the object cup strengthen, and connections from the novel item to the word *cup* weaken as do the connections between the dog and *cup,* but the connections between the object shoe and the auditory layer remain unchanged. During referent selection, if the model is given a novel referent selection trial with the word "*cheem*" presented to the network along with a cup, a dog, and a novel item, it will cycle activation from the two input layers up to the lexical layer and back down, when no auditory "match" to either the cup or shoe is found, it will inhibit those activations, diminishing the connections over multiple cycles (see Figure A2c). This leaves only the corresponding novel object node active and the connection between *cheem* and the novel object strengthen (see Equation 6).

Equation 6

$$\begin{aligned}\Delta w_{xy} = {} & a_x l_y (1 - w_{xy}) \\ & - .5 \cdot (1 - a_x) \cdot l_y w_{xy} \\ & - .5 \cdot a_x \left(1 - l_y\right) \cdot w_{xy}\end{aligned}$$

Here w_{xy} refers to the connection weight linking auditory unit *y* to lexical unit *x*. The change in connection weight is determined according to current activation of corresponding auditory and lexical nodes. The first line of Equation 6 indicated that if both an auditory nodes is on and its corresponding lexical node is on, the connection should be strengthened. If the auditory node is on, but not the corresponding lexical node (or vise-versa), connections decay as noted in the second and third lines of Equation 6. Weights are updated at each cycle according to Equation 7.

Equation 7

$$w_{xy}^{(t+1)} = w_{xy}^{(t)} + \eta \cdot \Delta w_{xy}^{(t)}$$

Here, η is the learning rate which mediates the connection weights on each cycle such that learning is very small and incremental.

Training the model

Simulations begin with a vocabulary acquisition phase, much like previous models of word learning (Horst, McMurray, & Samuelson, 2006; McMurray & Spivey, 2000). This captures the fact that 18- to 24-month-old participants in the current empirical studies came to the laboratory with a number of words in their productive vocabularies. Like the simulations reported by McMurray and colleagues (2012), during vocabulary acquisition, the model learns word-object pairs for a subset of all possible input nodes. In order to acquire a vocabulary, the model is presented with multiple ambiguous learning trials. On each vocabulary acquisition trial, a single auditory unit is active along with a variable number of visual items. In the current version, the numbers of items present on each trial, or the referential ambiguity level, is set to around 50% such that the number of items can vary from 1 to 30, but on average, there are 14 other competitor items present along with a single target label and referent. Over multiple cycles, the model builds associations between a specific auditory unit and specific visual unit using the Hebbian associative process in which “correct” word-referent associations are strengthened and “incorrect” spurious connections are weakened. One run for each of the thirty to-be-learned words is termed one epoch, or one full cycle of learning exposure.

After each learning epoch, the model is tested for its acquisition of the known lexical items. To test the model’s vocabulary at this point, it is given a 30 alternative forced-choice (30AFC) production task - all 30 possible known auditory units are activated slightly (an average of .03 for each node), but only a single visual unit is active, forcing the model to pick the name of the presented object. This allows for a rigorous test of what the model "knows" by requiring strong connections from a given visual node to the lexical layer and back to its respective auditory node and the complete inhibition of the activation of any irrelevant nodes and pruning of spurious connections. The vocabulary test is repeated for all 30 of the words the model learns as its starting vocabulary. The percent correct of 30 is recorded but the weights for the production test are not saved in order to allow for re-testing of the model at each point in vocabulary acquisition and capture a cross-sectional approach to development.

Periodically, the model is tested on its referent selection and retention abilities. During the referent selection task, the model is presented with 10, three-alternative forced-choice (3AFC) trials. Like the task with children, on each trial a single auditory unit is presented along with one novel item (selecting from the remaining 10 not used as known items) and two known items. Novel items are those which have not yet been presented to the model and thus, are still fully interconnected. Known items are those which have passed a 3AFC comprehension test in which the model is given three words from vocabulary acquisition and a label learned during vocabulary acquisition. This comprehension test is conducted immediately prior to referent selection and thus, also acts as a "warm-up" period like that used for children. This 3AFC criterion to determine if a word is known is less stringent than the production test, allowing for differences in what a child might comprehend versus what they say. During referent selection, the model receives five novel-name trials and five known-name trials. In both cases, the unit in each layer with the highest resulting activation when there is no further change in lexical layer activation is recorded as the choice. Following referent selection, the model receives a test of retention. Here, three novel items from referent selection are presented along with a novel name that had previously been mapped to one of the referents, just as is the case in empirical studies. Correct performance is again judged according to the visual nodes with the highest activation. Performance is recorded, but any weight changes that accumulate over the testing period are not saved (weights are returned to their pre-test states). As a result, learning can continue without the inflated values from testing – that is, while the model can learn within a set of test trials, it does not carry this learning back to the training trials to influence its vocabulary development at large. The model continues to acquire a vocabulary and is tested on referent selection every 500 epochs until it completes 80,000 total epochs. At this point the vocabulary has usually reached ceiling and referent selection performance is stable from epoch to epoch.

Because the model's lexicon starts at zero and increases as associations are strengthened and critically, as spurious associations are pruned over the course of multiple epochs, it allows an analysis of how referent selection and retention abilities change across vocabulary acquisition. In particular, the vocabulary knowledge of the model can be mapped onto the vocabulary knowledge of the 18- to 24-month-old child participants in the current work. Children typically have two clear points in their vocabulary acquisition; before and after the vocabulary spurt (McMurray, 2007). The model also has a division in its overall productive vocabulary; a plateau early in development that has a relatively low vocabulary and a second plateau at a high vocabulary level later in development. In this way, the model can give a developmental perspective of performance on a referent selection and retention task and predict how changes in current productive vocabulary might relate to children's ability to map and retain new word-referent pairs. It can also allow for detailed analyses of changes in the connection weights for specific word-referent pairs as a result of productive vocabulary. Simulation 1 presents this baseline developmental model of referent selection and retention performance across development and vocabulary levels.

Simulation 1

Prior empirical work suggests that as long as a child has at least 50 nouns in his or her productive vocabulary, the child should be able to succeed at referent selection (Mervis & Bertrand, 1994). Other work implies robust referent selection after 17-months-of age regardless of vocabulary (Halberda 2003; Spiegel & Halberda, 2011), but others have found that success on referent selection changes over development and vocabulary (Bion et al, 2013). Each of these studies, however, was slightly difference in terms of the task and stimuli making it difficult to compare across ages. Thus, I first used the HRN model to simulate children's baseline performance on referent selection and retention across development and vocabulary. This simulation provides insight on the changes in referent selection and retention performance at different points in vocabulary development and makes specific predictions regarding the relation between productive vocabulary knowledge and performance on a referent selection task.

Methods

Parameter values for each simulation are provided in Table A1. Ten models with the architecture described above were initialized to learn 30 vocabulary items. For each model, weights were initialized with random values between 0 and .25. As the models acquired their vocabulary, they were periodically tested on referent selection and retention. Referent selection consisted of ten 3-alternative forced-choice (3AFC) trials in which there were two familiar items and one novel item active. On half the trials, a novel label was also active and on the other half the trials, a known label was active. The retention test consisted of three novel items from referent selection and a corresponding novel label. Results were saved to an external file and any weight change on testing trials was discarded before continuing with training. This allowed the models to be tested repeatedly on referent selection and retention at different points in vocabulary development without worrying that the model could learn or develop from the testing period itself.

Productive vocabulary development was measured with the 30AFC production task described above. The referent selection test requires having a small set of words the model knows. In the current simulations, known items here were those that had been correctly identified by the model on a 3AFC comprehension trial. Thus, these words had been previously presented and mapped during vocabulary acquisition and had undergone some learning to prune away spurious connections. However, these words may not yet have undergone enough learning to pass the rigorous production test. A more rigorous test of comprehension might include multiple repetitions of the 3AFC test or a 10AFC test, but pilot work with the model suggests that these more stringent measures result in ceiling performance on comprehension test and equivalent comprehension and production performance. Thus, a single 3AFC comprehension test offers an optimal level of familiarity with a word but maintains the distinction between comprehension and production performance typically seen in children. Furthermore, this corresponds to the situation with children in that a known word in referent selection is typically one which the parents indicates the child understands, but might not yet produce on their own. Possible limitations of this approach are discussed in Chapter 6.

Results and discussion

Results from Simulation 1 are depicted in Figure A3. Average percent correct for the ten models in referent selection and retention are presented on the left axis and average productive vocabulary is presented along the right axis. As is clear in the figure, the model's performance on both kinds of referent selection trials and on retention trials is almost completely consistent across all points of vocabulary acquisition. Performance on Novel RS trials is at ceiling almost from the very beginning of the simulation, well above chance. Likewise, there is some increase in performance on Known RS trials when total vocabulary is low (roughly 8 of 30 possible words), but even during this period performance is still well above chance. In contrast, the model, like children, fails at retaining newly mapped word-referent pairs, performing at chance on retention trials across all vocabulary levels.

On the one hand, the results of Simulation 1 appear to provide a good fit to prior findings with children. For example, prior work suggests that both 18- and 24-month-old children can find well-known items in a referent selection task when asked (Suanda & Namy, 2013; Halberda, 2003; Horst & Samuelson, 2008), and children as young as 16- to 17-months of age can reliably select a novel referent from an ambiguous scene (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Halberda, 2003; Mervis & Bertrand, 1994). Finally, prior work suggests that children as old as 24-months-of-age do not retain word-object mappings from referent selection tasks (Bion et al., 2013; Horst & Samuelson, 2008). This then adds to a growing body of work in which the HRN model has demonstrated strong fits to empirical data on word recognition and referent selection and provides insight to the mechanisms that support the development and use of the early lexicon (McMurray et al, 2012; McMurray, Zhao, Kucker, & Samuelson, 2013).

On the other hand, however, limitations in the existing literature on referent selection and retention make it hard to give strong claims about the model's fit. In particular, the exact procedure and context used in previous studies of referent selection with 18-month-old children has varied greatly. Many limited the number of referents to two (Bion et al., 2013; Halberda, 2003), and almost all used looking measures with a dependent variable of overall looking time rather than a behavioral choice response (Gurteen et al., 2011; Halberda, 2006; Bion et al., 2013). In addition, the procedures used with 24-month-old children are very different from those used with 18-month-old children; with 24-month-old children, referent selection and retention is more typically measured in a complex and rigorous real-time, multi-object task (Horst & Samuelson, 2008; Kucker & Samuelson, 2012). Finally, the model enables examination of the relation between referent selection, retention and total productive vocabulary. The only prior work to directly examine the relation between total number of words known and referent selection performance is that of Bion and colleagues (2013). They examined 18-, 24-, and 30-month-old children's referent selection abilities for both known and novel items using a 2-item forced-choice looking task. They found increases in performance over age that were directly correlated with children's productive vocabulary at each point in development. However, this task was significantly different than the one presented to the model. First, Bion and colleagues used an easier 2AFC task whereas the model always has three items present on every trial. In addition, Bion and colleagues repeat the novel target item multiple times across trials, increasing the exposure to the item, whereas the model never repeats novel target items. Finally, the empirical paradigm included all known items on known referent selection whereas the model has a novel foil item present during known referent selection.

Thus, the model makes two unverified predictions regarding the development of referent selection and retention performance in children between 18- and 24-months of age. First, the fact that performance of the model on known and novel referent selection trials did not change drastically across a large change in the model's productive vocabulary suggests the possibility that as long as the child knows the specific known items presented on a given trial, their overall vocabulary level may matter little for performance in referent selection tasks. Second, the fact that the model is equally poor at retention when it has very few words in its vocabulary suggests that children younger than 24-months-of-age should also fail to retain word-referent mappings formed in referent selection tasks. I examine these predictions in Experiment 1.

CHAPTER THREE

EXPERIMENT 1: REFERENT SELECTION AND RETENTION IN 18-MONTH-OLD CHILDREN

Simulation 1 makes two predictions regarding the relationship between noun vocabulary development and performance in referent selection and retention between 18- and 24-months of age. First, Simulation 1 predicted that as long as a child knows the specific known items used in the task, overall vocabulary should not matter for referent selection performance. Although a small selection of prior studies have examined overall productive vocabulary (Bion et al., 2013; Mervis & Bertrand, 1994), these studies have used easier tasks, eliminated novel foil objects, or confounded learning by repeating words or objects. Second, Simulation 1 predicted that 18-month-old children, or children with lower overall vocabulary, should also fail at retaining newly mapped word-referent pairs. Only one prior study has analyzed both referent selection and retention in 18-month-old children (Bion et al, 2013), but the procedure included an easier 2AFC looking version of the task and repeated exposure to the critical novel items over trials.

The current experiment will thus fill this gap in the literature by testing younger 18-month-old children in a task like that used with 24-month-old children - a multi-object real-item referent selection task that compares forced-choice performance on known referent selection trials, novel referent selection trials, and retention abilities. Eighteen-month-old children typically have a much lower productive vocabulary than 24-month-old children. As we know, there are significant changes in vocabulary and language processing in the six month span from 18- to 24-months of age. As children develop, their efficiency at processing words increases (Fernald et al., 1998), and their vocabulary doubles (Fenson et al., 1994). It is possible that children are able to disambiguate an ambiguous word learning event without the use of prior vocabulary knowledge and thus, low vocabulary children would perform the same as high vocabulary children just as in Simulation 1. It is also possible that less complex tasks such as the looking measures, two-item tasks, or tasks without novel foils are not subject to vocabulary influences, again leading to identical behavior of children across development. Experiment 1 disentangles these possibilities by increasing the complexity of the task and making the procedure comparable to prior work with 24-month-old children.

Methods

Participants

Thirty-two 18-month-old children (16 females; *M* = 18 months, 24 days; range 17:22-19:25) with a mean productive vocabulary of 77.5 words (range = 1-375, median = 65 words) participated. Data for 8 additional children were not included in the analysis due to fussiness (2), more than three trials without a response (4), parent/sibling interference (1) or the parent neglecting to complete and return the vocabulary questionnaire (1). Children received a small prize for participating and did not participate in any previous condition of the experiment or a similar task.

Stimuli

Up to eight objects (from a pool of 16) known by roughly 66% of all 18-month-old children, according to the MacArthur-Bates Communicative Developmental Inventory Lex2005 database, were used as stimuli. There were also eight unknown, unfamiliar novel items (see Figure A4a). Prior to the start of the experiment, parents were given a list of the known items and asked to check which of the item names their child would be familiar with (e.g. “cat” versus “kitty”). Parents were also shown a picture with all the novel items and asked if their child would be familiar with or have a name for any of the items. Both known and novel items were replaced as needed to ensure children were familiar with the label for each of the known items and completely unfamiliar with each of the novel items. Four out of eight possible novel CVC words that conformed to the phonological rules of English but had no known referent (*cheem, dite, fode, lorp, pabe, stad, roke, yok*) were used as novel names. These words were drawn from a database of words used in previous word learning studies (Horst, NOUN database, 2009).

Procedure and design

During the entire procedure, the child was seated across a table from the experimenter in a booster seat next to his or her parent or on his or her parents’ lap. Parents completed the MacArthur-Bates Communicative Development Inventory (MCDI; Fenson et al, 1994) during the session and were instructed to avoid interacting with their child, offering minimal encouragement only if needed. For a full diagram of the trial procedure, see Figure A5 and Table A2.

Warm-up

The procedure began with three warm-up trials in which three known items were presented in a line on a white tray at equidistance from each other. On each of the three warm-up trials, the child was presented with three known items and asked to find a single known item by name. On each trial, the child could observe, but not reach the objects for three seconds prior to the experimenter prompting the child to get one of the items by name (“Can you get the shoe?”) and sliding the tray within reach of the child. The child was given time to make their choice and was re-prompted up to three times if necessary.

Children were praised or corrected as needed. Across all three trials target location and object were randomized and never repeated.

Referent selection

A series of eight referent selection trials immediately followed warm-up. On all trials, two known items (randomly chosen from the same three used during warm-up) and a single novel item (from the set presented to parents and deemed unknown) were present. On four alternating trials, children were asked to find a known item by name (e.g. "Can you get the *shoe*?"). These are the Known RS trials. On the other four trials, children were asked to find a novel item by name ("Can you get the *roke*?"). These are the Novel RS trials. In all cases, the location of the novel item was randomized and not repeated across trials. No correction or praise was given.

Break

Immediately following referent selection, children took a five-minute break during which they played in the playroom or colored quietly in the experiment room. No stimuli from the experiment were present during this time and children were actively engaged in free play with their parents. This delay period ensured that performance on the ensuing retention test was not simply immediate re-selection of the same items.

Retention

Following the break, children were presented with a single warm-up trial in the same manner as the previous warm-up and with the same three known items. This was meant to briefly re-engage the child with the task. Two retention trials immediately followed warm-up. On each retention trial, two novel objects that had previously been the targets on Novel RS trials were present along with a third novel object that had previously been present as a foil during a Known RS trial. Thus, all three items had previously been seen and were equally salient to the child, but only two had previously been mapped to a novel word. Like before, children were given a brief moment to view the items prior to the experimenter's prompt. Children were asked to retrieve one of the novel items by name. Items and prompts were not repeated across trials.

Preference trials

Following retention, children engaged in two preference trials during which one unnamed foil item from a Known RS trial was present along with one novel item named only during Novel RS and one novel item named during both Novel RS and retention. Items were placed on the tray and children were asked "Can you get one?" before the tray was slid forward for the child to make a choice. These trials assessed the saliency of the items and measured children's preference for items that were more or less familiar.

Coding

Naïve coders blind to the hypothesis and condition coded children's selections off-line. Eleven sessions (61.1%) were coded by a second coder for reliability. Inter-coder agreement was nearly 100% with only a single discrepancy across all of the children's selections. This discrepancy was settled via a discussion with a third coder blind to the hypothesis and condition.

Results and discussion

Overall

Results are reported in Table A3. As in prior work, overall children demonstrated excellent performance on Novel RS trials, selecting the correct novel referent 78.6% of the time, $t(31) = 9.40$, $p<.0001$, two-tailed t-test against chance (33%), d=1.66. This fits with the results of Simulation 1; 18-month-old children tested here selected the novel object in response to a novel name at levels similar to 24-month-old children in prior studies who presumably have higher vocabularies. However, contrary to the predictions from the model, these children performed very poorly on Known RS trials in which they were asked to select a known item from an array of known and novel objects. As a group, children chose the known target at chance levels (29.9%) on Known RS trials, $t(31) = -.55$, $p=.586$, $d=-.097$, *ns*. Instead of choosing the correct known item during these trials, children chose the novel foil item 70.3% of the time, significantly different from chance, $t(31) = 6.72$, $p<.001$, $d=1.189$. Remarkably, the items requested on Known RS trials are the same items parents confirmed familiarity with prior to test and that each child accurately selected during the warm-up trials. During warm-up, however there were only three known items present and no novel items, suggesting that at 18-months-of-age when both known and novel items are present, children's bias to novelty overrides their choices based on lexical knowledge. In addition, there were no differences in correct choices in the first half of the trials compared to the second half (31.25% on the first two Known RS trials to 31.25% average in the last two Known RS trials), $t(62)=0$, $p = 1.0$, *ns*, suggesting that poor performance was not due to fatigue. Not surprisingly, children were also poor on retention (32.75%), $t(28) = .036$, $p=.927$, *ns*, $d=-.0067$[1], consistent with the predictions of the model. Analysis of retention was conducted over only the subset of trials that tested on word-referent pairs the child had correctly selecting during the initial

[1] Three children failed to map the critical items during referent selection, thus the degrees of freedom diminished for this secondary analysis. If, however, analysis is done over all trials regardless of initial mapping, results remain the same and are not significantly different from chance (34.8%), $t(31) = -.225$, $p=.824$, *ns*, $d=.0397$.

referent selection trials. This was done to ensure the retention trials were testing recall only on the items most likely previously encoded and items to which children had previously demonstrated some form of initial association by previously selecting it. This approach to analysis on retention is also consistent with prior referent selection work with similar paradigms (Horst & Samuelson, 2008; Kucker & Samuelson, 2012).

During the preference trials, children chose the unnamed item at levels significantly different from chance (54.7%), $t(31) = 2.58$, $p=.015$, $d=.591$, confirming children's bias to attend to the most novel item present. Similarly, children selected the item that had been seen at a higher frequency throughout at a level marginally below chance on these trials (20.3%), $t(31) = -1.80$, $p=.0818$, $d=-.413$. These were items that had been named during novel referent selection trials and again asked for on a retention trial.

Low versus high vocabulary

To examine the possible influence of prior vocabulary knowledge on referent selection performance, children were split into high or low vocabulary groups according to a median split on total productive vocabulary. Note, however, that 18-month-old children overall still have a relatively lower vocabulary than 24-month-old children - an average of 78 words overall here in Experiment 1 as compared to an average of 300 words in the 24-month-old children in Kucker and Samuelson (2012). Thus, a median split of vocabulary of only the 18-month-old children into low and high groups is not equivalent to the low and high vocabulary groups in Simulation 1 of the model, which spans roughly 18- to 24-months of age. Nevertheless, a median split of the children from

Experiment 1 alone might give insight on the relation between overall vocabulary level and referent selection performance.

Thus, the data were split into a low and high vocabulary groups based on the media total productive vocabulary of 65 words. This revealed a significant increase in novel referent selection in children with a higher productive vocabulary; low vocabulary 70.6% correct, high vocabulary 90.4% correct, between groups, $t(30) = 2.32$, $p=.027$, $d=-.878$. Thus, as a group, children at 18-months-of age are indeed excellent at disambiguating a scene of real objects and selecting the novel object as the referent of a novel word, but children with more words in their productive vocabularies are better able to do this than those with fewer words.

In contrast, there were no differences between vocabulary levels on Known referent selection trials (low vocabulary 30.7% vs. high vocabulary 28.8%; $t(30)=.0884$, $p=.9229$, *ns*, $d=.0591$), suggesting that the bias to attend to novelty seen in the overall group is not just restricted to the lower vocabulary children. There were also no differences in retention across vocabulary levels; low vocabulary 31.25%, high vocabulary 34.6%; $t(27)= .2458$, $p=.8077$, *ns*.

On the preference trials, children chose the unnamed item more often regardless of vocabulary level; low vocabulary 57.89%, high vocabulary 50%; $t(30) = .591$, $p=.559$, *ns*. Likewise, they chose the item that was present during both referent selection and retention and thus, had previously been seen the most, at levels marginally below chance, again irrespective of vocabulary; low vocabulary 15.78%, high vocabulary 26.9%, $t(30) = 1.006$, $p=.3223$, *ns*.

Taken together, these results show that 18-month-old children do, in fact, select the novel referent when given a novel word, consistent with much prior work on disambiguation and referent selection and consistent with Simulation 1. On these novel referent selection trials, overall vocabulary does appear to lead to a slight improvement, but children overall still perform at levels well above chance, just like 24-month-old children on an identical task (Horst & Samuelson, 2008; Kucker & Samuelson, 2012). However, these children perform very poorly on known referent selection trials, choosing the novel item on those trials as well, a finding not predicted by Simulation 1. Furthermore, overall productive vocabulary in the 18-month-old children appears to play no role in known referent selection performance. This suggests that 18-month-old children might be choosing the novel target item because of an endogenous novelty bias and not due to the use of any prior lexical knowledge. These children's attraction to novelty was again confirmed in the preference trials where they also selected the most novel item most frequently. Thus, there appears to be a uniquely high novelty bias in 18-month-old children that is not apparent in work with 24-month-old children (Horst & Samuelson, 2008; Kucker & Samuelson, 2012).

These surprising results are counter the predictions of the model in Simulation 1. The model predicted above chance performance on novel referent selection for all vocabulary levels, such as seen here, but it also predicted high performance on known referent selection for even low vocabulary children, which was not found in the current experiment. This is because the HRN's selections in both novel and familiar referent selection trials are based on its prior word-referent associations (i.e. prior knowledge) alone. In its current form, it does not have any perceptual biases such as the attention to novelty seen in 18-month-old children. Thus, Simulation 2 implements a novelty bias in the model and then tests referent selection and retention performance for both low and high vocabulary levels.

Simulation 2

Simulation 2 used the same HRN model presented in Chapter 2. The architecture of the model was identical to Simulation 1 with the exception of the addition of a novelty bias. Here, an artificial boost was manually given to previously unseen items to give those items a temporary increase in activation when first presented, thereby simulating children's bias to attend to the unseen, and unknown items more than familiar items. Though the model does account for slow increases in familiarity (and thus decreases in novelty), the artificial novelty boost allowed for a greater difference between the activation of known and novel item. Like previously, the model began by acquiring a vocabulary and was periodically tested on its referent selection and retention performance, resulting in measures of performance that span a wide range of vocabulary levels. In this way, the model can be used to explore whether 18-month-old children's poor performance on known referent selection is indeed a result of increased activation and attention to unknown items. The model can also be used to ask if a novelty bias interferes with 24-month-olds abilities on such a task.

Methods

Implementing a novelty bias

Novelty is instantiated in the model via the activation of specific input nodes. Typically, at the start of a trial, all nodes that represent the presented items are activated equally. That is, the sum of the activations over the entire layer is normalized to 1.0 such that the inputs for the active units are equal, but higher than non-active units. Thus, in order to instantiate differences in the relative novelty of the presented items, novel items are given a novelty "boost" prior to normalization. In this way, there can be differential initial activation of individual nodes in a given layer. This activation works primarily in-the-moment with the resulting relative activation being first a response to the input itself, and second, dependent on the prior history or familiarity with that item/node. All items in the model get the same novelty boost above baseline the first time they are presented. The size of this boost is diminished over subsequent trials as a function of the number of times the specific item is present over the course of language acquisition (where *n* is the current novelty bias and *k* is the rate of novelty bias decay, see Equation 8).

