Distribution Agreement

The Role of Cross-Situational Word Learning in Children's Vocabulary Acquisition:
Theory, Behavior, and Mechanisms

By

Sumarga H. Suanda
Doctor of Philosophy

Psychology

____
Laura L. Namy
Advisor

____
Robert R. Hampton
Committee Member

____
Stella F. Lourenco
Committee Member

____
Lynne C. Nygaard
Committee Member

____
Phillip C. Wolff
Committee Member

Accepted:

____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____
Date

The Role of Cross-Situational Word Learning in Children's Vocabulary Acquisition:
Theory, Behavior, and Mechanisms

By

Sumarga H. Suanda
M.A., Emory University, 2009

Advisor: Laura L. Namy, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2012

Abstract

The Role of Cross-Situational Word Learning in Children's Vocabulary Acquisition: Theory, Behavior, and Mechanisms
By Sumarga H. Suanda

Over the first 6 years of life, children are reported to have amassed a vocabulary of about 14,000 words. What are the processes that underlie such prolific learning? Based on a wealth of evidence built up over the past 40 years, one process appears to be *fast mapping*, children's ability to draw on a host of referential cues to infer a word's meaning from a single exposure to a new word. More recently, a growing body of evidence has suggested that another process is *cross-situational word learning*, word learners' ability to determine word meaning not within a single encounter but across multiple encounters by tracking the cross-situational consistency between words and their candidate referents. Although the notion that children acquire their vocabulary at least in part through cross-situational learning is neither novel nor unintuitive, theoretical treatments and empirical investigations into this learning process are scarce relative to those of fast mapping.

The overarching goal of the research presented herein is to further investigate the nature of children's cross-situational learning capacities and to better understand the role of this type of learning in vocabulary acquisition. Two behavioral experiments that examine 5- to 7-year-olds' cross-situational word learning are reported. These experiments constitute the first empirical investigations of school-aged children's ability to acquire new word-to-referent mappings when the only cues to reference are the cross-situational co-occurrence statistics between words and their referents. These studies also examine whether some of the behavioral signatures previously observed in adults' cross-situational learning are also evident in children's learning. Additionally, a series of computational simulations that explore the candidate mechanisms underlying children's cross-situational word learning are reported. Although the results of these simulations do not provide a conclusive mechanistic account of children's learning, they do specify some conditions that readily account for the observed learning patterns and accurately predict other empirical phenomena. Collectively, the research presented here contributes to our understanding of the various processes that make children's impressive word learning possible.

The Role of Cross-Situational Word Learning in Children's Vocabulary Acquisition: Theory, Behavior, and Mechanisms

By

Sumarga H. Suanda
M.A., Emory University, 2009

Advisor: Laura L. Namy, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2012

Acknowledgments

First and foremost I want to thank Laura Namy, who for the past five years has been the best advisor, mentor, teacher, advocate and role model I could imagine. I am grateful to Laura for encouraging me from my very first day of graduate school to pursue areas of research that I found most interesting and for being completely invested and supportive of all my scholarly endeavors. I am also indebted to Laura for pushing me to see the big picture without overlooking the smaller pictures, for coaching me in how to conduct methodical and responsible science, and for continuing to teach me how to communicate and write about science with clarity.

I am also thankful for all the mentorship and opportunities given to me on my road to Emory. I thank Lisa Dierker, John Seamon, and particularly Henry Danso, at Wesleyan who introduced me to psychology, language acquisition and psychological research. I am grateful to Art Glenberg and the UW-Madison 2004 PREP summer program for getting me hooked on research and convincing me that graduate school in psychology would be in my future. I am especially grateful to Liz Brannon for giving me the opportunity to work in her lab at Duke. I couldn't have asked for a more perfect preparation for graduate school and a more enjoyable two years of research. I want to also thank Dukies Jessica Cantlon and Mat Fleck, two of my closest friends, for being so incredibly helpful in my graduate school applications and for continuing to be invaluable sources for advice, support, and friendship.

At Emory, I am indebted to an amazing group of faculty members who have been so generous with their time and who have supported me and taught me so much over these past 5 years. I especially want to thank Lynne Nygaard and Rob

Hampton for their willingness to serve on all of my committees, for providing insightful comments and helpful feedback on my research, and for supporting my work. I also thank Phil Wolff, Stella Lourenco, Philippe Rochat, Jack McDowell, David Edwards, Bob McCauley, Larry Barsalou and Mel Konner who through courses, exams and informal meetings have all improved my thinking and broadened my perspective immensely.

I want to thank members of the Language and Learning Lab (Brandi Kenner, Christina Tzeng, Anna Heilbrun, Jane Fisher, Nassali Mugwanya, Maggie Dancel, Ann Marie Finley, Katie Kimura, Arielle Friehling, Pat Finan, Cassie Hendrix, Kristie Denlinger, and Carlie Marks) for their cheerful spirits, invaluable assistance, and support of my work. I am especially indebted to Nassali Mugwanya, whose hard work and dedication is responsible for a large portion of the children's research described here. I couldn't have made it through these past 5 years without all my great Emory Psychology buddies, Virginia-Highlands neighbors, soccer teammates, squash partners, and all the Atlanta friends who all helped make the ATL such a fun home that is full of fond memories. I'll miss y'all.