Equation 8

$$n_x^{(t+1)} = n_y^{(t)} - \delta \cdot v_x^{(t)} \cdot (n_x^{(t)} - 1)$$

Activation of the visual units is multiplied by the novelty bias value only at the beginning of a trial, similar to the pre-shape in dynamic neural field models or the bias weight in connectionist models. The novelty bias value is equal to the just-previous value of that node multiplied by the decay rate as a function of the number of competitors present on the current trial (see Figure A6). Thus, the more novel an item is on a given trial, the higher the activation for that node and known items quickly fall over the course of multiple cycles back to an activation of 1.0 but novel items maintain an above 1.0 activation.

Referent selection and retention

The referent selection and retention test followed the same methods as those used in Simulation 1. The model was given 10, 3AFC referent selection trials each with two known items and a single novel item. Half the trials were known referent selection trials and half were novel referent selection trials. Novel items were those without any prior word-referent links or familiarity. Known items were those that had passed a 3AFC comprehension test. Retention consisted of three novel items from referent selection. Every 500 epochs, the model was tested on both referent selection and retention performance. Choices were noted, but weights were not saved in order for the model to continue to acquire a vocabulary and repeat testing later in development.

Results and discussion

Results from the simulation are depicted in Figure 7b. Overall the model captures the general trends of children in Experiment 1 and clarifies the results of Simulation 1. When the overall productive vocabulary of the model was lower (i.e. when the model "produces" roughly 30% of its total vocabulary, approximately epochs 10,000-27,000, see Figure A8a), performance on known referent selection was weak, hovering near chance, just as it was with children in Experiment 1. At the same time, the model still succeeded at novel referent selection and failed at retention. As the model's vocabulary grew (and increased to at least 90% of its total lexicon, approximately epochs 56,000-80,000), its ability to succeed at known referent selection also increased with no detriment to high novel referent selection performance (see Figure A8b). At this point in vocabulary development, the model remained at chance for retention, much like the prior work on 24-month-old children (Horst & Samuelson, 2008; Kucker & Samuelson, 2012).

In addition to analyzing the model's choices during referent selection, I can measure changes in connection weights in the model after the referent selection task. To do this, I saved the weights of the model, then gave the model the referent selection task and immediately following referent selection, computed the RMS difference between the initial weights and the weights following referent selection. RMS is sensitive to both positive and negative changes in weights, thus giving an idea of how much learning occurs in the course of a single trial. Figure A9 presents the changes in connection weights of the model without and with a novelty bias. In both panels it is clear that there is very little change in the connection weights following Known RS trials for either the original model (panel a) or the model with the novelty bias (panel b). This makes sense as the model has already passed a 3AFC comprehension test for the known words and been exposed to them a number of times and thus, the amount of learning or change after a single trial would be low. A comparison of changes in the connection weights for novel words after referent selection for Simulations 1 and 2 suggests there is more learning during referent selection when a novelty bias has been implemented. Though there are slight differences in the amount of learning in a single trial dependent on vocabulary level, the model still does not demonstrate retention. This suggests that the increase in weight change (i.e. learning during referent selection) when there is a novelty bias driving behavior might still not be enough to withstand a rigorous test for retention.

The model thus demonstrates that by implementing a novelty bias into a purely associative framework, I can account for 18-month-old children's poor performance on seemingly well-known words without altering high novel referent selection or changing retention rates. When vocabulary or prior word knowledge is not sufficient to support new learning, a novelty bias can instead drive selection, leading to high novel referent selection, but poor known referent selection. In other words, novelty can be so powerful that it can override the use of prior knowledge in new word learning scenarios. I next asked if novelty acts in the same manner when name-known items are removed and only novel items are present and how much can actually be learned from such novelty-driven word learning situations. These are the questions addressed in Experiment 2.

CHAPTER FOUR

EXPERIMENT 2: THE ROLE OF NOVELTY IN REFERENT SELECTION AND RETENTION

Experiment 1 revealed a strong influence of novelty during referent selection in 18-month-old children. The results suggested that these young children may not be utilizing their prior vocabulary at all, but indiscriminately attending to the most novel item present. Furthermore, the architecture of the general model relies exclusively on prior knowledge to learn word-referent pairs, but using this basic architecture, Simulation 1 could not capture 18-month-olds behavior on referent selection. It is only when a novelty boost was implemented in Simulation 2 that the model could accurately account for children's data. Furthermore, after the novelty bias was implemented in the model, there was an increase in the amount of change in connection weights during a single referent selection trial, suggesting more learning of novel word-referent pairs occurs in cases of novelty than without such a bias. However, neither children nor the model retain novel word-referent pairs encountered in a single referent selection trial. This suggests that children's novelty bias is necessary for referent selection, but that this does not lead to retention of that new word-referent link.

One reason for the failure to retain could be that when a child focuses exclusively on the novel object, they fail to encode the novel word-form or its relation to any of the objects. In other words, when a novel item is present, children do not even consider the other objects present or the labels for those objects. Not only are they not encoding the new word then, but they are also not adding the new word-referent pair to their lexicon. As such, novelty may not support long-term learning. This would predict that even without the competition of known items on a referent selection trial, 18-month-old children would behave similarly.

Though the results of Experiment 1 and Simulation 2 confirm children's use of novelty in referent selection, the task was not designed to tap into just novelty and instead, was designed to give children the ability to use the presence of known items to their advantage during disambiguation and mapping. Prior work has begun to more closely examine the role of novelty in word learning. Horst et al. (2011) examined 24-month-old children's use of relative novelty to map novel words to novel object during referent selection. Instead of an array of well-known objects paired with novel items such as in the typical referent selection task, they presented children with all novel items that differed only in the amount of visual familiarity. Horst and colleagues confirmed what they termed the "supernovel" effect in which children reject pre-familiarized novel items and choose the *most* novel item present as the referent when given a novel name. Similarly, Mather and Plunkett (2011) used a nearly identical looking task in 22-month-old children, finding that as children were repeatedly presented with a novel label and pairs of items that differed in familiarity, they eventually begin to look longer to the most novel item. Though 24-month-old children certainly use novelty during initial referent selection, none of these prior studies examined retention. Further, none explored children as young as 18-months who demonstrate a robust novelty bias that is not apparent in 24-month-old children. Whether or not 18-month-old children in Experiment 1 utilized the known items to disambiguate or encode new information, or if they would behave the same without known referents present is thus impossible to determine thus far.

Experiment 2 thus differentiates the role of novelty from prior knowledge and expands prior work to include 18-month-old children and test retention of words learned in an exclusively novelty-driven context. Here, I modeled the task after Horst et al. (2011) in which referent selection was composed of novel objects that varied only on the degree of familiarity the child had with each. Children between 18- and 24-months of age were given both novel referent selection trials in which all novel items differed in familiarity (called 'supernovel' trials in Horst et al., 2011) and known referent trials with all known items, followed by a test for retention of the new novel words. Importantly, the larger age range here allowed for comparisons to productive vocabulary. Another critical part of Experiment 2 is the addition of all known referent selection trials that did not include any novel foils. Embedding these all-known trials in a set of novel referent selection trials, allowed me to directly compare performance on known referent selection with novel competitors (such as in Experiment 1) versus known referent selection without a novel foil. Such a comparison tests the hypothesis that children have trouble ignoring the novel objet when asked for a known item, but when no novel distractors are present, children clearly demonstrate their current vocabulary knowledge. I also go beyond Horst et al. (2011) by including a test of retention. Finally, I simulate the results of Experiment 2 in the same HRN model, exploring fine-grained changes in performance and connection weights during referent selection and retention over the course of development.

Methods

Participants

Forty-three, 18-to 24-month-old children (19 females, *M* = 22 months, 15 days; range 17:22-25:20) with a mean vocabulary of 242.6 words (range = 1-657, median = 243) participated. This group could be further subdivided into two groups of children based on age and vocabulary level. In particular, the younger group of eighteen children (9 females) were between 17 ½ -months and 21 ½ -months of age with a mean age of 19 months, 18 days. This group had an average productive vocabulary of 95.2 words with a range of 1-304 words. Though this group of 18-month-old children were, on average, a few weeks older than the group of 18-month-old children in Experiment 1, $t(48)=4.073$, $p=.0002$, $d=-1.092$, the productive vocabulary of 18-month-old children in each experiment were comparable, $t(48)=.6912$, $p=.4928$, $d=-.1936$, *ns*, and consistent with the vocabulary typical of 18-month-old children (Fenson et al., 1994). The older group of 25 children (10 females) were between 21 ½ and 25-months of age with a mean age of 24 months, 17 days. This group had an average productive vocabulary of 348.7 words with a range of 35-657 words. This group of 24-month-old children was comparable in age and vocabulary to that of Horst and Samuelson (2008) and Kucker and Samuelson (2012) and had vocabulary typical for 24-months-old children (Fenson et al, 1994). Data for two additional children were not included due to fussiness (1) and experimenter error (1). Children received a small prize for participating and did not participate in Experiment 1.

Stimuli

A subset of the total possible known objects from Experiment 1 was used. To this set, an additional eight unknown, unfamiliar novel items were added (see Figure A4b).

Parents confirmed the status of each object as either known or novel prior to test. Items were replaced as needed. To the set of eight novel CVC words from Experiment 1, seven others were added (*koob, foo, belp, mel, zeb, blick, wif*).

Procedure and design

During the entire procedure, children were seated across a table from the experimenter in a booster seat next to their parents or on their parents' lap. Parents completed the MacArthur-Bates Communicative Development Inventory (MCDI; Fenson et al, 1994) during the session and were instructed to avoid interacting with their child, offering minimal encouragement only if needed. See Figure A10 and Table A2 for outlines of the procedure.

Familiarization

The procedure began with a short familiarization period in which children were presented with four novel items and were allowed to explore and manipulate the set for approximately one to two minutes. If an item was not explored by the child, the experimenter would point to it and say "look". If a child did not engage with the stimuli at all, the experimenter drew the child's attention to each item equally via pointing or holding the item. No names were given during this time. After the child had been familiarized with each of the four items, they were removed from the table and the child was given a second set of four items to explore in the same manner. After familiarization was complete, all items were removed from the table and the experiment proceeded directly into warm-up. The eight items familiarized during this period were later used as foils during novel referent selection trials.

Warm-up

Three warm-up trials followed familiarization. On these trials, three known items were presented on a white tray at equidistance from each other and asked the child was asked to find a single known item by name. Known items were randomly selected from a subset of six (selected from the array of 16 possible, Figure A4), that parents had previously confirmed to be known by the child. On each trial, the child could look at, but not reach the objects for three seconds prior to the experimenter prompting the child to get one of the items by name ("Can you get the *shoe*?") and sliding the tray within reach of the child to make their choice. The child was re-prompted up to three times if necessary. Children were praised or corrected as needed. Across all three trials target location and object were randomized and never repeated.

Referent selection

A series of 16 referent selection trials immediately followed warm-up. On eight alternating trials, children were asked to find a known item by name (e.g. "Can you get the *shoe*?"). These are the Known RS trials. On Known RS trials, three randomly chosen known items from the subset of six possible items were present along with a known name. Thus, known items and targets could repeat over the course of referent selection trials with the exception that the target was never the same two known trials in a row and the same three items were never presented back-to-back. On the other eight trials, children were asked to find a novel item by name ("Can you get the *roke*?"). There are the Novel RS trials. On Novel RS trials, three unnamed novel items were present. Here, one of the novel items was completely unknown and never-before-seen (the supernovel item) and the other two novel items were objects from pre-familiarization that also did not have names, but had been previously seen. Thus, items only differed on the amount of familiarity or relative novelty and not lexical knowledge. These unnamed foil items from familiarization showed up twice as foils – once in the first block eight referent selection trials and a second time when they were repeated on the second block of the referent selection trials. Supernovel items, however, never repeated over trials. In all cases, the location of the target item was randomized and not repeated across trials. No correction or praise was given.

Break

Immediately following referent selection, children took a five-minute break during which they played in the playroom or colored quietly in the experiment room. No stimuli from the experiment were present during this time and children were actively engaged in free play with their parents. This delay period ensured that the ensuing retention was not simply an immediate re-selection of the same items.

Retention

Following the five-minute break, children were presented with a single warm-up trial in the same manner as the previous warm-up and using objects from the same subset of known items. This was meant to briefly re-engage the child with the task. Four retention trials immediately followed warm-up. On each retention trial, two novel items that had previously been the targets on Novel RS trials were present along with a third item that had served as a pre-familiarized foil item during referent selection. Thus, all three items had previously been seen and were equally salient to the child, but only two had previously been mapped to a novel word. Like before, children were given a brief moment to view the items prior to the experimenter's prompt. Children were asked to retrieve one of the novel items by name. Items and prompts were not repeated across trials.

Finally, children were presented with four secondary referent selection trials (see Wilkinson, 2007; Wilkinson, Ross, & Diamond, 2003). Here, children were given two items that had previously been targets on Novel RS trials (one of which had been asked for a second time at retention) along with a third novel pre-familiarized item from referent selection that had no corresponding label and had not yet been used in retention. Critically, on secondary retention trials, children were now given a completely novel name that followed the same phonological rules of English, but had not been heard by the child previously. These trials were a less stringent measure of retention based on children's ability to use mutual exclusivity. If a child had already mapped a novel label to two of the novel items present, they should be significantly more likely to now choose an unnamed novel item as the referent for the new novel word.

Preference trials

Following retention, children were presented with five preference trials during which one known item and one novel item were present. No item had been previously seen or used in the task, but parents confirmed children's familiarity with the known items and lack of familiarity with the novel items prior to the experiment. Items were placed on the tray and children were asked "Can you get one?" before the tray was slid forward for the child to make a choice. These trials gave a measure of children's overall novelty versus familiarity bias in the absence of any word or label.

Coding

Naïve coders blind to the hypothesis coded children's final selections off-line. Data from twenty-two random subjects (51.1%) were re-coded for reliability purposes. Inter-coder agreement was high (96.27% across all selections), all discrepancies were settled via discussion with a third coder blind to the hypothesis and condition.

Results and discussion

Overall

Results of all trials and vocabulary groups are reported in Table A4. Consistent with prior work, children were overall very good at selecting the novel target item during Novel RS trials (69.7%), $t(42) = 13.198$, $p<.0001$, $d=1.662$, confirming the supernovel effect from Horst et al (2011) and children's use of novelty in a referent selection task. Because there are no known items to interfere with disambiguation here but performance was still well above chance, this suggests that children in Experiment 1 were likely relying primarily on novelty to select the referent. Also consistent with prior work, children overall accurately selected the target during Known RS trials, at levels significantly better than chance performance (74.1 %), $t(42) =12.516$, $p<.0001$, $d=1.909$. This differs significantly from Experiment 1 however, as here Known RS trials consisted of three known items and no novel foil items. Thus, children were not able to select the novel foil item as they did in Experiment 1 and instead, reliably selected the correct target. This further confirms children's novelty bias in Experiment 1.

When tested on their retention abilities, however, children were dismal. Overall, children selected the target only 35.6% of the time, *t*(40) =.614, *p*=.543, *ns*, *d*=.096.[2] However, when children erred on retention (which occurred over 60% of the time), they chose the other novel object that had been the target on Novel RS trials at levels significantly above chance (43.2%), *t*(42)=2.994, *p*=.005, *d*=.452. There were no differences in choices to the target versus the unmapped pre-familiarized item during retention, *t*(84)=1.0729, *p*=.2864, *ns*. This could be taken to suggest that children have some weak memory of which items had been the targets on Novel RS trials, but that these mappings were easily confused. It is also possible, however, that children are biased to choose the items to which they have less overall familiarity. In the retention trials, the third unnamed item was a novel pre-familiarized foil item that children previously explored and which appeared multiple times throughout referent selection. Thus, it is arguably more familiar that the Novel RS target items that were only seen twice across the entire session. In this light, children would be more likely to attend to the relatively more novel items that had been previously mapped during referent selection. During the secondary retention trials, children were at chance (36.63%), *t*(42)=.854, *p*= .3974, *d* = .130, *ns*.

Interestingly, when children were tested on their novelty versus familiarity preference *after* both referent selection and retention, all children regardless of age or vocabulary level demonstrated a robust novelty preference (63.3%), *t*(42)=3.49, *p*=.0014,

---

[2] This analysis was done across only trials on which a child had previously mapped the novel word during referent selection. Two children failed to map at all during referent selection and thus, were not included in the analysis. However, there were no differences is significance if the analysis is done across all retention trials (37.8%) *t*(42) =1.294 , *p*=.2029, *d*=.197, *ns*.

d=.521, versus chance (50%). It is possible that this reflects the baseline children's endogenous novelty bias children brought to the task; however because the novelty versus familiarity preference was assessed post-hoc, it is also possible that high novelty biases post-test are indicative of a training effect in which children were taught over the course of referent selection and retention to attend to novelty.

Low vocabulary children

When children are split by median vocabulary into low and high vocabulary groups, the results hold. Children in the low vocabulary group were significantly above chance on Novel RS trials, choosing the supernovel target 72.92% of the time, $t(21)=10.741$, $p<.0001$, $d=1.652$, and confirming that the supernovel effect (Horst et al., 2011) is not dependent upon productive vocabulary. Contrary to the results of Experiment 1 however, even children with a low overall vocabulary like the 18-month-old children of Experiment 1, could reliably find the target item during Known RS trials (68.7%) and did so at significant levels, $t(21)=8.036$, $p<.0001$, $d=1.71$. Of course, the critical difference here is that no novel foil items were present, only well-known familiar objects. These children also showed a marginal improvement in their performance as the experiment unfolded; performing at 63.64% during the first half of the Known RS trials and at 73.48% for the last Known RS trials, $t(42)=1.898$, $p=.065$, $d=-.352$. When tested on retention, however, low vocabulary children failed to select the target item above chance levels (29.2%), $t(19)= -.615$, $p=.546$, *ns*.[3] Low productive vocabulary then, does

[3]This analysis was done across only retention trials asking for an item to which the child originally mapped in referent selection. Two children failed to map at all during referent selection and thus, were not included in the analysis. However, there were no differences in significance if the analysis is done across all retention trials (30.68%), $t(21) =-.579$, $p=.569$, *ns*.

not seem to interfere with the ability to select the target during referent selection, but also results in no evidence of retention of newly mapped words from a supernovel context. When errors to the other named item on retention trials are broken down by vocabulary group, we see that the preference for the other named item is being driven primarily by lower vocabulary children who chose the other named item roughly 50% of the time, $t(21)=5.168$, $p<.0001$, $d=1.1$. Given the results of Experiment 1 and 18-month-olds' (and thus, low vocabulary children's) profound novelty bias, it is likely that children's choices and errors on retention were driven again by relative familiarity with the objects present.

High vocabulary children

Like the low vocabulary group, children in the high vocabulary group were also very good at Novel RS, choosing the target 66.35% of the time, $t(20) = 8.094$, $p<.0001$, $d=1.53$. These high vocabulary children were very good at Known RS trials as well, selecting the target 79.8% of the time, $t(20)=10.073$, $p<.0001$, $d=2.2$. During retention, high vocabulary children were marginally better than chance at selecting the target item, choosing the target 45.06% of the time, $t(20)= 2.10$, $p=.0485$, $d=.459$.[4] This suggests first, that productive vocabulary may play a role in retention for words mapped in a supernovel context, and second, that novelty biases may be overcome when vocabulary is robust enough.

Overall, the results of Experiment 2 show that despite catering to young children's endogenous novelty biases, low vocabulary children still cannot retain new word-referent pairs over a short break when novelty is the primary cue available during a

[4] This analysis was done across only trials which a child had previously mapped the novel word during referent selection. However, there were no differences in significance if the analysis is done across all retention trials (45.24%), $t(20) =2.04$, $p=.0545$, $d=.459$.

word learning task. Though children in the low vocabulary group have robust novelty biases, this strong bias does not lead to word learning. Only children with larger productive vocabularies showed evidence of retaining new word-referent pairs learned in a supernovel context. Thus, merely attending to an item because it is novel does not mean the child has encoded any new information in the single instance, or at least, has not gained enough robust knowledge to withstand time or a rigorous test of retention, unless there is also a robust underlying vocabulary. Consistent with the results of Simulation 2, novelty is most helpful in finding referents in the moment, but does not appear to play a critical role in long-term learning.

Simulation 3

Experiment 2 demonstrated children's high novelty bias can increase referent selection, but does not help in retention unless coupled with a high productive vocabulary. Simulation 2 confirmed the necessity of such a novelty bias in a HRN model, but did not explore moment-to-moment changes in referent selection or retention when lexical knowledge is not driving activation but instead, activation is reliant first on novelty. The current stimulation explores this possibility by using the same model to simulate the results from Experiment 2.

Methods

The model has the same architecture as outlined in Chapter 2 and used in Simulation 2, including a novelty bias. However, five auditory and visual units were reserved as supernovel items. These items were not trained during vocabulary development and were not used as foils during vocabulary acquisition. A separate set of five units remained as novel, pre-familiarized items. These novel units operated identically to the novel items in Simulations 1 and 2 in which they were present a small portion of the time during vocabulary acquisition, but not previously mapped to a specific label. Thus, the model had a total of 25 possible known words and five supernovel items.

The model began like Simulation 2 in which it acquired an initial vocabulary via multiple referent selection trials with various numbers of competitors and targets available. Like previously, every 500 epochs, the model was tested on its productive vocabulary. During this acquisition, the five novel items are present as foils a very small percentage of the time – 1%. This simulates the pre-familiarization period of Experiment 2. Referent selection proceeded with an identical set-up to Experiment 2 including implementation of the novelty bias; Known RS includes three items that have passed a 3AFC comprehension test along with a known label and Novel RS includes two novel items that were pre-familiarized along with one supernovel item and a novel label. The model cycles as normal until a single "choice" emerges on each trial. Retention is subsequently tested using two supernovel items from RS and a pre-familiarized foil and results are recorded but not saved so the model may continue to develop and be tested at different points in development for a cross-sectional approach.

Results and discussion

Results from the simulations are depicted in Figure A11. Overall, the model captures the general trends of children in Experiment 2. When the overall productive vocabulary of the model was lower (i.e. when the model "produces" roughly 30% of its total vocabulary, approximately epochs 8,000-32,000, see Figure A12a), performance on Known RS is well above chance (approximately 80%), but not yet at ceiling, just as it is with children in Experiment 2. At the same time, the model is also significantly above chance on Novel RS trials, demonstrating the supernovel effect seen in the empirical results of Experiment 2 and Horst et al. (2011).

As the model's vocabulary grew (and increased to at least 90% of its total lexicon, approximately epochs 48,000-80,000 in this model, Figure A12b), its ability to succeed on both Known and Novel RS also increased. However, children in Experiment 2 do not demonstrate a significant difference between ages or vocabulary level. This could partially be due to the approximate estimates of the model's vocabulary and the fact that the model does not map directly onto the age and vocabularies of children, but may represent a larger range of vocabulary development than what is tested here. This issue is discussed further in the limitations section of Chapter 6.

A major difference between the model and children in the current experiment is retention. Children with lower vocabularies were at chance for retention in the task, but the model performs above chance across the board, particularly at higher vocabulary levels. Though high vocabulary children in Experiment 2 did show marginal evidence of retention, the model performs higher on retention than low vocabulary children (Figure A12b). These differences might be due to differences in how the task is implemented in the model compared to the task presented to children. There are no perceptual biases in the model and thus, in order to simulate the results of Experiment 1, an artificial novelty bias was implemented. The associative nature of the model means that it inherently operates based on prior word-referent associations or lack thereof and thus, the novelty bias is critical to explain referent selection performance, particularly in younger 18-month-old children. Because activations, including the novelty boost, cycle between visual and auditory layers, this novelty bias could spread to both layers. When a single visual node is given activation in the input layer, that activation necessarily spreads to the lexical layer and back to the auditory input layer. Thus, when the respective auditory node for the visual unit is also active, its activation has received some of the benefit of the novelty boost. With children, however, it is unclear if a novelty-driven visual task requires the use of the auditory domain at all because children could select a referent based on novelty without encoding the word-form. Thus, it is possible that the model had more support for both word and referents during referent selection because of its architecture, leading to higher retention whereas children using novelty could be lacking in strong associations to the label. On the other hand, a close look at the changes in connection weights in the model after referent selection shows very little learning in the supernovel word-referent associations (Figure A13). This suggests that very small changes in a trial might be compounded with a novelty bias in addition to the novelty-driven task to support retention. The differences in the model are further explained in Chapter 6.

Conclusions about the role of novelty

Taken together, the results of Experiment 2 and Simulation 3 give support for the role of novelty in referent selection. However, the role of novelty in retention is much weaker. Experiment 1 showed an overriding novelty bias in low vocabulary 18-month-old children during referent selection. Simulation 2 demonstrated how a novelty bias is, in fact, necessary for the model to simulate children's behavior in the task. Experiment 2 pushed novelty further, using a paradigm that eliminated name-known referents. These results suggest there might be an interaction of overall productive vocabulary with retention in novelty-based tasks, but overall they give stronger support for novelty aiding in referent selection. Thus, children's endogenous novelty biases can indeed help referent selection, but this bias does not seem to help children retain novel words and the bias can overriding the use of prior knowledge at least during referent selection.

CHAPTER FIVE

EXPERIMENT 3: THE ROLE OF PRIOR WORD KNOWLEDGE IN REFERENT SELECTION AND RETENTION

Novel words are simply those which are not yet known. In young children, this process of transforming a novel word to a known one is understood to draw on both a child's novelty bias and prior vocabulary knowledge. The goal of the current thesis was to explore both. Experiments 1 and 2 and the respective simulations show a clear role for novelty biases in children's referent selection and that mappings induced by novelty alone do not robustly lead to retention. Experiment 3 then explores the role of prior vocabulary knowledge using two empirical studies and two simulations with the HRN model presented in Chapter 2.

Experiment 3a

Prior vocabulary knowledge is involved in referent selection and retention in two ways: knowledge for specific word-referent pairs and overall productive vocabulary level. Children can utilize their knowledge about specific labels for referents as a foundation to bootstrap new word-referent pairs, such as the theory of mutual exclusivity (Heibeck & Markman, 1987; Merriman & Bowman, 1989). The presence of known and novel items along with known or novel labels implicates a child's current vocabulary and knowledge and thus, competition of prior knowledge and current objects is inherent in ambiguous learning paradigms like those used in the current thesis. Many accounts of word learning have often recruited the use of prior knowledge and competition even if the procedure itself is not lexically-based (see Mather, 2013). However, few studies have explicitly explored how the strength of prior knowledge influences learning. Merriman and Schuster (1991) examined changes in disambiguation abilities in 2- and 4-year-old children in accord with high or low typicality of the words and referents present. Here, children were presented with pairs of pictures that showed prototypical exemplars of known words (e.g. 4-legged black and white cow) or atypical exemplars of known items (e.g. 8-legged purple cow) along with a novel object. Children were then either given a novel word, or a known word that was pronounced either typically (e.g. *apple*) or atypically modified (e.g. *japple*). In cases where atypical words or referents were used, children were much poorer at using mutual exclusivity during referent selection, suggesting the ability to align present objects with prior knowledge is critical for learning. However, the task used differed in procedure and age from the current thesis, was confounded with a pre-familiarization period to both novel and known stimuli, repeated stimuli over the course of the test, and did not test for retention. Furthermore, Merriman and Schuster (1991) did not systematically control for strength of word-referent knowledge because all known items had broader category representations for the child to draw on during the task. Thus, the findings cannot be directly integrated with the current literature on referent selection and do not give a clear picture of the role of specific word knowledge in word learning.

A handful of other work has identified that children's overall productive vocabulary can influence word learning by providing a solid foundation into which to integrate new words (Bion et al., 2013; Mervis & Bertrand, 1994). Here, a child's performance on a referent selection task was directly correlated with his or her productive vocabulary. Simulations 1 and 2 also support the influence of productive vocabulary knowledge on referent selection and retention over development. Again, however, the prior work differs in the procedure used and ages tested and do not manipulate specific vocabulary knowledge in the task. To critically examine the role prior vocabulary knowledge plays in the process of word learning, Experiment 3 explored referent selection in a context of either well-known or weakly-known competitor items across children at a wide range of productive vocabulary levels from 18- to 24-months-old.

The strongly-known foil items were ones that most 18-month-old children would likely have extensive experience with prior to the visit to the lab, such as a dog, a shoe, and a cup. In contrast, children were taught the names of the weakly-known foil items just prior to testing. In this way, children were expected to have widespread and variable experience with category examples for strongly-known foils including multiple encounters with the word, object, and the simultaneous presentation of both, but would be limited to a single exemplar, context, and exposure for the weakly-known foils. Thus, training for the weakly known words restricted not just the number of times the item was seen, but also the richness of information available to encode. This is known to lead to known differences in a child's representations and encoding of novel word-referent mappings (Perry & Samuelson, 2011; Smith & Yu, 2088; Wagar & Dixon, 2005; White & Morgan, 2008).

Training was set up to guarantee children were very familiar with the strong-known foils and ensure they actually learned the words for the weakly-known foils, even if they were only given limited knowledge of the weakly-known objects. Prior work has demonstrated that children can quickly learn the label for a novel object when presented in an ostensive naming context in which the name and item are semi-isolated and pragmatic cues are in place (Horst & Samuelson, 2008; Axelsson, Churchley, & Horst, 2012). Other work has suggested that competition during exposure to a new word can boost retention further (Hills, Maouene, Riordan, & Smith, 2010; McMurray et al., 2012) by helping to narrow the scope of possible referents. This competitive process can be heightened further when the competing items are already known, for example, when a to-be-named item is paired with a familiar cup (Horst, et al., 2006; McMurray, et al., 2012; Baldwin & Markman, 1989; Carey, 1978; Heibeck & Markman, 1987). Furthermore, it was critical to ensure that children were very familiar with the specific exemplar of the strong well-known objects that would be used in the study. Thus, training for the new "weakly-known" items was initially done in a competitive context in which all three well-known, strong items were present along with the to-be-trained weakly-known items.

The manipulation of word knowledge (strongly- vs. weakly-known) was done within subjects and allowed for a comparison of referent selection and retention in the more ambiguous context of weakly known words versus the less ambiguous context of well-known words. In the former case, children are presented with three items, none of which have a very rich representation in memory and only two of which have some minimal association with a label. This creates an environment that should be harder for children to disambiguate to determine which items are familiar, which are novel, which have names, and what labels map onto the objects (Merriman & Schuster, 1991). One would predict, then, that children of all ages would have difficulty during these referent selection trials since the objects would be equally unfamiliar. On the other hand, older children with higher vocabularies should have a rich representation of commonly-known toys, such as a car or dog, making it easy to process them when encountered during referent selection. When strong, well-known competing referents are used during referent selection then, children should be very good at quickly disambiguating the scene and perform much higher on referent selection, especially when supported by a rich overall vocabulary as well. Furthermore, because children are trained on the specific exemplar prior to test, this should ease the process further enabling children, particularly high vocabulary children, to override their novelty bias and retain new word-object pairs. These predictions are tested in two different referent selection and retention tests with 18- and 24-month-old children using both strong-known and weakly-known familiar items.

Methods

Participants

Forty-two, 18- to 24-month-old children (17 females, M = 22 months, 26 days; range 18:11-25:24) with a mean vocabulary of 242.9 words (range = 1-645, median = 184) participated. This group could be further subdivided into two groups of children based on age and vocabulary level. In particular, the younger group of fifteen children (7 females) were between 17 ½ months and 21 ½ months of age with a mean age of 19 months, 19 days. This group had a productive vocabulary of 129.4 with a range of 0-566 words. This group of children were slightly older than those in Experiment 1, $t(45)=4.915$, $p<.0001$, but the same age as the children in Experiment 2, $t(31)=.1478$, $p=.8835$, *ns*. In addition, their productive vocabulary was equivalent to 18-month-old children in both Experiment 1, $t(45)=1.538$, $p=.1310$, *ns*, and in Experiment 2, $t(29)=.5988$, $p=.5540$, *ns*. The older group of 27 children (10 female) were between 21 ½ and 25-months of age with a mean age of 24 months, 19 days. This group had an average productive vocabulary of 308.42 with a range of 36-645. This group of 24-month-old children was comparable in age and vocabulary to those that participated in Experiment

2; age *t*(31)=.1478, *p*=.8835, *ns*, vocabulary *t*(49)=.7745, *p*=.4423, *ns*. Data for three additional children were not included due to fussiness (2) and experimenter error (1). Children received a small prize for participating and had not participated in any previous referent selection experiment, including Experiments 1 or 2.

Stimuli

The same subsets of known and novel objects from Experiment 2 were used. In addition, three novel items were selected from the previous set of novel objects to be used as weakly known items (see Figure A4c). Parents confirmed the status of each object as either known or novel prior to test. Items were replaced as needed. The same set of novel CVC words from Experiment 2 was also used.

Procedure and design

During the entire procedure, the child was seated across a table from the experimenter in a booster seat next to their parents or on their parents' lap. Parents completed the MacArthur-Bates Communicative Development Inventory (MCDI; Fenson et al, 1994) during the session and were instructed to avoid interacting with their child, offering minimal encouragement only if needed. See Figure A14a and Table A2 for outlines of the procedure.

Training

The procedure began with a training period during which children were taught the names for three novel items and introduced to the specific exemplar and label for three known items. In line with prior work showing that competition plays a critical role in learning (Hills, et al., 2010), children were simultaneously given six items: 3 to-be-named novel items and 3 familiar known items. Three of the items were selected from the set of sixteen known items the parents had confirmed the child's knowledge prior to test. The three novel to-be-named items were selected from the larger set of unknown items, confirmed by parents to be previously unknown to the child, see Figure A3c. As the child explored each item, the experimenter would point and name the item the child was attending to. If the child did not engage with the stimuli, the experimenter drew the child's attention to each item as needed by holding up the item and/or pointing to it and repeating the name for the item multiple times in a conversational context (i.e. "Look! This is a *blick*! See the *blick*?"). Each item was named a minimum of three times over the course of the two-minute training period. The training period ended with a review in which the experimenter lined the six items up on the table out of reach of the child and systematically held each up and repeated the name. After the child had been trained on each of the items, they were removed from the table. Following training, the experiment proceeded directly into warm-up.

Warm-up/comprehension

After training, there were six warm-up trials in which the known items used in training were presented on a white tray at equidistance from each other. On the first three warm-up trials, the child was presented with the three previously-known items (e.g. shoe, car, dog) and asked to find a single known item by name. These items are referred to as strongly-known words. On the final set of three warm-up trials, the child was presented with the three just learned items from the training period (e.g. blick, zeb, daf) and asked to select the referent using the words that had just been taught. These items are referred to as weakly-known words. On each trial, the child could observe, but not reach the objects for three seconds prior to the experimenter prompting the child to get one of the items by name ("Can you get the *shoe*?") and sliding the tray with the items within reach of the child to make their choice. The child was re-prompted up to three times if necessary. Children were praised or corrected as needed. Because children were given a chance to select their choice in response to the prompt prior to any correction or praise being given, these trials also work as a comprehension check where children's errors or initial mistakes during warm-up provide an indication of the strength of the child's knowledge. Across all three trials target location and object were randomized and never repeated.

Referent Selection

A series of 16 referent selection trials immediately followed warm-up. On half the trials, children were asked to find a known item by name. These are collectively referred to as Known RS trials; four of these trials included two well-known referents (i.e. dog and shoe) along with a single novel item. These are the Strong Known RS trials. The other four Known RS trials included two weakly-known items the children had just learned the name for during training (i.e. blick and zeb) along with a novel item. These are the Weak Known RS trials. On the remaining 8 trials, children were asked to find a novel item by name, known as the Novel RS trials. Four of these novel referent trials had well-known, strong foils (e.g. dog and shoe) along with a never-before-seen novel object from the larger set depicted in Figure A4a and a novel label. These are the Strong Novel RS trials. The other four Novel RS trials had just-trained, weakly known foils (e.g. blick and zeb) with a never-before-seen novel object and a novel label. These are the Weak Novel RS trials. In all cases, the location of the novel item was randomized and did not repeat across trials. Known RS and Novel RS trials alternated over the course of referent selection.

Break

Immediately following referent selection, children took a five-minute break during which they played in the playroom or colored quietly in the experiment room. No stimuli from the experiment were present during this time and children were actively engaged in free play with their parents. This delay period ensured that the ensuing retention was not simply an immediate re-selection of the same items.

Retention

Following a five-minute break, children were given two warm-up trials in the same manner as the previous warm-up. The three strongly known items were used on the first trial and the three weakly known items were used on the second warm-up trial. This was meant to briefly re-engage the child with the task. Four retention trials immediately followed warm-up. On each retention trial, two novel items that had previously been the targets on Novel RS trials were present along with a third novel item that had served as a foil on a Known RS trial. Thus, all three items had previously been seen and were equally salient to the child, but only two had been mapped to a novel word. On two of the retention trials, target items were drawn only from the Strong RS trials and children were asked to recall the referent for a novel name initially presented on a Strong Novel RS trial. These are the Strong Retention trials. On the remaining two retention trials, the target was drawn from the Weak RS trials and children were asked to recall the referent for a novel word initially presented on a Weak Novel RS trial. These are the Weak Retention trials. In all cases, the target novel item was present along with another named item from the same strength class as well as an unnamed foil item from that same strength class. For example, if the target was initially named during a Strong Novel RS trial, then the named foil item was one from another Strong Novel RS trial and the unnamed foil was from a Strong Known RS trial.

Finally, children were presented with four secondary retention trials. Here, children were given two items that had previously been targets on the same strength Novel RS trials (one of which had been asked for a second time during retention) along with a third unnamed novel foil item, again from the same trial type. Two of the secondary retention trials drew objects from Strong RS trials whereas the other two secondary retention trials drew objects from Weak RS trials. Critically, children were now given a completely novel name that followed the same phonological rules of English, but had not been heard previously. Thus, these trials looked identical to the retention trials but were paired with a completely novel label. These trials were a less stringent measure of retention based on children's ability to use mutual exclusivity. If a child had already mapped novel labels to two of the novel items present, they should be significantly more likely to now choose an unnamed novel item as the referent for the new novel word. Critically, children had seen all items previously and thus, could not rely purely on novelty to select the referent.

Preference trials

Following retention, children were presented with five preference trials during which one known item and one novel item were present. No item had been previously seen or used in the task, but parents had confirmed children's familiarity with the known items and lack of familiarity with the novel items prior to the experiment. Items were placed on the tray and children were asked "Can you get one?" before the tray was slid forward for the child to make a choice. These trials gave a measure of children's overall novelty versus familiarity bias in the absence of any word or label.

Coding

Naïve coders blind to the hypothesis coded children's final selections off-line. Data from twenty-one random subjects (50%) were re-coded for reliability purposes. Inter-coder agreement was high (98.4% across all selections), all discrepancies were settled via discussion with a third coder blind to the hypothesis and condition.

Results and discussion

The focus of this experiment is the influence of strong versus weak word knowledge on referent selection and retention. Manipulating the strength of children's knowledge of the foil items within subjects allows for a direct comparison of children's performance in strongly-known and weakly-known trials with overall productive vocabulary and other individual differences held constant. I discuss each of the trial types – strongly-known and weakly-known – in turn, analyzing performance overall, in low vocabulary groups, and in high vocabulary groups for both referent selection and retention trials.

Strongly-known trials

Overall

Results of all trials and vocabulary groups are reported in Table A5. The configuration of Strong Known RS trials was similar to that of prior referent selection work with the exception that they were preceded by a training trial exposing children to the specific exemplar of the item used in the task. Given this, I would expect that higher vocabulary children would be especially able to succeed on Known RS trials. However, children overall failed to select the known target item and perform at chance (32.5%), despite the target being a very familiar, well-known and reviewed item, $t(41) = -.084$, p=.933, d=-.013, *ns*. Instead, children selected the novel foil item 66.27% of the time, $t(41) = 5.932$, p<.0001, d=.915, behaving much like the novelty-driven 18-month-old children of Experiment 1. As a group, children succeed at Strong Novel RS and did so with near ceiling performance (98.8% correct) regardless of age or vocabulary, $t(41)$=55.28, p<.0001, d=8.53, suggesting again, a high novelty bias. Overall, children still failed to retain the newly mapped words encountered in a well-known context, selecting the target 29.8% of the time, $t(41)$=-.60, p=.552, *ns*[5] and performing at chance on secondary retention trials (32.14%), $t(41)$=-.169, p=.8663, d=-.026, *ns*.

The vast discrepancy between the Strong Known RS results and prior work (Horst & Samuelson, 2008; Kucker & Samuelson, 2012) suggests that some component of the training protocol interfered with children's ability to correctly select the known referent when in an ambiguous context or when paired with a novel object. One possibility is that because children were immediately provided a name for any item they directed attention to during training, they then expected names to also be forthcoming during the main experiment (Yoshida & Burling, 2012). If so, children could have attended to the novel items more during referent selection in an effort to hear their names. However, I would expect then that in the course of the referent selection trials, in which a name is given *prior* to a choice and no name is given after, children would again quickly learn that now,

[5] This analysis was done across only retention trials to which the child had mapped the novel word during referent selection as per prior analysis. However, in this case, all children mapped during the RS trials and thus, analysis of only those trials to which children mapped is identical to analyzing all trials

no name will be given after a selection. This would result in a drop in performance across trials, which did not occur; correct selections were made at a rate of 36.95% on first half of trials and 27.4% of trials on second half, $t(82)=1.132$, $p=.2598$, *ns*.

Another possibility is that the poor Strong Known RS performance is related to children's exposure to only a single exemplar of each of the well-known items during training. Because category knowledge is critical for a robust representation of words (Mayor & Plunkett, 2010; Oakes, Coppage, & Dingel, 1997) and children often draw on their prior knowledge during word learning tasks (Merriman & Schuster, 1991), this restricted training protocol may have caused children to focus exclusively on specific single exemplars, thus limiting children's use of their prior knowledge. One other possibility is that reviewing the name for the well-known item could have resulted in children becoming habituated or bored with the item thereby ignoring it later during referent selection as they had already fully processed that specific exemplar. Future work is needed to test these possibilities.

Low vocabulary group

When children are split by median vocabulary into low and high vocabulary groups, the results hold. Children in the low vocabulary group selected the target during Novel RS trials at levels that were significantly difference from chance (97.5%), $t(19)=25.8$, $p<.0001$, $d=5.769$. Children in the low vocabulary group were at chance for selecting the target during Known RS trials, choosing the target item 21.25% of the time, $t(19)=-1.559$, $d=.1356$, *ns*. Instead, low vocabulary children selected the novel foil item the majority of the time (77.5%), and at levels well above chance, $t(19)=5.644$, $p<.0001$, $d=1.262$. This is consistent with the novelty bias demonstrated by children in Experiment

1. Because the Strong Known RS trials were similar to the Known RS trials during Experiment 1, it is not surprising that children also failed to retain the new word-referent pairs, selecting the target 27.5% of the time, *t*(19)=-.7168, *p*=.4822, *d*=-.1603, *ns*.[6] When errors to the other named item on retention trials are analyzed in children with low vocabulary, we still see selection at chance levels (37.5%), *t*(19)=.5118, *p*=.6147, *d*=.1145, *ns*. Children were at chance during the secondary retention trials as well (32.5 %), *t*(19)=-.0667, *p*=.9475, *d*=-.015. This demonstrates that even when trained on the specific exemplars of known items prior to referent selection, lower vocabulary children still demonstrate a robust novelty bias and fail to retain newly learned words.

High vocabulary group

Like the low vocabulary group, children in the high vocabulary group were also very good at Novel RS, with every single high vocabulary child choosing the target every single time (100%), *p*<.0001. However, contrary to prior work that demonstrated a robust ability to select a known item from an ambiguous array, high vocabulary children here performed very poorly on Strong Known RS trials, selecting the target at chance levels, only 42.8% of the time, *t*(21)=1.339, *p*=.1949, *d*=.2847, *ns*. Even these high vocabulary children demonstrated an overriding novelty bias, selecting the novel foil during Known RS (56.06%) instead of the known target, *t*(21)=3.091, *p*=.0055, *d*=.659, despite additional review of the strongly-known items during the training section of the experiment. Given the high novelty bias and performance equivalent to the low

[6] This analysis was done across a subset of the retention trials using only those which the child had previously mapped during referent selection, as per prior analysis. However, there is no difference in results when the analysis is done over all retention trials (27.5%), *t*(19)=-.716, *p*=.4822, *ns*.

vocabulary group, it is not surprising that these high vocabulary children still failed to retain newly mapped words (31.82%), *t*(21)=-.1526, *p*=.8802, *d*=-.0325, *ns*.[7] Children were at chance during the secondary retention trials as well (31.82 %), *t*(21)=-.1685, *p*=.868, *d*=-.0359.

Taken together, these results suggest that overall productive vocabulary appears to have little influence on referent selection and retention performance when competitors during referent selection are well-known. However, this conclusion runs contrary to the bulk of prior work showing minimally, a strong ability of 24-month-old and high vocabulary children to perform well on Known RS trials. This then support to the possibility that the pre-training on the specific exemplars of known items and/or the training on novel words prior to the task interfered with children's ability to use their prior knowledge during referent selection. Analyzing results on weakly known trials can help explain the results from strong known trials.

<u>Weakly-known item trials</u>

Overall

Prior work suggests that children should have more problems with the Weak RS trials as they are not able to bootstrap their prior knowledge as well in order to use mutual exclusivity to disambiguate the scene (Merriman & Schuster, 1991). The current results partially support this notion. Children overall performed well on Weak Novel RS trials, selecting the target novel object 83.9% of the time, *t*(41)=16.079, *p*<.0001, *d*=2.481, suggesting that competitor knowledge does not interfere with children's ability to

[7] This analysis was done across only those trials which a child had mapped the novel word during referent selection as per prior analysis. However, because all children mapped all words during referent selection, an analysis over all possible retention trials is identical.

disambiguate and select the novel target item. Performance was compromised on the Known RS trials, however. Children overall failed on Weak Known RS trials, performing at levels significantly below chance (12.9%), $t(41)=-5.815$, $p<.0001$, $d=-.8972$, and instead choosing the novel foil item at levels significantly above chance (83.5%), $t(41)=13.627$, $p<.0001$, $d=2.103$. This might suggest that children in fact took the novel foil item to be the referent for the given label, despite having been taught previously that the label referred to another item. This is certainly possible given the fleeting experience children had with the weak known items and the high number of errors or mistakes seen during the weak warm-up/comprehension trials (an average of 2.31 errors across all 3 trials, range 0-5 over all three trials) as compared to errors during the strong/well-known warm-up/comprehension trials (an average of .79 errors across all 3 trials, range = 0-65 over all three trials), between groups $t(82)=5.724$, $p<.0001$, $d=-1.249$. The poor performance on Weak Known RS trials could also be evidence for an overriding novelty bias, much like that seen in Experiment 1. In this case, when the competing foil items are weakly known or possibly never initially encoded, children resort to using novelty instead of prior knowledge.

If children are indeed selecting the target item based primarily on novelty and not prior knowledge on Weak RS trials, we would suspect that, like in Experiment 2, they would be unable to retain new word-object pairs. Quite surprisingly, however, children overall demonstrate an ability to retain new word-object pairs mapped in an ambiguous context with competing items that are very weakly known, selecting the target significantly more than chance, 63.1% of the time, $t(42)= 4.71$, $p<.0001$, $d=.727$.[8] Children were also slightly above chance during the secondary retention trials (46.4%), $t(41)=2.153$, $p=.0327$, $d=.332$. These results suggest that the specific word knowledge during referent selection matters for retention, but in the direction opposite than that predicted. This finding is further clarified by splitting children into low and high vocabulary groups.

Low vocabulary group

When children are further divided into low and high vocabulary according to a median split, there are slight differences between the groups. Children in the low vocabulary group selected the target during Weak Novel RS trials and did so at significant levels (82.5%), $t(19)=10.24$, $p<.0001$, $d=2.29$. On Known RS trials, these children were significantly *below* chance for selecting the target, choosing it only 13.33% of the time, $t(19)=-3.313$, $p=.0037$, $d=-.7409$. Instead, they selected the novel foil item the majority of the time (84.17%) and well above chance levels, $t(19)=8.688$, $p<.0001$, $d=1.94$. This is not surprising given the very weak knowledge children had for the known referents and the strong novelty bias present in lower vocabulary children.

Children in the low vocabulary group thus demonstrate an overriding novelty bias and poor knowledge for the just-trained weakly known items. The combination would predict poor retention for words mapped in a weakly known context. However, these low vocabulary children selected the target during retention 65% of the time, significantly

[8] Again, analysis was done over only trials which the child had previously mapped. However, again, results were still significantly above chance when analysis is done over all items on retention, 47.62%, $t(41)= 2.87$, $p.=0065$, $d=.442$.

more than chance, *t*(19)= 3.57, *p*=.002, *d*=.797[9], though children were at chance on the secondary retention trials (47.5%), *t*(19)=1.299, *p*=.2096, *d*=.29. This counter-intuitive retention result suggests that during referent selection, children were doing more encoding than blindly attending to the most novel object present. In order to retain, children must have encoded some information about the word-form as well as made a minimal association between the word and object. One possibility is that when disambiguation is harder, such as when all items present are weakly known, it takes more cognitive 'work' to select the correct novel referent on the Novel RS trials. Thus, it is also possible that more information could be gained from a single trial due to the increase in processing needed to solve the referential ambiguity of a single referent selection trial. This idea is elaborated in the discussion.

High vocabulary group

Like the low vocabulary group, children in the high vocabulary group were also very good at Novel RS, choosing the target 85.23% of the time, *t*(21)=12.3, *p*<.0001, *d*=2.29. Despite their high vocabulary knowledge, however, these children still performed poorly on Known RS trials, selecting the target only 12.5% of the time, a level significantly below chance, *t*(21)=-5.197, *p*<.0001, *d*=-1.108. On these trials, children selected the novel foil instead (82.95%), *t*(21)=10.487, *p*<.0001, *d*=2.236. High vocabulary children appear to be behaving similarly to the 18-month-old children on Weak RS trials, thus suggesting these children will also demonstrate retention and leading to the conclusion that overall vocabulary knowledge does not matter for referent

---

[9] Again, analysis was done over only the subset of trials which the child had previously mapped. However, again, results were still significantly above chance when analysis is done over all possible retention trials, 52.5%, *t*(19)= 2.541, *p*=.0199, *d*=.568.

selection in the context of weakly known words. However, these high vocabulary children do retain, selecting the target 61.36% of the time, *t*(21)= 3.06, *p*=.0059, *d*=.653[10]. The high vocabulary children were not significantly different from the low vocabulary children on retention, *t*(40)=.2810, *p*=.7801, *ns*, but did select the target at marginally significant levels during the secondary retention trials (45.45%), *t*(21)=1.915, *p*=.0692, *d*=.408. This suggests that high vocabulary children may have encoded some minimal information during referent selection, but that knowledge was too weak to withstand the break period and the subsequent rigorous tests of retention.

Preference trials

Despite a strong novelty preference in the preference trials of Experiment 2 when children were given a known item paired with a novel item, they choose the familiar and novel items equally (49.8%), *t*(40)=-.0506, *p*=.960, *d*=-.0079, *ns*. Some literature suggests children around 2-years-old have an endogenous novelty bias, while other work proposes massive individual differences in which children often fluctuate and have no consistent preference (Cohen, 2004; Houston-Price & Nakai, 2004; Hunter & Ames, 1988). Though it is unclear what mechanism is driving children's preference here, it is clear that novelty vs. familiarity preference after a novelty-based task (Experiment 2) or after a knowledge-based task (Experiment 3a) is different, with the task itself being likely to have influenced that preference. A post-hoc analysis of novelty vs. familiarity

[10] Again, analysis was done over all retention trials and items. When analysis is just done over those items to which children had correctly selected on referent selection, children are no longer significantly above chance, but now marginally above chance, 43.18%, *t*(21)= 1.493, *p*=.1502, *d*=.318, *ns*.

preference with performance on the task was not significant across both referent selection and retention.

Simulation 4a

Overall, the results of Experiment 3a suggest that children's relative knowledge about objects might matter for learning. In this case, when weakly known items were paired with a novel item, children demonstrated sufficient learning during a single referent selection trial in order to demonstrate later retention for those word-referent pairs. Prior simulations have described this process of referent selection as a competition between known and novel referents and words. In order to explore the effect of relative knowledge further, the model from Chapter 2 was used to simulation the empirical results from Experiment 3a. Specifically, ostensive naming in a competitive environment was used to train the model on new weakly-known words. These were then paired with novel items during referent selection. Like previously, both selection behavior and connection weights were analyzed after referent selection and retention.

Methods

The model has the same general and specific architecture as outlined in Chapter 2 and used in Simulation 2, including a novelty bias. Here, however, the model has a total of 27 words it can learn during the initial vocabulary learning phase (as opposed to 30). The additional 3 units are held out from vocabulary learning to be used as weakly known items trained at a later time. The model began like the baseline model, acquiring an initial vocabulary via multiple referent selection trials with multiple and variable competitors and targets available. During this acquisition, novel items are present as foils a very small percentage of the time – 1%. This is needed to keep these units integrated enough to be used later in the referent selection task. This simulates the real-life exposure children get to multiple objects to which they do not yet know the name for words which do not yet have a referent. To-be-trained items are never present as foils, however. This allows for a test on the effect of ostensive naming during training that is not confounded with prior experience. Prior to each referent selection test, the model is given three known items plus the three to-be-trained items along with a single novel name. In order to simulate ostensive naming in the context of multiple other items, the target object here is given a slight boost of .03 before normalization. The model then cycles through to map the respective label and object (see Horst, 2008). This is repeated for each of the six objects present; the three held-out units and three known units.

Referent selection then proceeds in a manner identical to the prior simulations but with a distinction between strong and weak trials. Strong trials have two well-known items that were learned during general vocabulary acquisition and again reviewed during training along with a novel item. Weak trials have two of the three newly-trained, but weakly-known items along with a novel object. The model cycles as normal until a single "choice" emerges on each trial. Retention is tested as well and results are recorded but not saved so the model may continue to develop and be tested at various points in vocabulary acquisition.

Results and discussion

Results from the simulation are depicted in Figure A15. Overall, the model performs similarly to children on most aspects of the task. However, the first critical question is if the model can learn the new weakly-known words during ostensive naming training. Although the architecture of the model guarantees that there are changes in word-referent associations during training, the model does not actually demonstrate accurate knowledge for these newly learned words during a 3AFC comprehension test. This is not necessarily surprising given that the model's overall vocabulary takes thousands of epochs to gain a handful of words and ostensive naming occurs over the course of a single epoch with a single repetition of the word-referent pair. Though this is a limitation to the model, it might also map onto children who do not always demonstrate knowledge of these new words. As we know, knowledge is a slow accumulation over time and thus, the model and children, can demonstrate some minimal honing and pruning of connections without those connections yet being sufficient to hold up to a rigorous comprehension test. Furthermore, as we will see, differences during referent selection in strongly versus weakly known words suggest that the model treated well known and weakly known words differently from each other and from novel unnamed items.

On Strong Known RS trials, the model initially simulated the behavior of children with performance hovering near chance in the early stages of vocabulary acquisition (Figure A16a). However, as the model gains a vocabulary, its performance on Strong Known RS trials increases and overcomes the artificial novelty bias. Performance on Weak Known RS trials, however, remains poor in both the model and children across all ages and vocabulary levels (Figure A16b). Like previously, low vocabulary and early development are defined roughly when the model produces 30% of its total vocabulary, approximately epochs 10,000-21,000, and high vocabulary is when productive vocabulary is over 90%, roughly epochs 48,000-80,000 in this model, see Figure A16c.

Critically, the model also shows excellent performance on Novel RS trials for both strong and weak competitors at all ages. Though Weak Novel RS in the model is lower than Strong Novel RS, both are above chance, just as they are in children. Strong Novel RS is at ceiling as a result of the strong novelty bias combined with strong prior word-referent associations for the well-known strong competitors which helps the model prune out the unnecessary connections for known items.

When the model is tested on retention of the newly mapped words, it fails across the board. Although children here show a surprising ability to retain words mapped in a weakly-known context, the model fails at retention. Despite this failure, a fine-grained analysis of the connection weights in the model after referent selection give evidence for differential changes in learning for novel words mapped in a strong context versus novel words mapped in a weak context (see Figure A17). Here, the highest degree of change in connection weights is seen when weakly known items are associated following Strong RS trials. The second highest is for the connection of novel words mapped on Weak RS trials. Everything else, including the amount of change for novel words mapped in a strong context, falls close to floor indicating little learning. This suggests that the model can hone relevant connections and prune away spurious connections most efficiently when competing items are weakly connected themselves. Thus, an analysis of the change in connection weights for words mapped in the context of strong foils versus weak foils shows evidence for differences in learning.

Experiment 3b

Although the literature suggests that a competitive atmosphere during training for new words would result in the highest learning rates, Experiment 3a demonstrated that such a training atmosphere was detrimental to children's subsequent referent selection abilities. One reason for children's poor performance on Known RS trials during the referent selection task in 3a, is that children became too habituated or bored with the familiar stimuli and thus, resorted to using only novelty regardless of their underlying productive vocabulary skills. Similarly, training on only a single exemplar could have forced children to have a narrow view of the category and not be able to align the current exemplar to their prior knowledge, particularly with children in the high vocabulary group. Thus, the next step was to eliminate the competitive component and extra repetition of already known items that might have set up such gross differences between the well-known and weakly-known trials and masked the predicted vocabulary effects. Experiment 3b made two changes to the training period – using ostensive naming in isolation in order to train the children on the three new word-object pairs, and removing training on already known word-object pairs. Each to-be-trained item was presented one at a time and children did not see the well-known items prior to test.

Methods

Participants

Forty, 18- to 24-months-old children (19 females, *M* = 21 months, 17 days; range 17;29-25;15) with a mean vocabulary of 189.4 words (range = 0-673, median = 141) participated. This group could be further subdivided into two groups of children based on age and vocabulary level. In particular, the younger group of twenty-five children (13 females) were between 17 ½ months and 21 ½ months of age with a mean age of 19 months, 25 days. This group had an average productive vocabulary of 140.28 words with a range of 4-573 words. This group was equivalent to the younger group in Experiment

3a in both age, *t*(40)=.7046, *p*=.4854, *ns*, and vocabulary, *t*(38)=.2294, *p*=.8198, *ns*. The remaining 15 children were closer to 24-months-old and between 21.5 and 25-months. This older group of fifteen children (6 female) were between 21 ½ and 25-months of age with a mean age of 24 months, 13 days. This group had an average productive vocabulary of 271.33 words with a range of 0-673 words. This group was equivalent to the older group in Experiment 3a in both age, *t*(40)=.7427, *p*=.4620, *ns*, and vocabulary, *t*(39)=.5725, *p*=.5703, *ns*. Data for three additional children were not included due to fussiness (2) and experimenter error (1). Children received a small prize for participating and had not participated in Experiments 1, 2, or 3a.

Stimuli

Stimuli were identical to those used in Experiment 3a.

Procedure and design

The procedure was identical to that of Experiment 3a with one exception. Instead of a competitive training context with labels for six different items ranging in well-known to new, children were now taught the names for only the new items. Names were also taught one at a time. The experimenter placed a single to-be-named item on the table, directed the child’s attention to it, and gave its corresponding label (e.g. “Look! This is a *blick*. See the *blick*?). The child was then able to explore the item for roughly 30 seconds during which time the experimenter labeled the item 3-6 more times. Following this, the experimenter retrieved the item from the child, named it one final time while pointing to it, then put the item away and repeated the training with each of the remaining to-be-named weakly known items. See Figure A14b and Table A2 for outlines of the procedure.

Coding

Like previously, naïve coders blind to the hypothesis coded children's final selections off-line. Data from twenty-five subjects (62.5%) were re-coded for reliability purposes. Inter-coder agreement was high (98.4% across all selections), all discrepancies were settled via discussion with a third coder blind to the hypothesis and condition.

Results and discussion

Strong/well-known trials

Overall

Results of all trials and vocabulary groups are reported in Table A6. Like Experiment 3a, when given a novel label along with two strong, well-known items, children here were very good at selecting the novel target item (94.38%), $t(39)=32.37$, $p<.0001$, $d=5.118$. This was comparable to children's high performance in Experiment 3a on Strong Novel RS trials, between groups, $t(80)=2.00$, $p=.0488$, $d=-.439$. When asked to select a well-known item by name during the Strong Known RS trials, children as a group failed (39.58% correct), $t(39)=1.087$, $p=.2836$, $d=.172$, *ns*, initially suggesting that removing additional training on already known items and teaching new words in isolation using ostensive naming have little effect on improving later known referent selection performance.

Again in line with the results of Experiment 3a, when tested for their retention of words mapped in a strong, well-known context, there was still no evidence that the children learned the word-referent pair (36.25%), $t(39)= .547$, $p=.587$, $d=.0866$, *ns*.[11]

[11] Like previously, only a subset of retention trials were analyzed. However, if all trials on retention are analyzed for retention, children are still at chance (33.75%), $t(39)= .1367$, $p=.8919$, $d=.0216$, *ns*.

Furthermore, children had no preference for any items present, also choosing the other named item at chance levels (40%), $t(39)=1.224$, $p=.228$, $d=.196$, *ns* and performing at chance on the secondary retention trials (38.75%), $t(39)=1.043$, $p=.3035$, $d=.165$, *ns*. This demonstrates that children, at best, encoded only minimal information about word-referent pairs mapped in the context of strongly-known words.

Low vocabulary group

When performance is broken down further by age group, differences in performance emerge. On Novel RS trials, children in the low vocabulary group chose the novel target item 93.75% of the time, $t(19)=19.7544$, $p<.0001$, $d=4.417$. On Known RS trials, these children still demonstrate an overriding novelty bias, selecting the well-known target only 20% of the time, significantly below chance, $t(19)=-2.2015$, $p=.0403$, $d= -.4923$, and erring to the novel item 80% of the time, $t(19)=7.96$, $p<.0001$, $d= 1.78$. During retention, low vocabulary children selected the target at chance levels, 30% of the time, $t(19)= -.394$, $p=.6978$, $d=-.073$, *ns*. They also selected the target at chance during secondary retention trials (32.5%), $t(19)=1.043$, $p=.305$, $d=.165$, *ns*.[12] This means that the significant retention seen overall is not driven by low vocabulary children. Furthermore, it suggests that using ostensive naming to train children on the labels for new objects in a non-competitive atmosphere, changed the dynamic of the Weak RS trials such that children are no longer able to retain the newly mapped words.

[12] Like previously, only a subset of trials were analyzed. However, if all the trials on retention are analyzed, children are still at chance (37.5%), $t(19)= .56187$, $p=.5808$, $d=-.073$, *ns*.

High vocabulary group

When performance is broken down by age group, children with a high vocabulary performed above chance on Novel RS trials, choosing the target 95% of the time, $t(19)=27.025$, $p<.0001$, $d=6.043$. On Known RS trials, these children also showed above chance performance, choosing the well-known target item 59.2% of the time, $t(39)=3.02$, $p=.0071$, $d=.675$, an improvement from Experiment 3a. This demonstrates that children with higher overall vocabularies are either able to overcome their novelty biases to select the correct referent the majority of the time, or that a higher productive vocabulary combined with a lack of training for specific known exemplars allowed these children to correctly select well-known items.

On the retention trials, high vocabulary children selected the target 42.5% of the time, a level equal to chance, $t(19)= 1.046$, $p=.3089$, $d=.234$, *ns*[13]. Children's errors during retention were equally distributed; they chose the other named item only 32.5% of the time, $t(19)=-.067$, $p=.9475$, $d=-.015$, *ns*. During the secondary retention trials, high vocabulary children were again at chance, selecting the target 45% of the time, $t(19)=1.49$, $p=.1515$, $d=.334$, *ns*. This confirms the conclusion from Experiment 3a and prior work, that high vocabulary children can perform well on referent selection trials, but do not retain newly mapped word-referent links.

[13] Like previously, only a subset of retention trials were analyzed. However, if all the trials on retention are analyzed, children are still at chance (35%), $t(19)= .272$, $p=.788$, $d=.182$, *ns*.

Weakly-known trials

Overall

Prior work suggests that when competing items are less well-known and objects are harder to disambiguate, performance will falter. However, in the current experiment, overall, children were successful at disambiguating a scene with weakly known items and were able to select the target during Novel RS trials at levels significantly greater than chance (84.38%), $t(39)=15.513$, $p<.0001$, $d= 2.453$. This was not significantly different from the results of Experiment 3a; between groups t-test, $t(80)=.0975$, $p=.9226$, *ns*. Performance on the Weak Known RS trials was again poor, despite the change in the training procedure designed to boost knowledge. Overall, children chose the target known item 20.8% of the time, significantly less than chance, $t(39)=-3.35$, $p=.0018$, $d=-.53$. There was no significant difference in overall performance between Experiment 3a and 3b on Known RS trials, $t(80)=1.584$, $p=.1171$, $d=-.35$, *ns.* This suggests that children's novelty bias may still be playing a role in referent selection regardless of how well known the competing items are. Such focus on novelty would predict poor retention akin to Experiments 1 and 2.

However, like the counterintuitive results of Experiment 3a, when children were tested for retention of words mapped in a weakly-known context, they overall succeeded and selected the target 52.5% of the time, $t(39)= 2.646$, $p=.0117$, $d=.418$[14]. The children here also chose the other named item on referent selection trials 26.25% of the time, a level no different from chance, $t(39)=-1.258$, $p=.216$, *ns*, and selected the target on

[14] Like previously, a subset of trials which the child had previously mapped were analyzed. However, if all the trials on retention were analyzed, children are still significantly above chance, choosing the target 46.25% of the time, $t(39)= 2.296$, $p=.0271$, $d=.363$.

secondary retention trials at chance levels (38.75%), $t(39)=1.043$, $p=.304$, *ns*. Thus, the results at this point are in line with those of Experiment 3a, showing the surprising ability of children to retain newly mapped words learned in highly ambiguous contexts.

Low vocabulary group

When results are analyzed according to a median split of vocabulary, children in the low vocabulary group perform similarly to the low vocabulary children from Experiment 3a. On the Novel RS trials, children in the low vocabulary group selected the target 81.25% of the time, $t(19)=8.93$, $p<.0001$, $d=1.997$. During Known RS trials, these same children perform much more poorly, choosing the target at almost floor levels (9.17%), significantly below chance, $t(19)=-8.24$, $p<.0001$, $d=-1.843$. This suggests not that children are just choosing randomly, but that they are actively ignoring the newly learned known target in favor of selecting the novel foil, which they choose 84.5% of the time, $t(19)=15.02$, $p<.0001$, $d=3.36$. The overriding novelty bias here would suggest that children would not be encoding enough lexical information during referent selection to retain new word-referent pairs. This is confirmed for children with a low vocabulary who choose the target on retention 42.5% of the time, a level no different from chance, $t(19)=1.045$, $p=.3089$, $d=.126$, *ns*.[15] There was no preference for the other named item during retention trials either with children choosing it 47.5% of the time, $t(19)=1.708$, $p=.1039$, $d=.382$, *ns*. Children also choose the target on secondary retention trials at levels no different from chance (30%), $t(19)=-.327$, $p= .747$, $d=-.073$, *ns*. These results differ from the general trend in Experiment 3a, suggesting that differences in the training

[15] Analysis was done over a subset of trials which the child had previously mapped. However, if the analysis is done over all retention trials, results are still at chance (37.5%), $t(19)=.5619$, $p=.5808$, $d=-.088$, *ns*.

protocol between the two experiments had cascading effects on children's ability to learn and retain new word-referent pairs during referent selection.

High vocabulary group

Children in the high vocabulary group were, unsurprisingly, also very good at Novel RS. On these trials, children chose the correct target 87.5% of the time, significantly above chance, $t(19)=14.17$, $p<.0001$, $d=3.167$. However, on Known RS trials, children with a high vocabulary were now no longer significantly below chance, but at chance for choosing the correct newly-learned known item (32.5%), $t(19)=-.099$, $p=.9282$, *ns*. This was not significantly different from high vocabulary children in Experiment 3a, $t(76)=1.60$, $p=.1145$, *ns.*

If retention is broken down further by vocabulary group, we see that the significance overall is being driven by high vocabulary children who choose the correct target 62.5% of the time, $t(19)= 2.467$, $p=.0272$, $d=.579$[16]. Here we also see chance performance on secondary retention trials (35%), $t(39)=.2723$, $p=.7883$, $d=.061$, *ns*. These results suggest that overall vocabulary matters for retention, but only when novel words are mapped initially in the context of weakly-known competitors. It is likely that the amount of competition during these referent selection trials played a critical role, an idea elaborated on in the discussion.

Simulation 4b

The results of Experiment 3b confirm that relative knowledge and familiarity for objects is critical for referent selection and retention. However, the empirical results

---

[16] Results were again analyzed across a subset of trials, but remain significantly higher than chance if all trials are analyzed (55%), $t(19)= 2.73$, $p=.013$, $d=.613$.

suggest differences in learning based on the ostensive definition paradigm used for training. This difference led to changes in referent selection and later retention. Because the model can give more fine-grained analysis of how strong or well-known words are, it was again used to simulate children's performance in the task. Specifically, the training was updated to include ostensive definition without competitors and then the model was tested on the resulting strength of knowledge for the just-trained words. Referent selection and retention using strongly- versus weakly-known words were again tested.

Method

The model is run identically to that in Experiment 3a except that during training of the to-be-named items, only a single word and a single object are active. The model systematically cycled through each to-be-named item one at a time, then continued with referent selection. Architecture and computations are identical to 3a.

Results and discussion

Results from the simulations are depicted in Figure A18. The model's performance on Simulation 4b is similar to that of Simulation 4a. One difference, however, is that here, the model is able to learn the newly trained items enough to demonstrate correct selection on a 3AFC comprehension test. This is evidence that the change in training does indeed result in subtle changes in knowledge for the new items.

Like previously, low vocabulary in the model is defined as when the model has roughly 30% of its total vocabulary, or approximately epochs 7,000-24,000 in the current simulation (Figure A19a). High vocabulary is when the model has accumulated over 90% of its total vocabulary, epochs 48,000-80,000 (see Figure A19b). On Strong Known RS trials, the model initially performs barely above chance, but performance rises as vocabulary increases, similar to the pattern seen in children in Experiment 3b. Like all previous models, on Strong Novel RS trials, the model is still at ceiling across all vocabulary levels.

Performance on weak Known RS is slightly higher here than it was in Simulation 4a. Now, instead of performing below chance, the model hovers at chance. In addition, the ostensive naming training without competitors in Simulation 4b resulted in higher performance on Weak Novel RS. This brings Weak Novel RS closer to the ceiling performance of the Strong Novel RS trials, suggesting that sequential ostensive naming in the model results in relatively stronger knowledge for the newly-trained/weakly-known items.

Finally, and most importantly, Simulation 4b results in slight differences in retention for words mapped in a strong context versus a weak context. As vocabulary increased, retention performance began to diverge such that the model demonstrated higher retention for words mapped during Weak Novel RS than for words mapped during Strong Novel RS. Though the increased performance on Weak Novel RS could be responsible for the change in retention, the model does not perform above chance for the novel words mapped during Strong Novel RS trials, which it mapped at ceiling levels. This suggests that it is not performance on RS alone that accounts for the rise in retention.

A further analysis of the connection weights of strong known words, novel mapped words, and weak known words further supports the difference in strong versus weak-mapping contexts (Figure A20). Like Simulation 3a, the model here has the largest change in weight connections (i.e. the most learning in a single trial) for weakly-known words after a Strong RS trial, followed by novel words mapped on Weak RS trials. Weakly known words also have significant change in their connection weights after Weak RS trials, suggesting that when these words were initially trained in an ostensive naming context before test, they were not fully encoded in the lexicon. Learning continued during referent selection when these items were present. Both strong known words and novel words mapped in a strong context fall near floor in weight changes after referent selection, indicating either full consolidation is complete (likely in the case of strong known words) or very little advances in learning in a single trial (for novel words mapped in a strong context).

General discussion

Comparing the empirical results of Experiment 3a and Experiment 3b reveals very few differences in performance, but together, these experiments provide insight to how knowledge influences referent selection and retention. First, there is no evidence in the empirical results that children trained in a competitive environment learned the new words better or more robustly than children trained with ostensive naming of the new objects one at a time. There were the same amount of errors on warm-up/comprehension for the newly-trained/weakly known words, with an average of 2.31 errors/trial and 2.43 errors/trial on competitive training and ostensive naming respectively, $t(80)=.3267$, $p=.7447$, $d=-.0718$, *ns*. Regardless, children with a high vocabulary appeared to perform better on Weak Known RS trials after ostensive naming training. Though not significantly different from children with competitive training, $t(76) = .80$, $p=.426$, *ns* or from chance, these children were now no longer significantly below chance. This lends support to the idea that sequential ostensive naming training on new objects prior to referent selection, conditions children to attend to the novel object, a novelty bias that can be overcome when overall vocabulary knowledge is high enough. These results may also suggest that sequential ostensive naming is better for long-term learning, especially when taken together with the analysis of connection weights, which show incomplete initial learning for words trained with sequential ostensive naming. More likely however, the results are due to the training diminishing children's endogenous novelty bias ever so slightly or a lack of conditioning during training; two possibilities explained further in Chapter 6.

The computational simulations help explain the moment-to-moment changes in children's word-referent associations over learning and referent selection. The model makes two things clear; learning continues incrementally and changes as overall vocabulary rises, and the amount of learning taking place during a single referent selection trial depends on the knowledge the child, or model, has about items present at that moment. In both Simulation 4a and 4b, the model performs almost perfectly on referent selection, unless specific word knowledge is weak or novelty biases are high. Furthermore, more processing and changing of weights during referent selection means more learning takes place in a single instance. This can then predict a higher probability of retention for new word-referent pairs.

Taken together, the results of Experiment 3 lend evidence to the role relative knowledge plays in referent selection and retention. Specifically, Novel RS is equally high regardless of if competitors are well-known or weakly-known, but retention differs among these two competition contexts. Counter-intuitively, an initial referent selection context that consists of less well-known competitors results in higher rates of retention.

That is, when all elements present are relatively unknown to the child, either because they are completely novel or because the child just learned the label for them, the child is able to learn enough to demonstrate retention of the novel word-object pair at a later time. One possible cause for retention, explained further in the following chapter, is the level of competition and cognitive processing required to initially disambiguate an unknown array of items. The less knowledge a child has about a group of items as a whole, the more processing, or changes in connection weights that are needed in order to disambiguate the novel target from the known foil items. The more a child spends processing the competitors as well as the target, the more information or knowledge the child is likely to have encoded during that single fleeting moment. The more encoding that occurs during a single Novel RS trial, the higher the changes in connections are and subsequently, the higher the rates of retention. In this way, the less stark the contrast is between novel and known items, the more learning that can occur. These results also hint at differences between vocabulary levels, but the weak effects here will certainly need to be examined further.

CHAPTER SIX

GENERAL DISCUSSION

Children have traditionally been viewed as amazing word learners. In just six months, from 18- to 24-months-of age, they acquire an average of over two hundred new words. This feat is especially remarkable given the complexity and ambiguity involved in the acquisition of even a single new word. In particular, research has been concerned with how children solve the problem of referential ambiguity and map a single word to its correct referent. Prior work has shown a clear role for novelty and knowledge in children's ability to solve the problem of referential ambiguity. The goal of this thesis was to further analyze how children use novelty and knowledge in the moment to find the correct referent for a novel word and how those elements can help children extend learning over time in order to demonstrate retention of newly mapped word-referent pairs. In addition, I explored ways in which novelty and knowledge interact in the process of word learning via a computational model. Here, I review the findings on the role of novelty in referent selection and retention, the role of knowledge in referent selection, limitations of the current thesis, and future directions.

The role of novelty

A novel word or object is simply one which has not yet become known. Thus, a lack of prior knowledge or representation in memory can be used to define a novel word, particularly in the context of children's language development. In this way, novelty is relative to the amount of knowledge for that word or object or other words or objects present. Understanding the role of novelty is now thought to be a critical part of understanding the mechanisms behind word learning (Mather, 2013). Moreover, prior work has demonstrated that children at 10-months-of-age can detect changes in novelty given novel words even though they are far from being proficient at word learning (Mather & Plunkett, 2010a), and novelty detection without words is known to occur even younger (Fantz, 1964). Thus, because word learning begins with identifying a novel word or object, novelty detection has been hypothesized to be an early indicator to more complex cognitive processes, such as language development (Mather & Plunkett, 2010a; Mather & Plunkett, 2012). The current thesis explored the role of novelty in both referent selection and retention across variable ages in development.

Referent selection

Novelty has previously been known to help children attend to the correct object during referent selection (Horst et al, 2011; Mather & Plunkett, 2012; Mervis & Bertrand, 1994). The current thesis supports this idea further, demonstrating that novelty aids in referent selection performance at all ages tested (18-24-months-old), at all vocabulary levels (a productive vocabulary of 0-675 words), across various conditions (novelty- or knowledge-based tasks) and in most contexts (competitors are novel, semi-known, or well-known). In each of these cases, children indiscriminately attend to the most novel object present. On trials in which they are given a novel label, this helps immensely in getting the child to the correct referent. However, on trials when the target is a known item, this bias to attend to the novel item means children actually perform quite poorly. The default to novelty, especially in younger children and those with lower vocabularies (see Experiment 1), even when a focus on novelty is not needed or justified (e.g. when children are asked for a known item), suggests that in some cases, children's endogenous novelty bias does not act as a useful lexical constraint. In fact, children's novelty bias can override the use of lexical principles, such as mutual exclusivity and the use of vocabulary knowledge, during referent selection. In other words, the novelty bias children bring to a task might help them attend to the correct object, but when novelty drives disambiguation, children may not be engaging in the lexical components of the task. In this way, the mechanism underlying behavior may be perceptually driven instead of lexical and such a mechanism does not guarantee that the child encodes any lexically relevant information during referent selection. By attending almost exclusively to the unknown object, children's attention is less likely to be divided to encode the novel label (Sinnett, Costa, & Soto-Faraco, 2006). Moreover, there is some evidence that attending to an item due to novelty (sometimes regarded as a more implicit, perceptual process), draws on a different cognitive system than searching for a referent after being given a given a word, a more explicit processing task (Snyder, et al., 2008; Ackles, 2008; Friedrich, 2011; Batterink & Neville, 2011).

The computational model gives an even more precise explanation of children's use of novelty in referent selection. The model, which typically uses prior knowledge and word-referent associations to build new mappings, can instead select its choice on a given trial based on a novelty bias given to the visual input. That is, novelty and knowledge can be disassociated to some extent in the model. Notably, the model needs an artificial novelty boost in order to simulate the novelty-driven performance of children on a referent selection task. In this way, novelty overrides knowledge in an associative word learning framework. Novelty thus has an influence on behavior that operates in a context that is different than the influence of knowledge. As the model matures and vocabulary rises, object familiarity increases and the novelty bias diminishes. This results in performance akin to 24- and 30-month-old children that draws on prior word-object associations and vocabulary knowledge to drive referent selection behavior.

Retention and learning

Despite novelty driving disambiguation during referent selection, novelty alone does not help children retain newly mapped words. Experiment 1 confirmed that novelty drives behavior in even young children, but as the subsequent results of Experiment 2 illustrate, when the task caters to children's novelty bias and their natural tendency to attend to the novel item, children still cannot encode enough information to demonstrate retention. The failure to retain could be due to a failure to encode the novel word-form or a failure to robustly associate the word and referent. If a child has a high overall vocabulary, the results suggest they might be starting to overcome this failure and make some advances toward retention.

The model can begin to differentiate why children fail to retain and give insight to the changes in word-referent associations over the course of the trials. An analysis of the weight changes following novel word mapping in the referent selection trials of Experiment 2 showed very little change and thus, very little learning after a single referent selection trial. These results suggest that children have difficulty retaining newly mapped words in Experiment 2 due to a lack of forming a robust word-referent association or pruning of irrelevant links during the single referent selection trial. However, unlike 24-month-old children, when the model has learned most of its vocabulary, it retains newly mapped words. This is due to the model having already pruned away many irrelevant connections, making the encoding of new word-referent pairs more efficient and robust. A closer look at the behavioral results shows a trend toward retention as vocabulary rises; low vocabulary children chose the target on retention 30.7% of the time, compared to high vocabulary children who selected the target on retention 45.2% of the time, marginally above chance. The model's significantly high retention rate is likely indicative of this trend given that the mapping of the model's vocabulary to children is only an approximation (see the limitations sections for an elaboration on the imperfections of the model's vocabulary). As prior work has demonstrated, increases in overall vocabulary knowledge can bootstrap retention (Bion et al, 2013), potentially predicting that 30-month-old children with a high vocabulary would be able to retain words mapped in a novelty-based context.

In addition, a separate set of pilot simulations in the model (not reported in the current thesis) suggest that familiarity or a lack of novelty can help boost retention. In these familiarization simulations, the novel objects appeared as competitors during initial vocabulary acquisition at a higher rate than in the current thesis (increased from 1% to 17.5%). This resulted in a rise in retention as vocabulary increased. This is consistent with prior empirical work as well (Kucker & Samuelson, 2012) and demonstrates how knowledge is a slow increase in familiarity with initially novel items.

Taken together, the empirical work and associative model suggest that relatively high novelty overrides knowledge-based selection until the child or model acquires a robust vocabulary. Only when overall productive vocabulary is high, can children's endogenous bias to novelty be overridden to support robust referent selection and even retention.

The role of knowledge

Vocabulary knowledge has previously been thought to help in referent selection and retention in two ways; competition between specific known items present (Horst et al., 2010; Markman & Wachtel, 1988; Markman et al., 2003; Merriman & Bowman, 1989) and overall vocabulary knowledge (Bion et al, 2013; Mervis & Bertrand, 1994). In the former case, children use their knowledge about labels for the other presented items to reject those items as possible referents for a novel word (Halberda, 2003; Heibeck & Markman, 1987). In the latter, children come to a word learning task able to process words more efficiently given a strong lexical foundation of what objects and categories words can refer to (Bion et al, 2013; Fernald et al., 1998). In both cases, children draw on their prior lexical knowledge as part of the process of adding a new word to their vocabulary. Prior work would suggest that stronger knowledge in terms of either specific word knowledge or overall vocabulary would support referent selection and retention. The results of the present thesis, however, suggest a more complicated story of how knowledge influences word learning and proposes that knowledge is not as easily defined or measured as we think. The process of drawing on knowledge is inherent in referent selection, but performance on known referent selection trials in all three experiments was low, likely due to interference from novelty. Despite this poor performance, the type of competition was critical for later retention, especially when coupled with high overall vocabulary knowledge.

Referent selection

During referent selection, children are presented with an array of items that often span from some well-known referents to unfamiliar novel items. In order to select a single item to the exclusion of the others, there must be some level of competition either between the items themselves or competition between the items present and the child's current vocabulary. We already know that competition between known and novel words helps learning (Gupta, 2008; Horst et al., 2010; Zosh et al., 2013). The current thesis expanded these competitive processes by varying the strength of knowledge children had for the competing items on any given trial and testing children across a wide range of overall productive vocabularies (Experiment 3).

First, variability in the strength of specific word knowledge during a referent selection task was analyzed. Prior work suggested that the less a child knows about all items present, the harder it is to disambiguate a scene and select a novel referent (Merriman & Schuster, 1991). But, if a child can easily detect the familiar items as name-known, finding the sole unnamed item is relatively easy (Carey, 1978; Hidaka & Smith, 2010). However, contrary to these predictions, the current thesis demonstrates that all children from 18- to 24-months-old are equally good at novel referent selection regardless of their knowledge of the competing foil objects. When the competing items are well-known items, such as a dog or shoe (Experiment 1 and 3), children select the novel target in response to a novel label. When competing items are weakly-known items, such as a just-trained item (Experiment 3) or a pre-familiarized item (Experiment 2), children are also very good at selecting the novel item in response to a novel label. Thus, it does not initially appear that the strength of children's knowledge about the specific foils on a trial matters for selecting the novel referent in response to a novel word.

However, performance on known referent selection trials in which children are asked for a known item by name, suggests a different conclusion. Overall, children show a novelty bias, selecting the novel foil over the correct known referent in most cases (but see high vocabulary children in Experiment 3b). When name-known items are less well-known (the weakly-known items in Experiment 3), the detriment to known referent selection performance is greatest. Specifically, when asked to find a newly-learned item by name, children often err and choose the unnamed novel object instead (Experiment 3). If children are instead asked to find a well-known item by name, they sometimes show slight improvement (Experiment 3b), yet still err and select the novel object the majority of the time (Experiment 1). Thus, consistent behavior on referent selection across variable knowledge levels appears to be due to children's prevailing default to novelty and not necessarily changes in specific word knowledge.

This null effect on Known RS trials diverges sharply from previous work that has demonstrated above chance performance on Known RS for children around 24-months-of-age, who presumably know competing items well (Bion et al, 2013; Horst & Samuelson, 2008; Kucker & Samuelson, 2012). As discussed in Chapter 5, the results of the current studies are possibly due to the specific training protocol used that pre-familiarized (and potentially habituated) children with well-known items prior to test (Experiment 3a). If training on already known words is eliminated as it was in Experiment 3b, and children are split according to high or low overall productive vocabulary, then performance on Known RS trials aligns with that of prior research for 24-month-old children. The fact that the high vocabulary group showed strong performance on the Known RS trials of Experiment 3b suggests that overall productive vocabulary knowledge plays a role in referent selection abilities after all.

The idea that overall productive vocabulary influences referent selection is well supported in the literature (Bion et al, 2013; Mervis & Bertrand, 1994). Children in the high vocabulary group of Experiment 3b demonstrated their prior specific word knowledge and selected the correct known item during referent selection, but only if the referent was well-known. In all other cases, when children's overall vocabulary level is low (Experiment 1 and Experiment 3), or when children are trained on specific known exemplars prior to test (Experiment 3a), they fail to select a known referent when there is a novel foil also present. If the novel foil is eliminated, children do not show this novelty bias and can demonstrate their knowledge about items in the task, such as in the Known RS trials of Experiment 2 where only known items were present. When specific knowledge for competing items is weaker and overall vocabulary levels lower, novelty appears to play a more prominent role in referent selection. Accordingly, the novelty-biased behavior when competitors are weakly-known in Experiment 3 would suggest that children are again not robustly encoding the new word-referent pair in a single referent selection trial and should fail to retain. This hypothesis, however, was not supported.

Retention and learning

Children's novelty-driven performance on referent selection would predict dismal retention. Both Experiment 3a and 3b, however, yield the counter-intuitive result that words mapped in the context of weakly-known competitors result in higher retention, especially when supported by a high overall productive vocabulary. These weakly known items were words that children demonstrated they had minimally learned as the majority of children still succeed on the warm-up/comprehension trials. But, they were also words on which children overall made significantly more errors than they did on warm-up/comprehension trials for strongly-known words. It is likely then, that children's processing of these weakly known words was not as efficient or quick as prior well-known strong words (Fernald et al., 1998). When combined with what we know about competition between known and novel items, this suggests that disambiguating a novel object from an array of weakly-known items is much harder than selecting a novel from well-known items or using novelty alone to disambiguate. This means that when presented with an ambiguous word learning scenario where all things are more weakly known, children would have to do more "work" find the novel referent. Though children's behavioral choices during referent selection did not reveal differences between novel referent selection of strongly-known and weakly-known items, the simulations of the model did suggest such differences. Here, the model shows significantly higher changes in connection weights after referent selection only for weakly-known competitors or weakly-known words, but not for strongly-known words or strongly-known competitors. That is, when a word is already well-known, minimal strengthening of connections or pruning of irrelevant connections occurs in a single trial and the model can even prune away these competing known items quickly when given a novel label. However, when words or objects are weakly known, the model has to cycle more to eliminate competitors and arrive at a choice. This indicates the model is learning more in a single trial when the competitors are weakly known.

It could be the case then, that more processing is required of children when competitors are weakly-known in order for them to correctly select the referent for a novel word. The competitive process that occurs in-the-moment of referent selection would not have sufficient prior knowledge to readily bootstrap for quick selection. Thus, in order to disambiguate the objects and make a choice in this more ambiguous environment, children would be forced to attend to every object present and process more of the information given. Extra processing could then result in more encoding during a single referent selection trial, aiding in later retention. This idea is similar to choosing a single fork from an array of silverware, or choosing a single fork from an array of plates. When presented with a soup spoon, a spork, a fork, and a small salad fork and asked to select the fork, a more fine-grained analysis of all items, including the target, would be required. Attention to the specific features of each of the items might be necessary in order to determine which object is the one with the longer tines, straight sides, and longer handle unique to the dinner fork (Desimone & Duncan, 1995). However, when there is only one fork next to a row of plates and bowls, a very quick glance at all items present will easily reveal which item is the fork – the smaller flatter of the objects present – without needing to encode or notice the specific features of any item present. In this way, the more similar all objects are (either perceptual or conceptual space) or less familiar they are, the more processing that is necessary to even succeed at disambiguation. More processing could mean there is more information gained during initial word-object mapping, aiding in the robustness of initial encoding and later retention. Though this explanation might suggest retention in Experiment 2 in which all items were less familiar, Experiment 2 differed critically in that children had no prior labels for competing items. Because success on the task requires children to integrate a new word into their lexicon, novelty-driven disambiguation without the ability to use the lexicon is unlikely to lead to learning unless other lexical support, such as a high overall vocabulary, is available. To this end, Simulation 3 suggested that in novelty-driven referent selection, retention might occur when coupled with high overall vocabulary. High vocabulary children are those that have successfully mapped and encoded many words previously, often in many different situations and context. The scenario of comparing multiple weakly-known items is very much in line with children's everyday encounters as well as children are learning the names for hundreds of words simultaneously, presumably having many weakly-known referents present at any single word learning exposure.

This explanation is akin to other work on word learning for weakly-known non-solid substances. In this line of work, a dynamic field model suggests that when words are learned in a vocabulary domain of less-well known words, new word-object pairs may be mapped more robustly and thus, retained over time (Kucker, Samuelson, & Spencer, 2012). In that model, retention is due to less competition in the nearby prior vocabulary rather than the competition present in-the-moment. Regardless, the role of competition found in both the prior work and the current thesis support the associative model used in the current thesis and propose that both competition within the known vocabulary and competition in-the-moment matter. Further work will need to further explore the interaction of competitive environment and overall vocabulary levels, however.

Theoretical Implications

One contribution of this work is to highlight the continuity of novelty and knowledge. In particular, we have repeatedly seen in these experiments that minimal familiarity with items can be a detriment to learning and increasing familiarity might lead to better word learning. This is consistent with the slow associative account of word learning (McMurray, et al., 2012; see also Yu & Smith, 2007). What is more critical to learning is prior vocabulary knowledge, which can overcome novelty biases and help to bootstrap retention of new words.

Despite the positive results of the current thesis in demonstrating retention in knowledge-driven context, the results also confirm that knowledge is relative and not always easily defined. The discrepancy between children's performance on all known item warm-up trials and their performance on referent selection when a novel foil item was present across each of the experiments suggests that vocabulary knowledge between 18- and 24- months-old, even for well-known items, is weak and susceptible to interference. The high selection of novel items on Novel RS and subsequent poor retention also demonstrates the continual nature of knowledge. Vocabulary knowledge is not a set construct that can be measured equivalently across context, but rather, our understanding of a child's knowledge is limited to a single point in time in which they behave in a particular task in a specific context. Thus, even evidence of retention in the tasks here is not hard evidence for learning, but instead a suggestion that knowledge is slowly being gained over development.

Limitations

Though this thesis adds to the current literature and our understanding of how novelty and vocabulary knowledge influence word learning, there are still limitations to the current work. Word learning is complex, involving multiple cognitive processes, lexical principles, and pragmatic cues. The current thesis explored just two of many factors known to influence word learning. Though novelty and knowledge are thought to be particularly important in young children's referent selection and retention, and there has been a recent push to explore these facets further (Mather, 2013), there is also a strong history of examining the influence of social and pragmatic processes on word learning. Hebbian-based models are well-regarded in the field of development, but the simple associative model used here cannot account for these pragmatic cues. This particular model also does not yet include components of similarity or categorization, which could be relevant in future extensions of the current work.

The model is also limited in its estimate of productive vocabulary. The model demonstrates two clear plateaus in vocabulary development. The early plateau seen when the model knows roughly 8 of the possible 30 words could be thought to map onto word learning prior to the vocabulary spurt. The second plateau later when the model hits ceiling would then map onto the thought to be post-vocabulary spurt period. However, the model only has a lexicon of 30 words as opposed to the thousand-plus limit of a given child's vocabulary. Thus, by the most conservative estimates, producing just 8 words in the model is almost 30% of the total vocabulary. Children at 18-months of age produce roughly 90 words, which is roughly 12% of the words on the MCDI. Further, it is unclear if the model's total vocabulary is representative of the MCDI, child-sized lexicon, an adult-size lexicon, or none of the above. Thus, dividing the model into pre- and post-spurt is just one of many approaches to defining vocabulary level in the model. Furthermore, we know that there are networks of words and biases in vocabulary that matter (Hills et al., 2007; Samuelson & Smith, 1999), but are not implemented in the model. Thus, parallels between low and high vocabulary in the model are currently limited to being very general pictures of trends and estimates and are far from estimating individual behavior.

An additional limitation in the current thesis is the task itself. Because of the binomial nature of the empirical task, a child either retains or not. Retention implies the child fully encoded the new word robustly into his or her lexicon in order to maintain it over time and withstand the rigorous test of retention. A failure to retain, however, tells us very little about what the child encoded during the single mapping instance. It could be that the child robustly encoded both word and object, but failed to make a link between the two, or if could be they encoded the referent, but not the word-form, or vice versa. The model and prior work suggest that children do get minimal information about both word and referent from a single word-referent pair, but this knowledge is too weak and minimal to be retained over time or withstand the rigorous test of retention. However, neither previous work nor the models make specific predictions about how the word or referents are encoded independently or how novelty versus knowledge influences less-robust associations.

Along these same lines, the empirical task is also limited in its one-shot approach. Children are given only one chance to initially map a word and object and only a single chance to demonstrate retention. Further, children are burdened with learning up to eight new words (11 in the case of Experiment 3) within the span of roughly fifteen minutes. Multiple repetitions of trials would avoid this single-shot approach, but would also confound results and lead to inflated performance as trials repeated (see Mather & Plunkett, 2009). Thus, the current results could very well be driven by initial reactions and primarily perceptual responses and less of the underlying lexical processes. However, such initial biases are clearly important to take into account for lexical development, especially in the case where novelty may prevent initial encoding and thus, protract the learning process. These biases are thus directly in line with some of the goals of this thesis. Important as well is that the model is not all or nothing, but instead behaves based on the slower building of associations and probabilistic qualities of connections between words and referents. The model can then fill in details about children's behaviors during referent selection and retention.

Future directions

The results of the current thesis provide an interesting story of the interaction of novelty and knowledge in word learning. However, these topics are relatively new to the field of word learning and thus, there is much that future work can explore. For instance, the results of Experiment 1 suggest that children's high novelty bias prevents them from demonstrating knowledge they already have about words and referents in the world. By diminishing this novelty bias, it could be possible to boost performance on known referent selection trials. Pre-familiarization with the novel objects is known to boost retention in older 24-month-old children (Kucker & Samuelson, 2012). Pilot work, not presented here, with 18-month-old children in an identical pre-familiarization task used in Kucker & Samuelson (2012) however, still shows poor performance on known referent selection trials (with a 1 minute pre-familiarization: $t(17)=1.538$, $p=.1425$, *ns*; 5 minute pre-familiarization: $t(8)=.8225$, $p=.4346$, *ns*. Critically, pre-familiarization in 18-month-old children also did not boost retention regardless of if children received one minute of familiarization (44%), t(16=1.017, $p=.3242$, *ns*, or up to five minutes (33.3%), $t(9)=.0254$, $p=.9804$, *ns*. Future work might take a note from Bion and colleagues (2013) and Mervis and Bertrand (1994) though and take a longitudinal approach to examining how overall productive vocabulary is related to novelty biases. It could very well be that as novelty biases diminish, vocabulary acquisition takes off and pre-familiarization can aid in retention. Because the current work did not measure children's novelty bias apart from the word learning task, this cannot yet be answered.

In the current thesis, novelty often overrides prior knowledge during referent selection, especially in younger children with lower vocabularies (Experiment 1), in novelty-driven tasks that eliminate name-known competitors (Experiment 2), and when children are trained on specific exemplars of already known items that will later be used in the task (Experiment 3a). However, prior work has suggested that robust category representations of targets and competitors prior to test may aid in referent selection and possibly retention (Mayor & Plunkett, 2010). Thus, a replication of Experiment 3 but with training on multiple exemplars of the objects should alter performance on referent selection and retention. Specifically, we should find an increase in known referent selection and retention performance for words mapped in a well-known/strong context as the objects would be more easily aligned with children's prior knowledge and thus, limit the use of novelty. We should also see equivalent or lower retention for words mapped in a weak context as children's knowledge for the competing items would be stronger and referent selection require less processing. However, children in this latter case would still have relatively less prior specific knowledge to draw on in order to bootstrap the new word-referent pair.

As the results have all suggested, novelty and knowledge interact to influence word learning. The inherent problem with referent selection tasks such as these, however, is that novelty and name-unknown are confounded (Mather, 2013). The unnamed items are almost always more novel than the named. Though Experiment 2 and prior work by Horst and colleagues (2011) removed the use of named items, it is still the case for Experiments 1 and 3 as well as other work using referent selection that the perceptual novelty of items are unequal. One way to begin to divide novelty and naming is to increase the familiarity of the target novel items but never introduce a name. Then, present this highly familiar, but unnamed novel item in the context of weakly known, but named items (such as the competitors in Experiment 3). This paradigm, however, would still leave the big theoretical question of how we might equate familiarity for objects with and without labels. It is possible, that having a name would make an item more salient or familiar despite less time or exposures to it.

Finally, future work will exploit the model further. Specifically, it will be critical to explore how the model can account for the particular differences in retention seen in various age groups in Experiment 3a and 3b. One way to do this might be to alter the inhibition rate, or the degree to which competitors are suppressed during processing, across the layers. Changing the inhibition rate in the lexical layer will simulate changes in how competition generally plays out, whereas changing inhibition in the input layer might simulate how much or little objects compete with each other. In could be that lower inhibition in the input layer simulates the processing of more dissimilar objects that require less competition to settle, whereas higher inhibition simulates very similar objects that must cycle longer in order to settle to a single choice. That is, it is harder to eliminate and prune away irrelevant connections when they are more closely aligned with other items present or equally unknown. In addition, overall vocabulary is inherently critical to referent selection and retention in the model. The model demonstrates retention for weakly known items, but only after a robust productive vocabulary has been acquired. Similarly, with a larger vocabulary, the model will always succeed at known referent selection, though empirically, younger children show retention for weak competitors and even higher vocabulary children fail at known referent selection at times. Future work will also track individual word-referent pairs and changes in connection weights across multiple levels of vocabulary knowledge.

Conclusions

Research over the last few decades has demonstrated that children are remarkable word learners. They quickly and accurately seem to map novel words and referents and build a massive vocabulary in a short period of time. Novelty and prior vocabulary knowledge are known to play critical roles in this process of word learning. The current thesis specifically explored the role of novelty and knowledge in selecting a referent in response to a novel word, and making a robust association between that word and referent over time. The results suggest that in many cases, children's endogenous novelty bias can indeed drive referent selection, but that it may not act as a lexical principle and that it does not aid in retention of new word-referent pairs. Furthermore, when the amount of prior knowledge a child has for items present varies greatly, such as when an unknown item is paired with a well-known item, novelty can override the use of knowledge during disambiguation. However, when disambiguation is more challenging and children must spend more time processing items present, such as when an unknown item is paired with a weakly known item, children do in fact draw on prior knowledge and make an initial robust word-object association. In addition, the results suggests that both increases in referent selection and retention performance increase as overall vocabulary and knowledge increase, potentially due to decreases in endogenous novelty biases. Thus, children can use novelty to learn a new word, but their fragile prior word knowledge is, in fact, most critical for learning.

APPENDIX
TABLES AND FIGURES

Table A1. Input values and parameters for all simulations of the model.

Parameter	symbol	computational role	E1	E2	E3a	E3b
Input units		number of nodes in each input layer	40	40	40	40
Known words/ objects		total number of words to acquire during vocabulary acquisition	30	25	27	27
Trained words/ objects		items held out until training proceeding referent selection	0	0	3	3
Novel words/ objects		unnamed objects present a small amount of time during vocab acquisition	10	10	10	10
Supernovel words/objects		novel objects held out until referent selection	0	5	0	0
Lexical units		number of nodes in lexical/decision layer	500	500	500	500
Weight size		initial random weights of associations	.25	.25	.25	.25
Learning rate	η	rate at which weights increase on each cycle	.0005	.0005	.0005	.0005
Referential ambiguity		average number of competitors during vocabulary acquisition	.5	.5	.5	.5
Novelseen		amount of time novel items are present as foils during vocabulary acquisition	.01	.01	.01	.01
Feed-forward temperature	τ_{ff}	input to lexical layer	.01	.01	.01	.01
Feedback temperature	τ_{fb}	lexical to input layer	.01	.01	.01	.01
Decay		rate at which spurious connections die away	.5	.5	.5	.5
Novelty bias starting value	n_{xy}	boost to un-seen items	2.5	2.5	2.5	2.5
Novelty bias decay rate	δ	rate at which novelty boost decays at each exposure	.01	.01	.01	.01

Table A2. Outline of experiment procedures.

	Experiment 1	Experiment 2	Experiment 3a	Experiment 3b
Silent pre-fam.	0	1-2 min	0	0
Training	0	0	3 known, 3 novel	3 novel
Warm-up	3	3	3 known & 3 new	3 known & 3 new
Known RS	4	8	8	8
Novel RS	4	8	8	8
Break	5 min	5 min	5 min	5 min
Retention	2	4	4	4
Second. Ret.	0	2	2	2
Preference	2, RS/Ret novels	5, Nov. vs. Fam	5, Nov. vs. Fam	5, Nov. vs. Fam.

Note: Certain events occurred in some experiments, but not others. Numbers indicate the number of trials of a given type with a 0 indicating that type did not appear in that experiment.

Table A3. Comparison of empirical results of Experiment 1 with prior work with 24-month-old children and split by vocabulary.

	Experiment 1 Overall 18 mo	Experiment 1 Low vocabulary (<65 words)	Experiment 1 High vocabulary (>64 words)	Kucker & Samuelson (2012) 24 mo	Horst & Samuelson (2008) 24 mo
Known RS	29.9%	30.7%	28.8%	71%***	72%***
Know RS error to novel	70.3%**	69.7%***	71.2%***	28%	----
Novel RS	78.6%***	70.6%	90.4%	90%***	70%**
Retention	32.8%	31.25%	34.6%	45%	34%
Preference to unnamed	54.7%***	57.9%***	50%***	----	----

Note: The first column depicts performance overall in Experiment 1, the second and third columns are the results of Experiment 1 split by median vocabulary into low or high groups, and the fourth and fifth columns indicate prior 24-month-old results. Chance is 33% in all cases

****p*<.001 Note: RS indicates referent selection trials

Source: Kucker, S.C. & Samuelson, L.K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped word. *Infancy*, *17*(3), 295-323.

Source: Horst, J.S., & Samuelson, L.K. (2008). Fast mapping but poor retention. *Infancy*, *13*(2), 128-157.

Table A4. Empirical results of Experiment 2, split by age group and median vocabulary.

	E2 Overall 18-24 mo	E2, 24 mo (17.5-21.5 mo)	E2, 24 mo (21.5-2 5 mo)	E2 Low vocab (<243 words)	E2 High vocab (>242 words)
Known RS	74.1%***	67.8%***	78.6%***	68.7%***	79.8%***
Novel RS	69.7%***	68.9%***	70.2%***	72.9%***	66.3%***
Retention	35.6%	31.9%	49%**	29.2%	45.1%[#]
Retention – error to other named	43.2%**	47.2%**	40.3%	50%***	36.1%
Secondary Ret.	36.6%	39.8%	34.3%	41.7%	31.3%
Novelty Pref.[&]	63.3%**	65.6%[#]	61.7%[#]	63.7%[#]	62.9%[#]

Note: Chance is 33% in all cases except where indicated by [&] where chance is 50%

[#]*p*<.05 **p*<.05 ***p*<.01 ****p*<.0001

Table A5. Empirical results of Experiment 3a, split by age and median vocabulary.

	E3a Overall 18-24 mo	E3a, 18 mo (17.5-21.5 mo)	E3a, 24 mo (21.5-25 mo)	E3a Low vocab (<184 words)	E3a High vocab (>183 words)
Strong Trials					
Known RS	32.5%	30%	33.95%	21.25%	42.8%
Known RS – error to novel	66.27%***	68.33%***	65.12%***	77.5%***	56.06%***
Novel RS	98.8%***	96.67%***	100%***	97.5%***	100%***
Retention	29.8%	23.3%	33.3%	27.5%	31.8%
Retention – error to other named	34.5%	46.67%	27.7%	37.5%	31.8%
Secondary Ret.	32.14%	33.3%	31.48%	32.5%	31.8%
Weak Trials					
Known RS	12.9%***	19.44%	9.26%***	13.33%*	12.5%***
Known RS – error to novel	83.5%***	77.22%***	87.04%***	84.7%***	82.95%***
Novel RS	84.5%***	76.67%***	87.96%***	82.5%***	85.23%***
Retention	63.1%***	83.3%***	51.9%#	65%*	61.36%*
Retention – error to other named	38.1%	26.7%	44.4%	40%	36.4%
Secondary Ret.	32.14%	33.3%	31.48%	32.5%	31.8%
Novelty Pref.&	49.4%	47.4%	48.4%	49.8%	48.9%

Note: Chance is 33% for all trials except where indicated by & where chance is 50%

#*p*<.05 **p*<.01 ***p*<.001 ****p*<.0001

Table A6. Empirical results of Experiment 3b, split by age and median vocabulary.

	E3b Overall 18-24 mo	E3b, 18 mo (17.5-21.5 mo)	E3b, 24 mo (21.5-25 mo)	E3b Low vocab (<141 words)	E3b High vocab (>140 words)
Strong Trials					
Known RS	39.8%	34.33%	48.33%	20%	59.17%*
Known RS – error to novel	60.4%***	65.67%***	51.67%	80%***	40.83%
Novel RS	94.37%***	95%***	93.3%***	93.75%***	95%***
Retention	36.25%	38%	33.33%	30%	42.5%
Retention – error to other named	40%	38%	43.3%	47.5%	32.5%
Secondary Ret.	38.75%	34%	46.67%	32.5%	45%
Weak Trials					
Known RS	20.8%*	16.67%**	27.78%	9.17%***	32.5%
Known RS – error to novel	71.5%***	77.33%***	62.22%*	84.58%***	58.75%**
Novel RS	84.38%***	80%***	91.67%***	81.25%***	87.5%***
Retention	52.5%#	50%	56.67%*	42.5%	62.5%*
Retention – error to other named	26.25%	28%	23.3%	22.5%	30%
Secondary Ret.	32.5%	32%	33.3%	30%	35%
Novelty Pref.&	55.79%	55.8%	55.78%	58.75%	52.83%

Note: Chance is 33% for all trials except where indicated with & where chance is 50%

#$p<.05$ *$p<.01$ **$p<.001$ ***$p<.0001$

Figure A1. General architecture of the Hebbian Normalized Recurrent Network (HRN) model

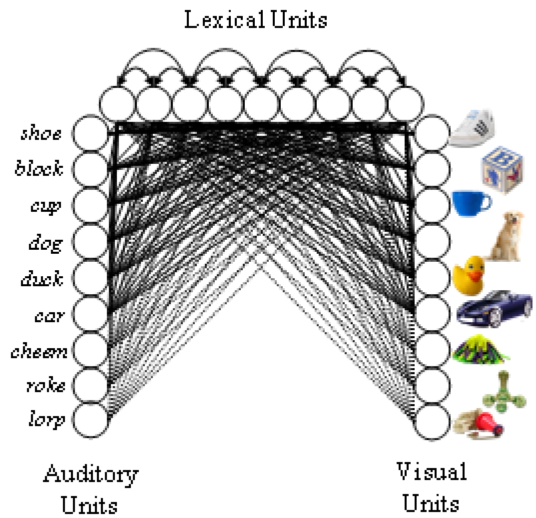


Figure A2. Cycling of the HRN model. An example of input to the model during a known referent trial is presented in panel (a), changes in connections are presented in panel (b). In this example, the word *cup* is presented along with the objects cup, dog, and car. Panel (c) depicts a novel referent selection trial. Connections between the correct word and referent are strengthened, whereas connections for items present without their corresponding label are weakened and items and words present remain unchanged.

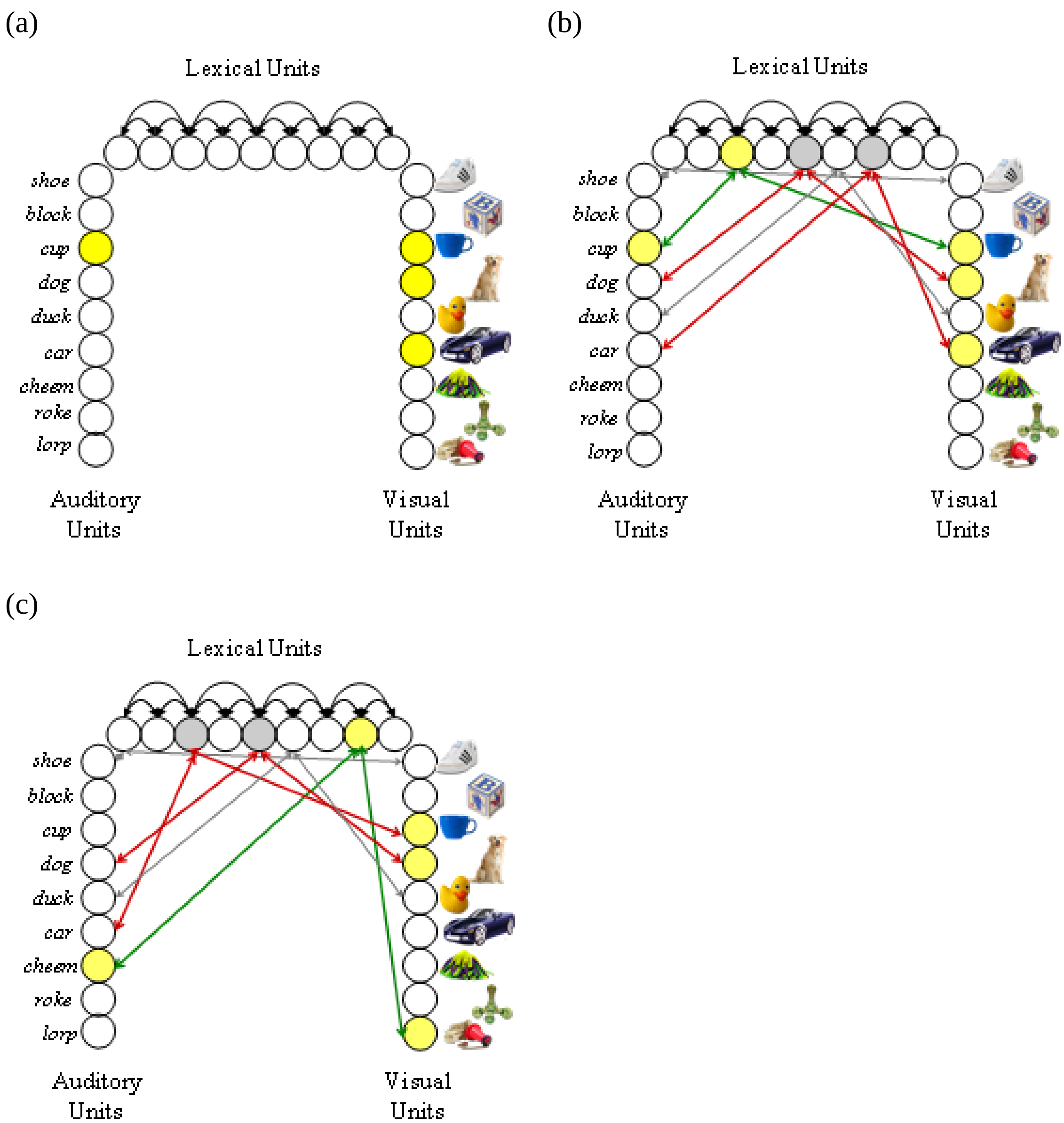


Figure A3. Results of Simulation 1. Performance of the model on referent selection and retention, plotted as percent correct to the target along the left y-axis. Overall productive vocabulary of the model is plotted along the right y-axis.

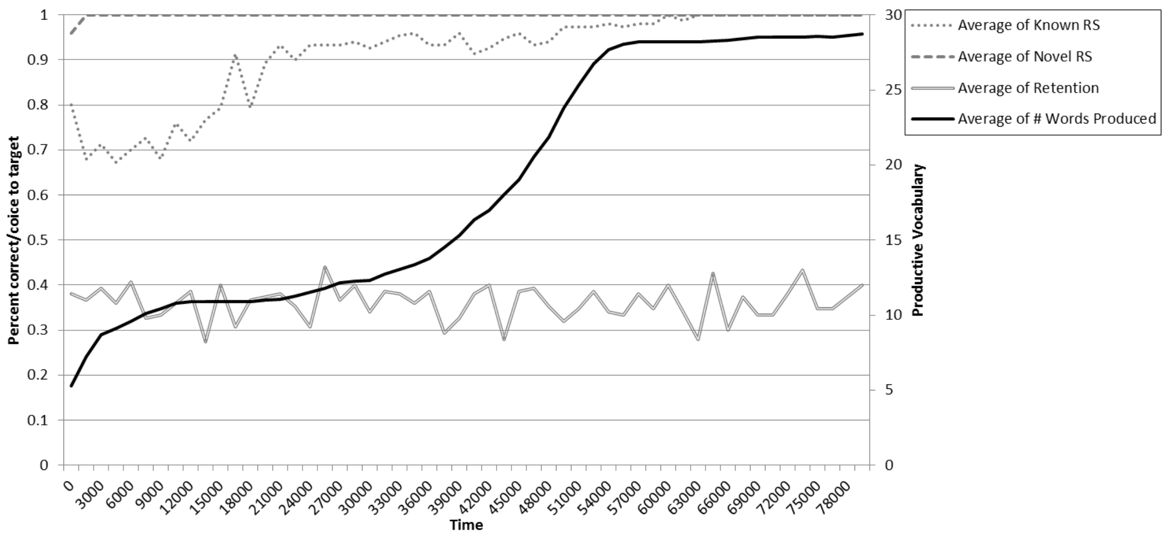


Figure A4. Stimuli used in Experiments 1, 2, 3a, and 3b. Novel items from Experiment 1 are depicted in panel (a), novel items added as supernovel in Experiment 2 are depicted in panel (b), novel items trained as weakly known items in Experiment 3 are depicted in panel (c), and known items are depicted in panel (d).

Figure A5. Referent selection and retention paradigm used in Experiment 1.

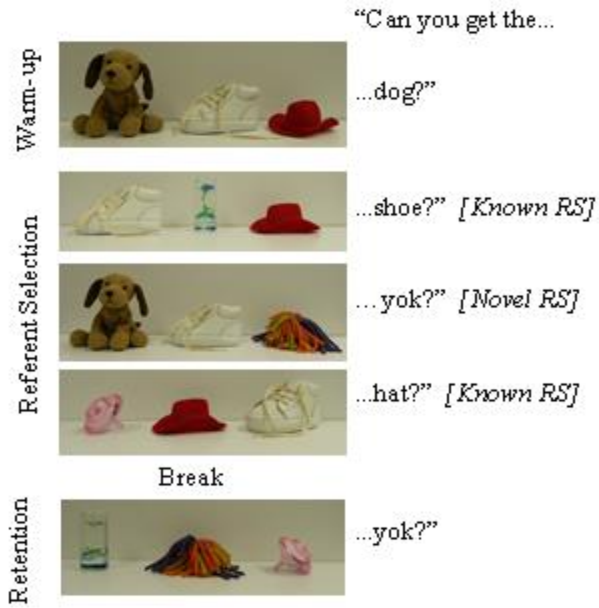


Figure A6. Decay rate of the novelty bias of visual nodes over the course of multiple trials as a function of the number of competitors present on a given trial.

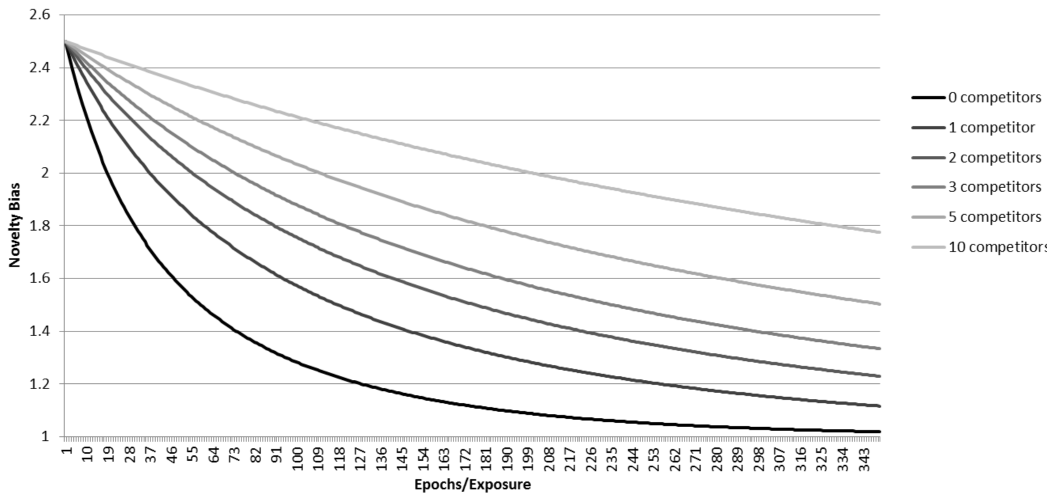
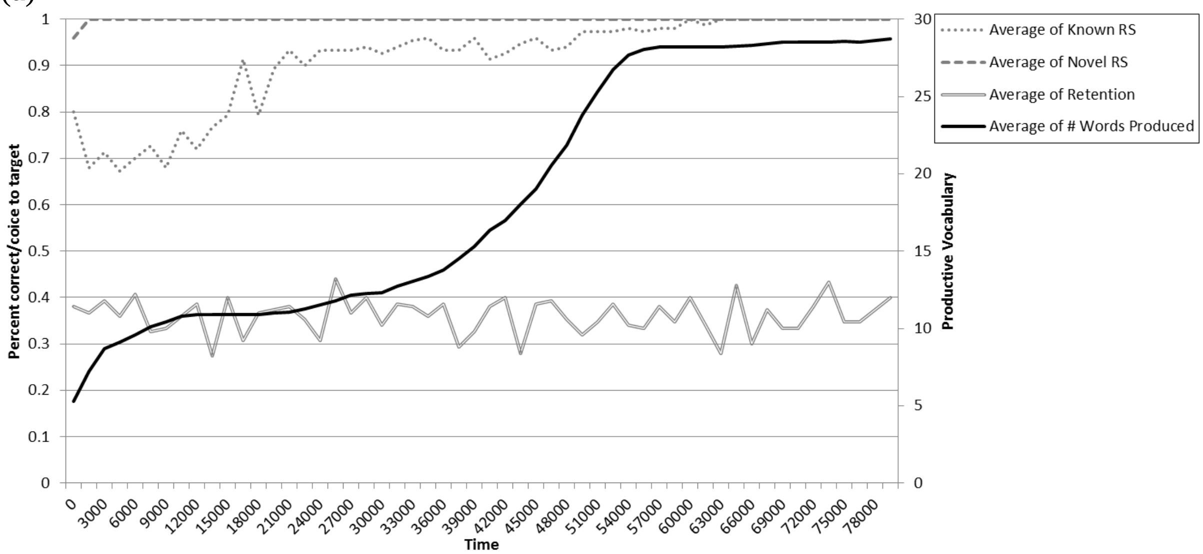


Figure A7. Results of Simulation 2 with a novelty bias compared to the baseline model of Simulation 1. Panel (a) depicts Simulation 1 without a novelty bias, panel (b) shows Simulation 2 with a novelty bias. Performance of the model on referent selection and retention, plotted as percent correct to the target along the left y-axis. Overall productive vocabulary of the model is plotted along the right y-axis.

(a)



(b)

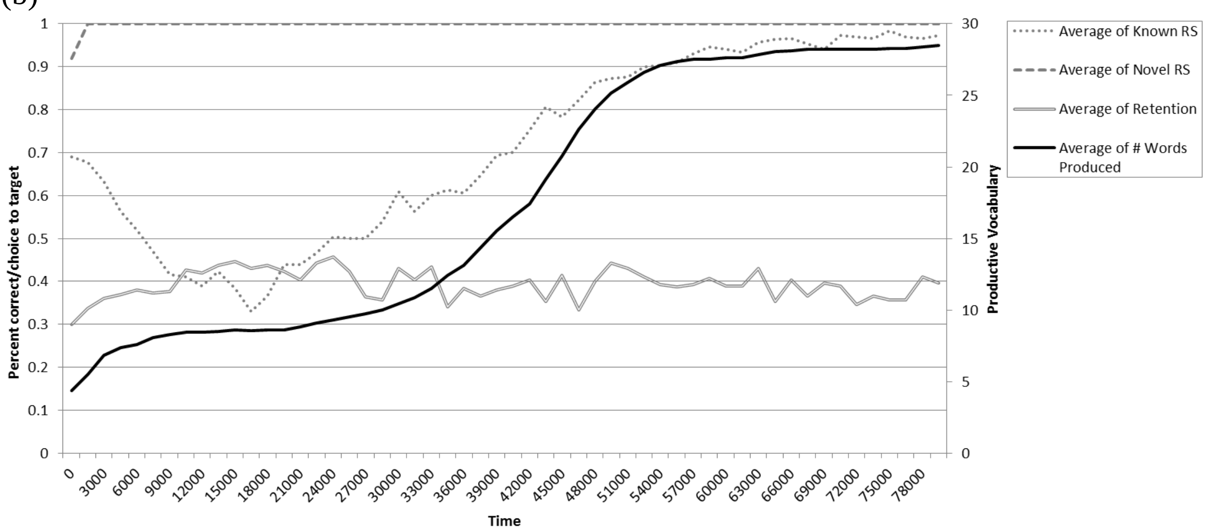
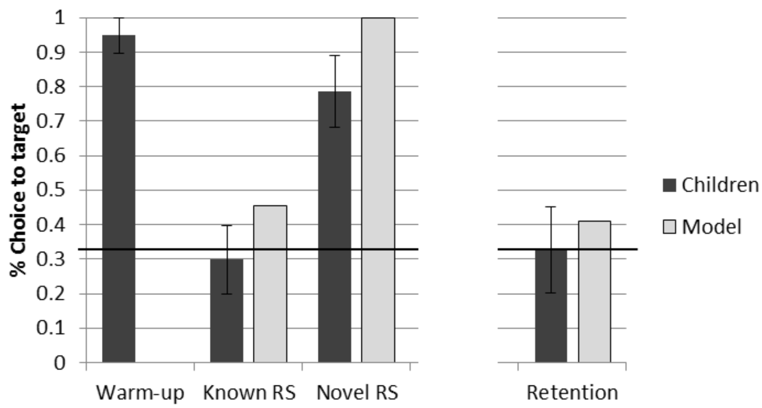
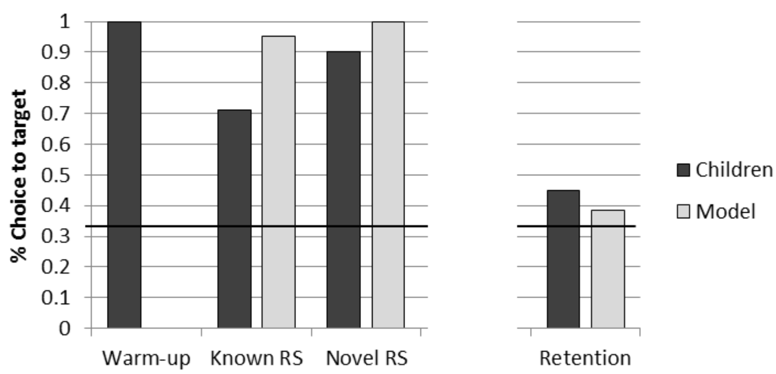


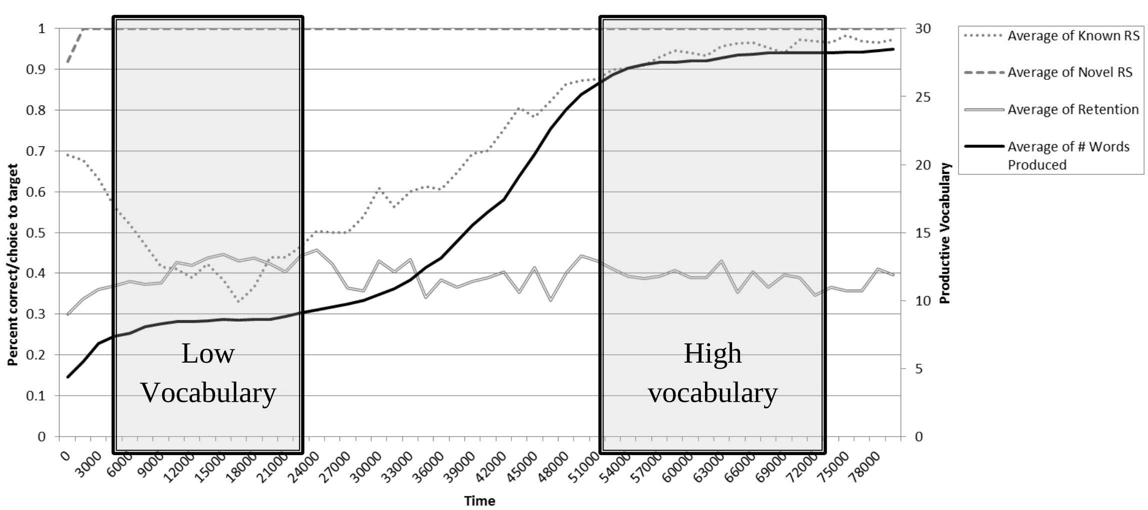
Figure A8. Comparison of referent selection and retention performance of Simulation 2 and low and high vocabulary children. Low vocabulary 18-month-old children in Experiment 1 vs. Simulation 2 are depicted in panel (a) and high vocabulary 24-month-old children in Kucker and Samuelson (2012) vs. the model is depicted in panel (b). Note: Low vocabulary in the model included the epochs of development where the model produced between 8 and 10 words, or the first plateau in vocabulary acquisition. High vocabulary in the model included epochs where the model produced more than 28/30 words after hitting its vocabulary spurt. Panel (c) designates the low and high vocabulary splits according to the epochs and results of the model.



(a)



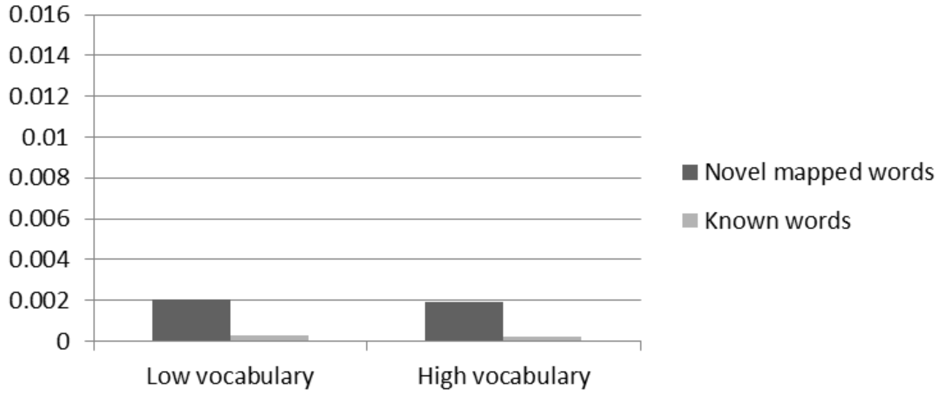
(b)



(c)

Figure A9. Analysis of the change in connection weights in the model after referent selection for low and high vocabulary groups. Simulation 1 without the novelty bias is depicted in panel (a). Simulation 2 with the novelty bias is in panel (b)

(a)



(b)

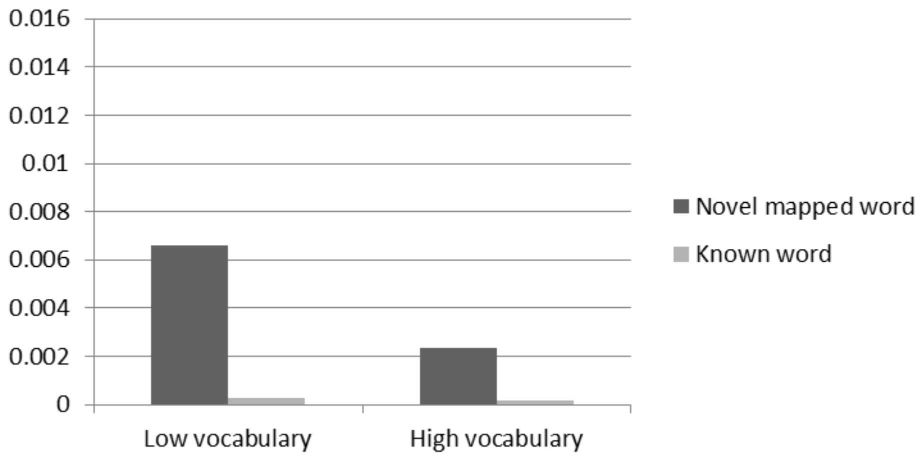


Figure A10. General procedure of Experiment 2

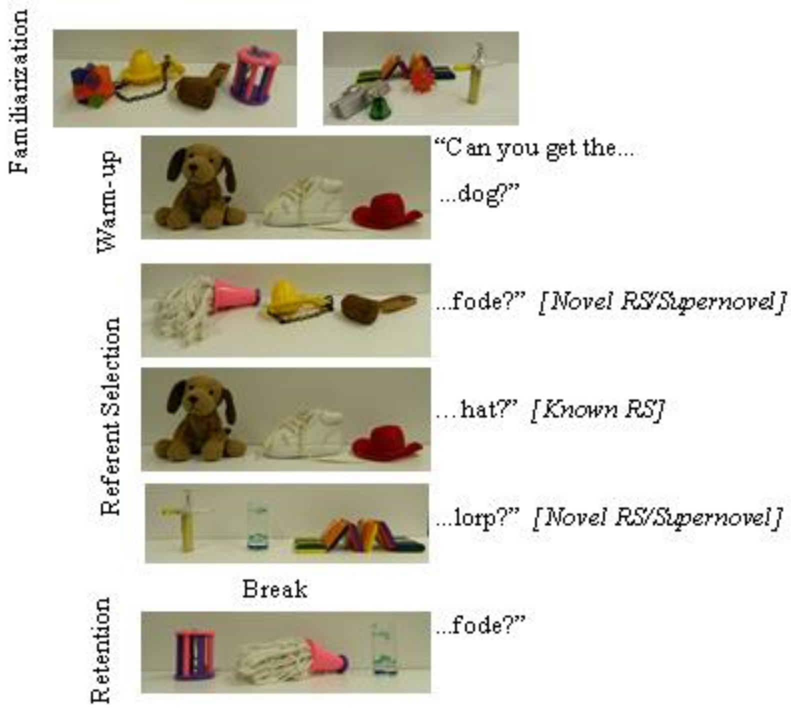


Figure A11. Results of Simulation 3. Performance of the model on referent selection and retention, plotted as percent correct to the target along the left y-axis. Overall productive vocabulary of the model is plotted along the right y-axis.

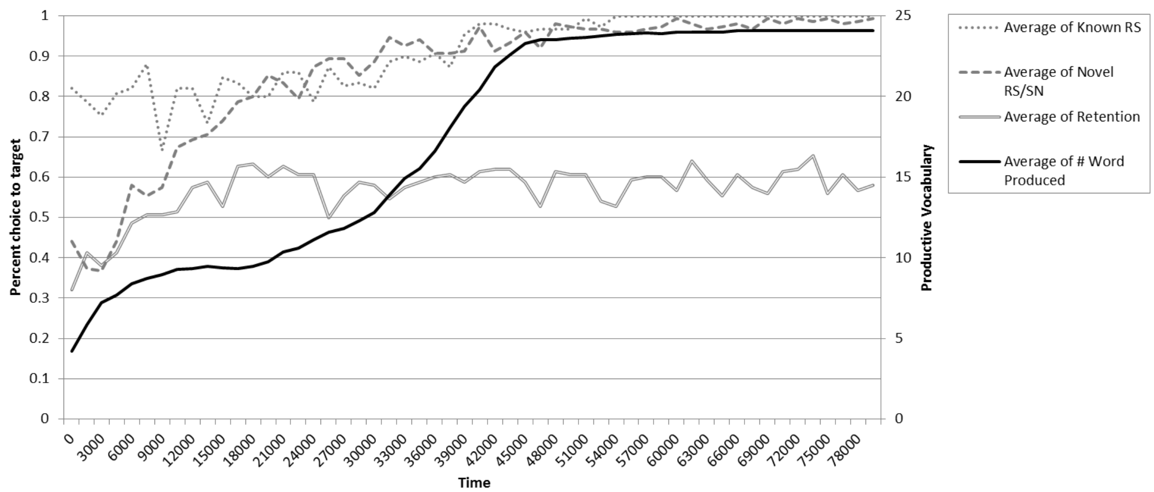
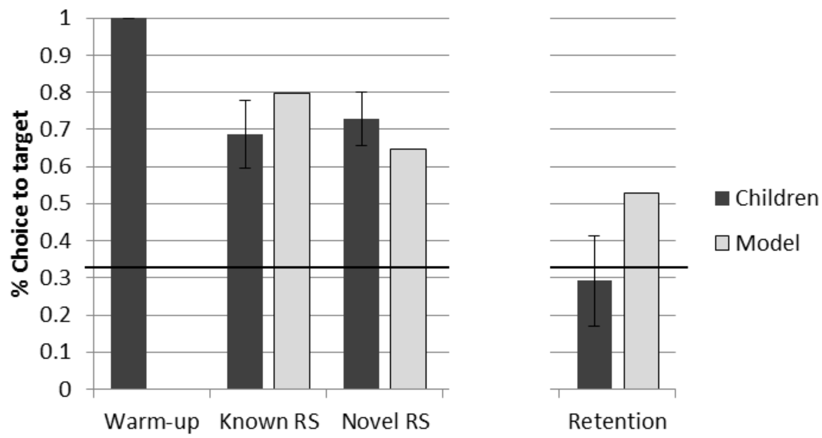
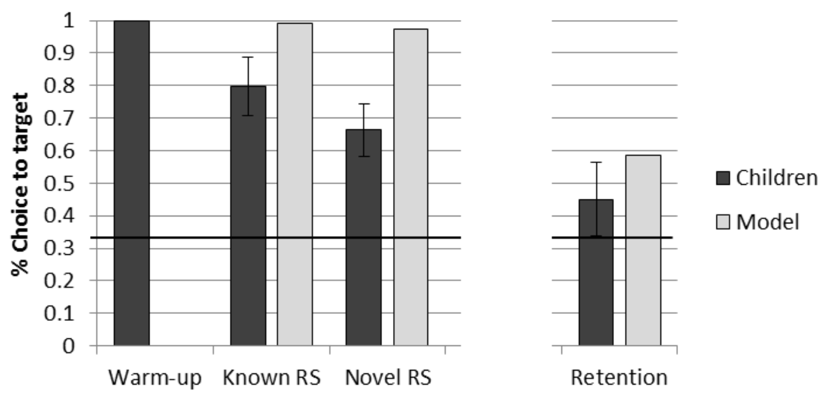


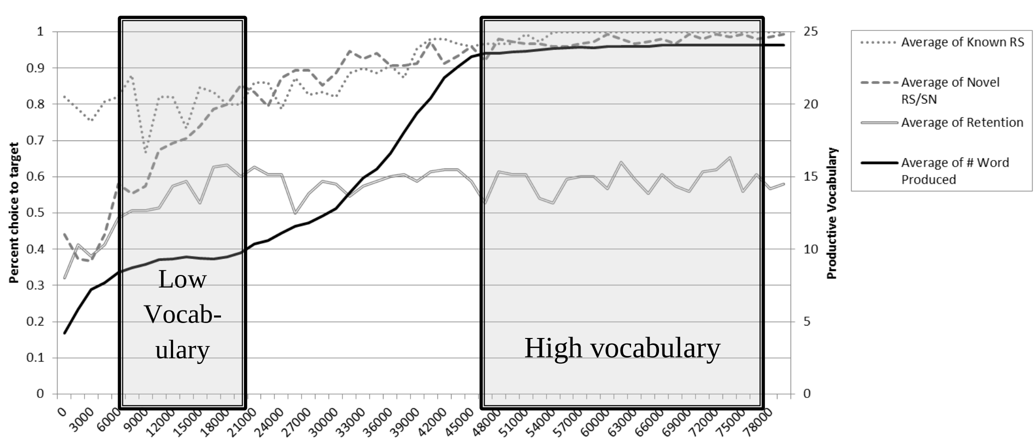
Figure A12. Comparison of referent selection and retention performance in Simulation 3 and low and high vocabulary children in Experiment 2. Note: Low vocabulary children and model results are depicted in panel (a). Low vocabulary in the model included the epochs of development where the model produced between 8 and 10 words, or the first plateau in vocabulary acquisition. High vocabulary is depicted in panel (b) included epochs where the model produced more than 23/25 words after the vocabulary spurt. Panel (c) designates the low and high vocabulary splits according to the epochs and results of the model.



(a)



(b)



(c)

Figure A13. Analysis of the change in connection weights after referent selection in Simulation 3 for low and high vocabulary groups.

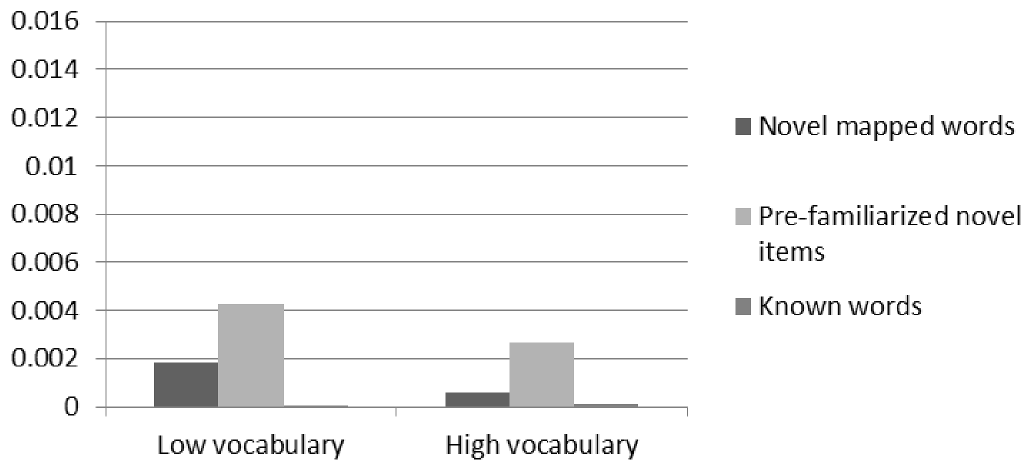


Figure A14. General procedure of Experiment 3. Experiment 3a is depicted in panel (a) and Experiment 3b in pictured in panel (b)

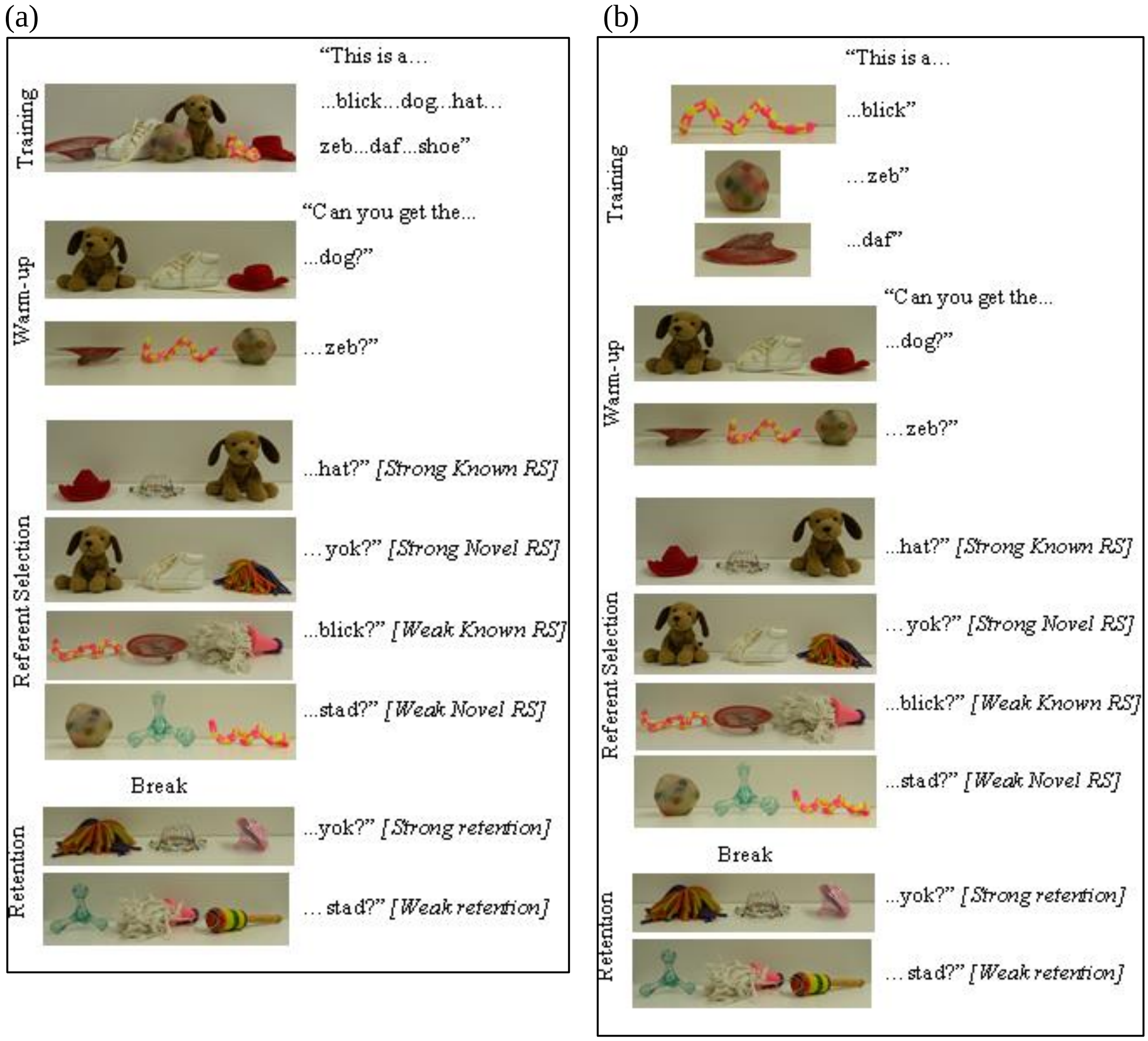
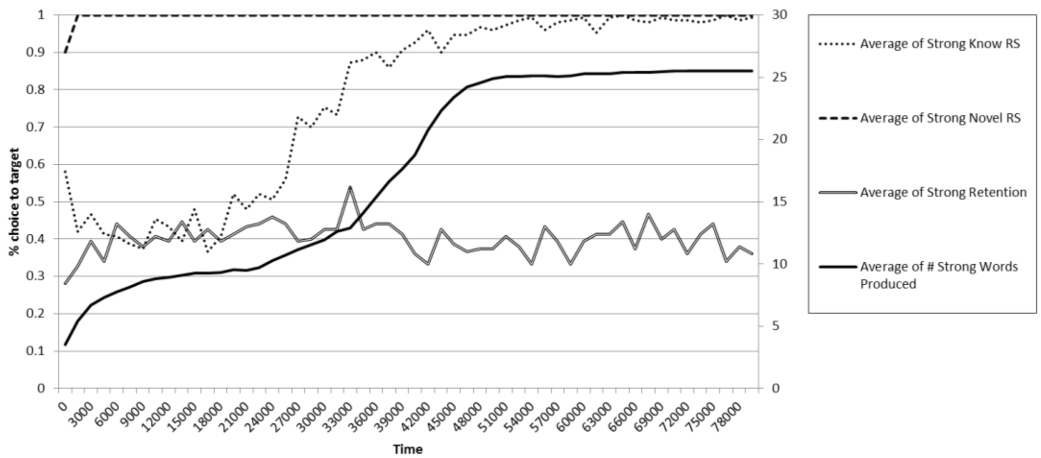


Figure A15. Results of Simulation 4a. Performance of the model on referent selection and retention, plotted as percent correct to the target along the left y-axis. Overall productive vocabulary of the model is plotted along the right y-axis. Panel (a) depicts performance on the Strong RS trials whereas panel (b) depicts Weak RS trials.

(a)



(b)

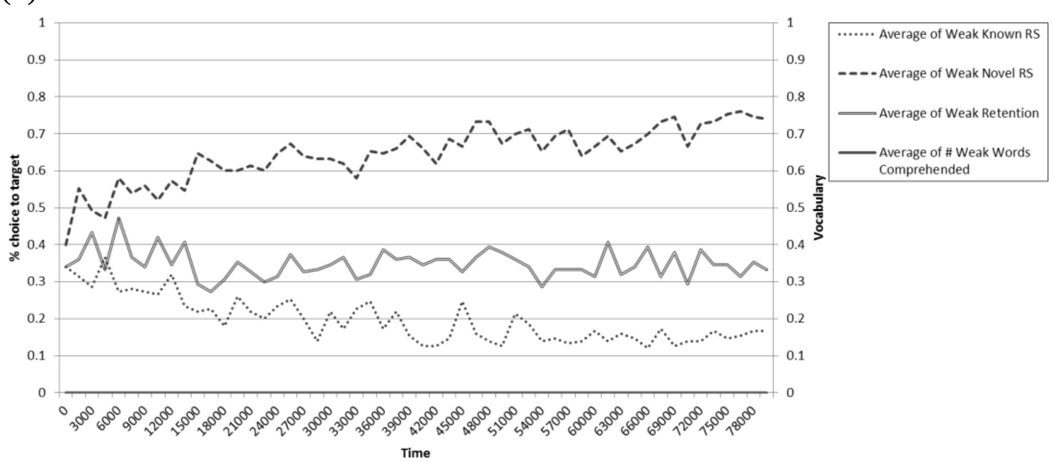
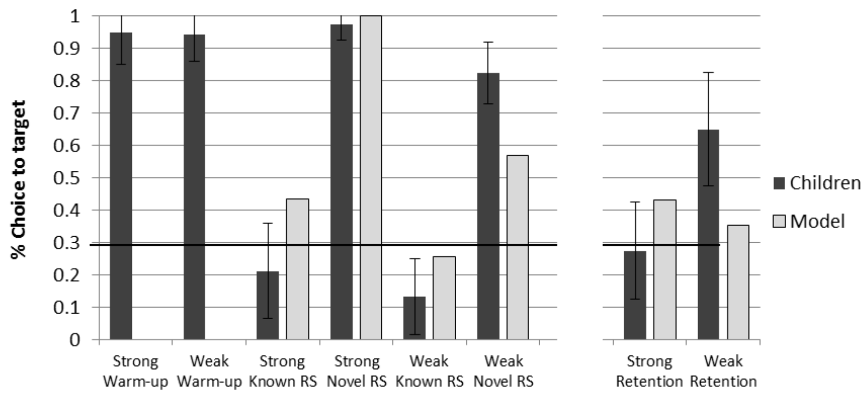
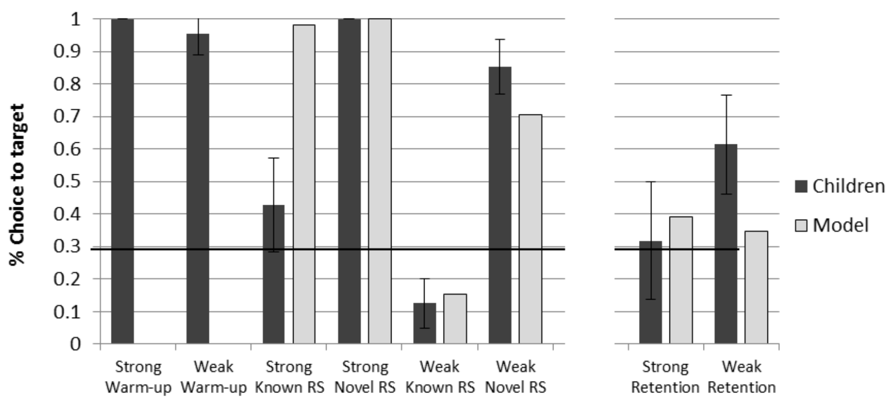


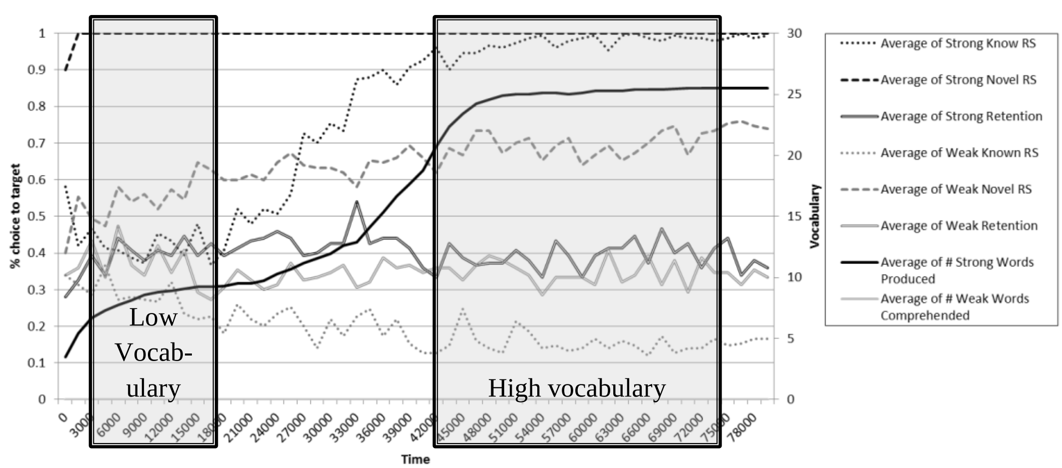
Figure A16. Comparison of referent selection and retention performance in Simulation 4a and Experiment 3a. Low vocabulary children and model results are in panel (a) and high vocabulary children and model are depicted in panel (b). Panel (c) shows the vocabulary groups in the model Note: Low vocabulary in the model included the epochs of development where the model produced between 8 and 10 words, or the first plateau in vocabulary acquisition. High vocabulary included epochs where the model produced more than 25/27 words after the vocabulary spurt.



(a)



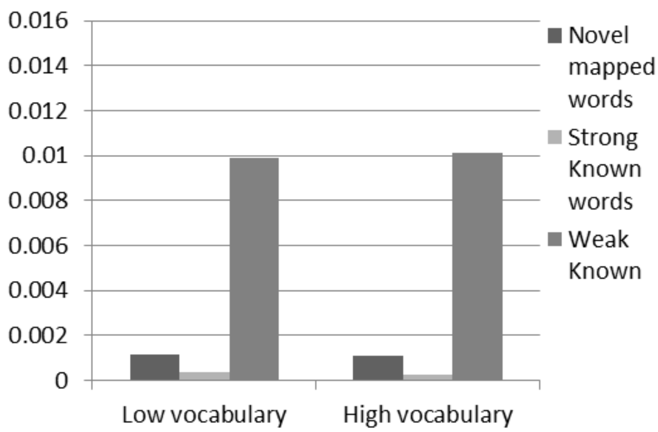
(b)



(c)

Figure A17. Analysis of the change in connection weights after referent selection in Simulation 4a for low and high vocabulary groups. Change in connection weights for words after Strong RS trials are depicted in panel (a) and change in connection weights for words after Weak RS are depicted in panel (b)

(a)



(b)

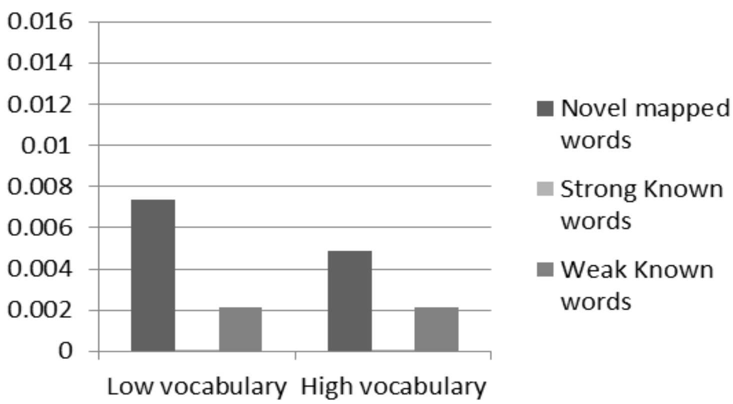
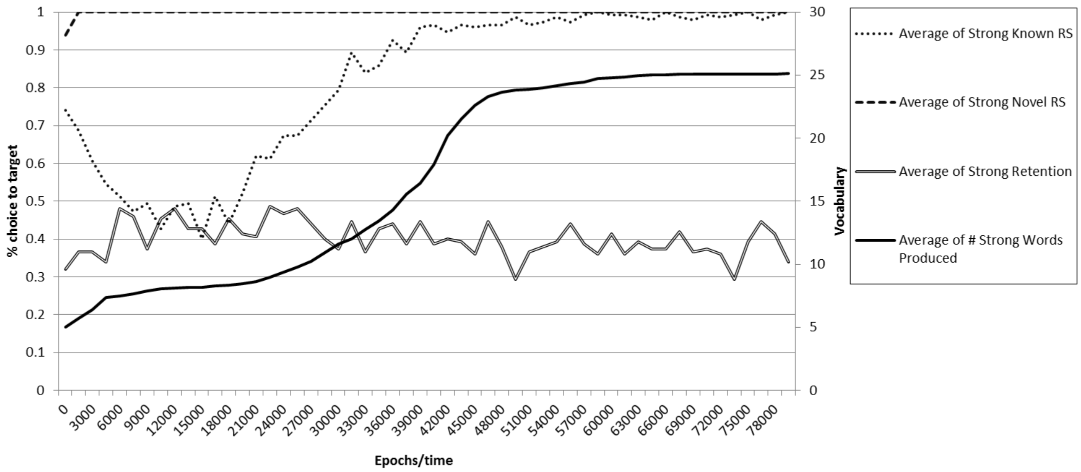


Figure A18. Results of Simulation 4b. Performance of the model on referent selection and retention, plotted as percent correct to the target along the left y-axis. Overall productive vocabulary of the model is plotted along the right y-axis. Panel (a) depicts performance on the Strong RS trials, whereas panel (b) depicts the Weak RS trials.

(a)



(b)

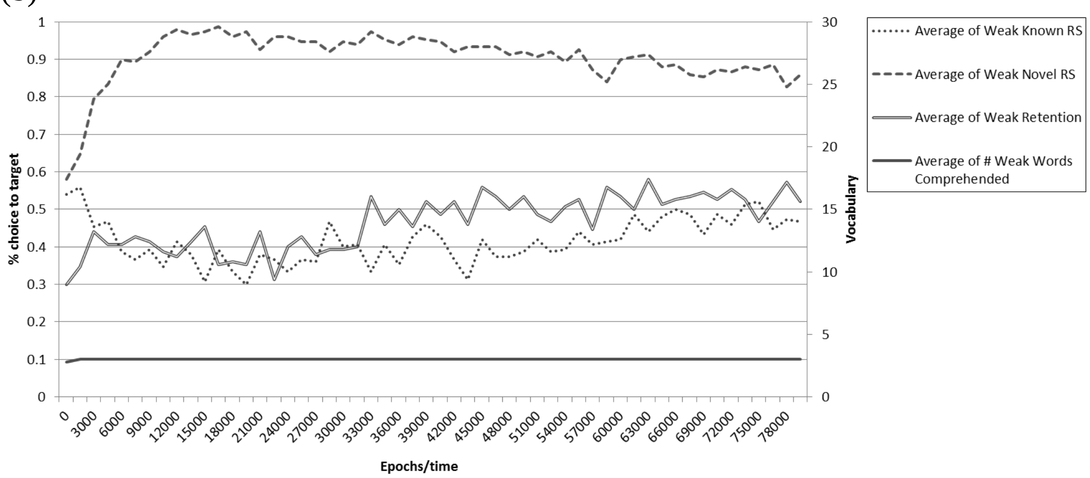
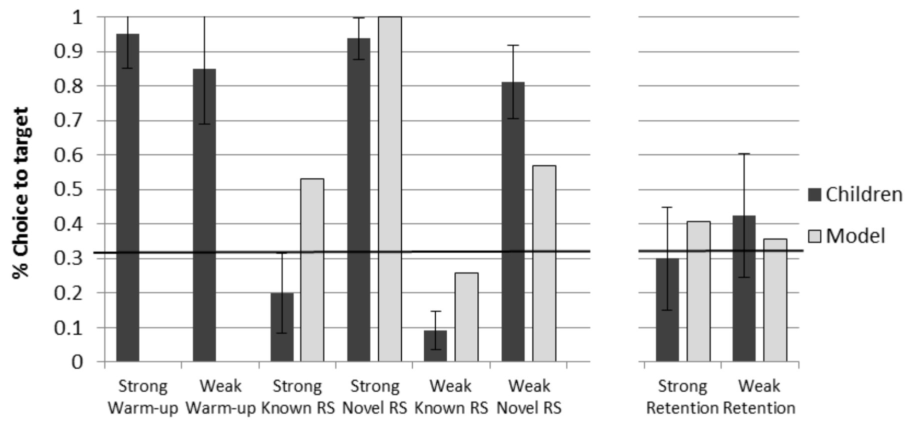
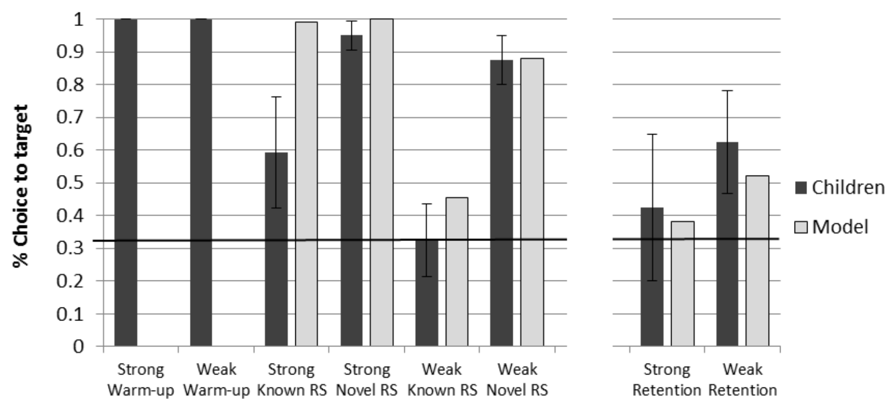


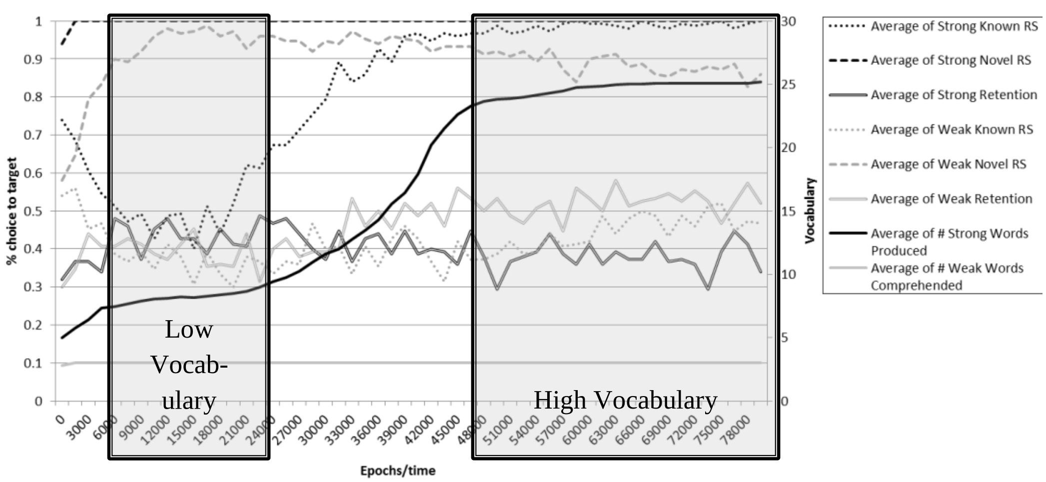
Figure A19. Comparison of referent selection and retention performance in Simulation 4b and Experiment 4b. Low vocabulary children and model are in panel (a) and high vocabulary children and model are depicted in panel (b) in Experiment 3b. Panel (c) shows where the vocabulary groups fall in the model. Note: Low vocabulary in the model included the epochs of development where the model produced between 8 and 10 words, or the first plateau in vocabulary acquisition. High vocabulary included epochs where the model produced more than 25/27 words after the vocabulary spurt.



(a)



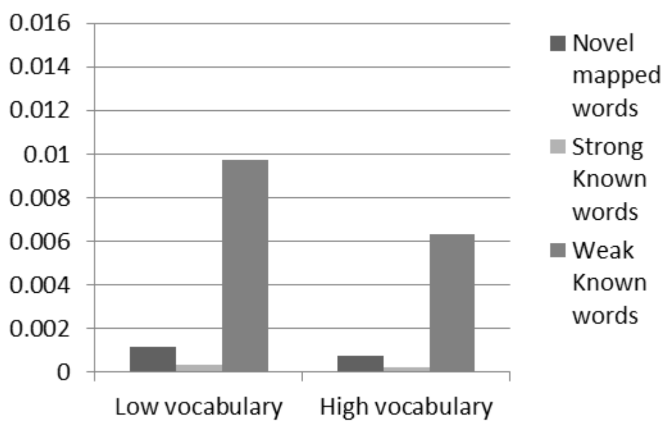
(b)



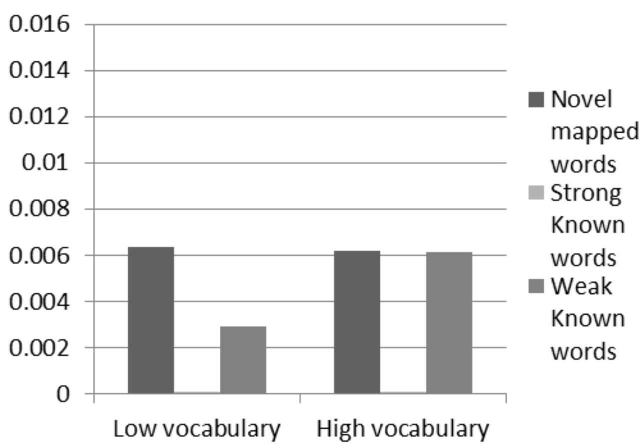
(c)

Figure A20. Analysis of the change in connection weights after referent selection in Simulation 4b for low and high vocabulary groups. Changes in connection weights after Strong RS trials are depicted in panel (a) whereas changes in connection weights after Weak RS trials are depicted in panel (b).

(a)



(b)



REFERENCES

Ackles, P.K. (2008). Stimulus novelty and cognitive-related ERP components of the infant brain. *Perceptual and motor skills, 106*(1), 3-20.

Alt, M. (2011). Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *Journal of Communication Disorders, 44*(2), 173-185.

Axelsson, E.L., Churchley, K., & Horst, J.S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology, 3*(88), doi: 10.3389/fpsyg.2012.00088

Baldwin, D.A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 874-890.

Baldwin, D.A., & Markman, E.M. (1989). Establishing word object relations: A first step. *Child Development, 60,* 381-398.

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience, 23*(11), 3181-3196.

Bion, R.A.H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39-53.

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Science*, *2*(2), 67-73.

Bornstein, M.H., & Mash, C. (2010). Experience-based and on-line categorization of objects in early infancy. *Child Development, 81*(3), 884-897.

Capone, N.C., & McGregor, K.K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research, 48*(6), 1468-1480.

Carey, S. (1978). The child as word learner. *Linguistic theory and psychological reality*, Cambridge, MA: MIT Press.

Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, *6*(3), 184-205.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15*, 17-29.

Carpenter, M., Nagell, K.., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and the communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i-174.

Civan, A., Teller, D.Y., & Palmer, J. (2005). Relations among spontaneous preferences, familiarized preferences, and novelty effects: Measurements with forced-choice techniques. *Infancy*, *7*(2), 111-142.

Clark, E.V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*(2), 417-431.

Cohen, L.B. (2004). Uses and misuses of habituation and related preference paradigms. *Infant and Child Development, 13*(4), 349-352.

Desimone, R. & Duncan, J. (1995). Neural mechanism of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*(5), 630-641.

Dollaghan, C. (1985). Child meets word: Fast mapping in preschool children. *Journal of Speech and Hearing Research, 28*(3), 449-454.

Eimas, P.D., Siqueland, E.R., Juszyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, 171*(3968), 303-306.

Fantz, R.L.(1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146*(3644, 668-670.

Fenson, L., Philip, P., Reznick, J., Bates, E., Thal, D.J., & Pethick, S.J. (1993). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), i-185.

Fernald, A., Pinot, J.P., Swingley, D., Weinberg, A., & McRoberts, G.W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*(3), 228-231.

Friedrich, M. (2011). Early word learning: Reflections on behavior, connectionist models, and brain mechanisms indexed by ERP components. In J. Guendouzi, F. Loncke, & M.J. Williams (Eds.) *The handbook of psycholinguistic and cognitive processes: Perspectives in communication disorders*. New York: Psychology Press.

Golinkoff, R.M., Hirsh-Pasek, K., Bailey, L.M., & Wenger, N.R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology, 28*(1), 99-108.

Goodman, J.C., McDonough, L., & Brown, N.B. (1998). The role of semantic context and memory in the acquisition of novel nouns. *Child Development, 69*(5), 1330-1344.

Gupta, P. (2008). The role of computational models in investigating typical and pathological behavior. *Seminars in Speech and Language, 29*(3), 211-225.

Gurteen, P.M., Horne, P.J., & Erjavec, M. (2011). Rapid word learning in 13- and 17-month-olds in naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology, 109*(2), 201-217.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87*(1), B23-B34.

Halberda, J. (2006). Is this a dax I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology, 52*, 310-344.

Heibeck, T.H., & Markman, E.M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58*(4), 1021-1034.

Hidaka, S., & Smith, L.B. (2010). A single word in a population of words. *Language Learning and Development, 6*(3), 206-222.

Hills, T.T., Maouene, J., Riordan, B., & Smith, L.B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language, 63*, 259-273.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*(4), 341-348.

Horst, J.S. (2008) Turning novel names into known names: Understanding referent selection and retention in 24-month-old children and neural networks. (Unpublished doctoral dissertation). University of Iowa, Iowa City, IA.

Horst, J.S., McMurray, B., & Samuelson, L.K. (2006). Online processing is essential for learning: Understanding fast mapping and word learning in a dynamic connectionist architecture. *Proceedings from the 28th Meeting of the Cognitive Science Society*, 339-344.

Horst, J.S., Samuelson, L.K., Kucker, S.C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition, 118*(2), 234-244.

Horst, J.S., & Samuelson, L.K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*(2), 128-157.

Horst, J.S., Scott, E.J., & Pollard, J.A. (2010). The role of competition in word learning via referent selection. *Developmental Science, 13*(5), 706-713.

Hunter, M.A., & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5*, 69-95.

Jaswal, V.K., & Markman, E.M. (2001). Learning proper and common names in referential versus ostensive contexts. *Child Development, 72*(3), 768-786.

Jaswal, V.K., & Markman, E.M. (2003). The relative strengths of indirect and direct word learning. *Developmental Psychology, 29*(4), 745-760.

Jusczyk, P.W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Science, 3*(9), 323-328,

Kagan, J. (2009). Categories of novelty and states of uncertainty. *Review of General Psychology, 13*(4), 290-301.

Kucker, S.C., & Samuelson, L.K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy, 17*(3), 295-323.

Kucker, S.C., Samuelson, L.K., & McMurray, B. (in prep). Slowing down fast mapping: Redefining the dynamics of word learning.

Kucker, S.C., Samuelson, L.K., & Spencer, J.P. (2012). When less structure in language can help: Fast mapping non-solid substances. Poster session presented at the 42nd Annual Jean Piaget Society Conference, Toronto, Ontario, Canada.

Landau, B., Smith, L.B., & Jones, S.S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*, 299-321.

Lederberg, A.R., & Spencer, P.E. (2008). Word-learning abilities in deaf and hard-of-hearing preschoolers: Effect of lexicon sixe and language modality. *Journal of Deaf Studies and Deaf Education, 14*(1), 44-62.

Markman, E. (1991). Whole object taxonomic and mutual exclusivity. In S.A. Gelman & J.P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development.* Cambridge, MA: Cambridge University Press

Markman, E.M., & Wachtel, G.F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, *20*(2), 121-157.

Markman, E.M., Wasow, J.L., & Hansen, M.B. (2003). Use of mutual exclusivity assumption by young word learners. *Cognitive Psychology, 47*(3), 241-275.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 385*, 813-815.

Mather, E. (2013). Bootstrapping the early lexicon: How do children use old knowledge to create new meanings? *Frontiers in Psychology, 4*(96), doi: 10.3389/fpsyg.2013.00096

Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy, 14*(1), 60-76.

Mather, E., & Plunkett, K. (2010a). Novel labels support 10-month-olds' attention to novel objects. *Journal of Experimental Child Psychology, 105*(3), 232-242.

Mather, E., & Plunkett, K. (2010b). Mutual exclusivity and phonological novelty constrain word learning at 16 months. *Journal of Child Language, 38*(5), 933-950.

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science, 36*(7), 1157-1177.

Mayor, J., & Plunkett, K. (2010). A neuro computational account of taxonomic responding and fast mapping in early word learning. *Psychological Review, 117*(1), 1-31.

Mayor J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science, 14*(4), 769-785.

McGregor, K.K., Rohlfing, K.J., Bean. A., & Marschner, E. (2008). Gesture as a support for word learning: The case of under. *Journal of Child Language, 36*(4), 807-828.

McGregor, K.K., Sheng, L., & Smith, B. (2005). The precocious two-year-old: Status of the lexicon and links to the grammar. *Journal of Child Language, 32*(3), 563-585.

McMurray, B. (2007). Defusing the childhood naming explosion. *Science*, *317*, 631.

McMurray, B., & Spivey, M. (2000). The categorical perception of consonants: The interaction of learning and processing. *Proceedings of the Chicago Linguistic Society, 34*(2), 205-220.

McMurray, B., Horst, J.S., & Samuelson, L.K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*(4), 831-877.

McMurray, B., Zhao, L., Kucker, S.C., & Samuelson, L.K. (2013). Pushing the envelope of associative learning: Internal representations and dynamic competition transform association into development. In L.Gogate & G. Hollich (Eds.) *Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence.* Hershey, PA: IGI Global.

Merriman, W.E., & Bowman, L.L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*(3/4), i-129.

Merriman, W.E., & Schuster, J.M. (1991). Young children's disambiguation of object name reference. *Child Development, 62*(6), 1288-1301.

Mervis, C.B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development, 65*(6), 1646-1662.

Mundy, P.C. (1984). The effects of stimulus preferences on infants attention to novel stimuli. *Child Study Journal, 14*(4), 287-300.

Munro, N., Baker, E., McGregor, K., Docking, K., & Arculi, J. (2012). Why word learning is not fast. *Frontiers in Psychology, 3*(4), doi: 10.3389/fpsyg.2012.00041

Naigles, L.R., & Swenson, L.D. (2007). Syntactic supports for word learning. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development,* Malden, MA: Blackwell Publishing.

Oakes, L.M., Coppage, D.J., & Dingel, A. (1997). By land or by sea: The role of perceptual similarity in infants' categorization of animals. *Developmental Psychology, 33*(3), 396-407.

Oviatt, S.L. (1982). Inferring what words mean: Early development in infants' comprehension of common object names. *Child Development, 52*(1), 274-277.

Perry, L.K., Samuelson, L.K., Malloy, L.M., & Schiffer, R.N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science, 21*(12), 1894-1902.

Perry, L.K., & Samuelson, L.K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology, 2*(345), doi: 10.3389/fpsyg.2011.00345

Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*, 819-865.

Samuelson, L.K., & Smith, L.B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*, 1-33.

Sinnett, S., Costa, A., Soto-Faraco, S. (2006). Manipulating inattention blindness within and across sensory modalities. *The Quarterly Journal of Experimental Psychology*, *59*(8), 1425-1442.

Smith, L.B., Jones, S.S., Landau B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*(1), 13-19.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558-1568.

Snyder, K.A., Blank, M.P., & Marsolek, C.J. (2008). What form of memory underlies novelty preferences? *Psychonomic Bulletin & Review, 15*(2), 315-321.

Southgate, V., & Meints, K. (2000). Typicality, naming, and category membership in young children. *Cognitive Linguistics, 11*(1/2), 5-16.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Psychology, 109*, 132-140.

Suanda, S.H., & Namy, L.L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science, 36*(3), 545-559.

Suanda, S.H., & Namy, L.L. (2013). The organization of words and symbolic gestures in 18-month-olds' lexicons: Evidence from a disambiguation task. *Infancy, 18*(2), 276-288.

Swingley, D. (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development*, *6*(3), 179-183.

Swingley, D., & Aslin, R.N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54*, 99-132.

Tarbi, E.C., Sun, X., Holcomb, P.J., & Daffner, K.R. (2011). Surprise? Early visual novelty processing is not modulated by attention. *Psychophysiology*, *48*(5), 624-632.

Uehara, I. (2000). Transition from novelty to familiarity preferences depending on recognition performance by 4-year-olds. *Psychological Reports, 87*(3), 837-848.

Vlach, H.A., & Sandhofer, C.M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology, 3*(4), doi: 10.3389/fpsyg.2012.00046

Wagar, B.M., & Dixon, M.J. (2005). Past experience influences object representation in working memory. *Brain Cognition*, *57*(3), 248-256.

White, K.S., & Morgan, J.L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114-132.

Wilkinson, K.M., (2005). Disambiguation and mapping of new word meaning by individuals with intellectual/development disabilities. *American Journal on Mental Retardation, 110*(2), 71-86.

Wilkinson, K.M. (2007). Effect of "missing" information on fast mapping by individuals with vocabulary limitations associated with intellectual disability. *American Journal on Mental Retardation, 112*(1), 40-53.

Wilkinson, K.M., & Mazzitelli, K. (2003). The effect of "missing" information on children's retention of fast-mapped labels. *Journal of Child Language, 30*(1), 47-73.

Wilkinson, K.M., Ross, E., & Diamond, A. (2003). Fast mapping of multiple words: Insights into when "the information provided" does and does not equal "the information perceived". *Journal of Applied Developmental Psychology, 24*(6), 739-762.

Woodward, A.L., Markman, E.M., & Fitzsimmons, C.M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology, 30*(4), 553-556.

Yoshida, H., & Burling, J.M. (2012). Highlighting: A mechanism relevant for word learning. *Frontiers in Psychology, 3*(262), doi: 10.3389/fpsyg.2012.00262

Yoshida, H., & Hanania, R. (2011). If it's red, it's not Vap: How competition among words may benefit early word learning. *First Language, 33*(1), 3-19.

Yu, C., & Smith, L.B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414-420.

Zosh, J.M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science, 17*(1), 20-28.