Finally I wouldn't be where I am without the love, cheering and backing of my family (Marjorie, Endo, Ata Suanda, Kirsten Wisneski, James and Lucia LaVilla-Havelin, Fran Havelin, Ralph Marash). I want to thank my parents, Marjorie and Endo Suanda, especially, for always giving me their love, believing in me, and encouraging me in everything I've done. I dedicate this dissertation to them.

Table of Contents

Chapter 1. Introduction……………………………………………………………………………...1

Chapter 2. Fast Mapping and Cross-Situational Approaches to Children's Word Learning………………………….………………………………………………………………..4

2.1 A Fast Mapping Approach to the Mapping Problem………………………………....5

2.2 A Cross-Situational Approach to the Mapping Problem……………………………...9

2.2.1 Cross-situational word learning: Behavioral evidence……………………….10

2.2.2 Cross-situational word learning: Computational evidence………………...16

2.3 Outstanding Questions…………………………………………………………………...18

2.3.1 Scaling from adult word learning to children's word learning……………..19

2.3.2 The nature of the underlying learning mechanisms…………………………..21

Chapter 3. The Effect of Contextual Diversity on Children's Cross-Situational Word Learning……………………………………………………………………………………………24

3.1 Background…………………………………………………………………………………..24

3.2 Method………………………………………………………………………………………….29

3.2.1 Participants………………………………………………………………………………29

3.2.2 Stimuli……………………………………………………………………………………….30

3.2.3 Design……………………………………………………………………………………….31

3.2.4 Procedure…………………………………………………………………………………33

3.2.4.1 Familiarization phase………………………………………………………………33

3.2.4.2 Learning phase………………………………………………………………………..35

3.2.4.3 Test phase……………………………………………………………………………...36

3.2.5 Coding………………………………………………………………………………………37

3.3 Results..38

3.4 Discussion..42

3.4.1 Cross-situational word learning in children's vocabulary growth............43

3.4.2 Statistical learning in adult and child learners...45

3.4.3 Why does contextual diversity aid learning? 47

3.4.4 Conclusion...49

Chapter 4. Probing the Precision of Children's Mappings in Cross-Situational Word Learning...50

4.1 Background..50

4.2 Method..51

4.2.1 Participants...51

4.2.2 Stimuli, Design & Procedure..51

4.3 Results..52

4.4 Discussion...55

Chapter 5. Children's Cross-Situational Word Learning: Hypothesis Testing or Associative Learning? 59

5.1 Background..59

5.1.1 Hypothesis Testing..61

5.1.2 Associative Learning...63

5.1.3 Current Endeavor..65

5.2 Models...66

5.2.1 Hypothesis Testing (HT) Model..66

5.2.2 Associative Learning (AL) Model...70

5.3 Simulation 1: The role of contextual diversity in cross-situational word learning……………………………………………………………………………………………73

5.3.1 Method…………………………………………………………………………………73

5.3.2 Results…………………………………………………………………………………75

5.3.2.1 HT model results…………………………………………………………75

5.3.2.2 AL model results…………………………………………………………76

5.3.2.3 Correlations between models and children's performance……………76

5.3.3 Discussion……………………………………………………………………………80

5.4 Simulation 2: Item selection as a function of cross-situational co-occurrence statistics……………………………………………………………………………………………81

5.4.1 Method…………………………………………………………………………………81

5.4.2 Results…………………………………………………………………………………82

5.4.2.1 HT model results…………………………………………………………82

5.4.2.2 AL model results…………………………………………………………83

5.4.2.3 Correlations between models and children's performance……………84

5.4.3 Discussion……………………………………………………………………………87

5.5 Simulations 3 and 4: A Modified Associative Learning Model……………………88

5.5.1 Method…………………………………………………………………………………90

5.5.2 Results…………………………………………………………………………………91

5.5.2.1 Contextual diversity effects in MAL models……………………………91

5.5.2.2 Item selection effects in MAL models……………………………………92

5.5.2.3 Correlations between MAL models and children's performance……93

5.5.3 Discussion……………………………………………………………………………94

5.6 Simulation 5: Accounting for other empirical data……………………………………..95

5.6.1 Method……………………………………………………………………………………..96

5.6.2 Results and Discussion………………………………………………………………97

5.7 General Discussion………………………………………………………………………….98

5.7.1 Empirical evidence for the proposed parameters that improve model fit.

……………………………………………………………………………………………………99

5.7.2 Implications for the hypothesis testing vs. associative learning debate…….

………………………………………………………………………………………………102

5.7.3 Conclusion………………………………………………………………………………105

Chapter 6. General Discussion……………………………………………………………107

6.1 Contributions of a Cross-Situational Approach…………………………………108

6.1.1 Cross-situational co-occurrence statistics as an additional cue to

learning……………………………………………………………………………………109

6.1.2 The role of cross-situational learning in the discovery of fast mapping

cues……………………………………………………………………………………………112

6.1.3 Cross-situational learning as a common theme underlying vocabulary

acquisition research programs……………………………………………………116

6.1.4 Cross-situational learning as statistical learning: Towards a unified

account of language acquisition…………………………………………………117

6.2 Limitations to the Cross-Situational Approach…………………………………121

6.2.1 Can cross-situational learning scale up to real-world learning

environments? ……………………………………………………………………………121

6.2.2 Is cross-situational learning applicable to the acquisition of non-object labels? 123
6.2.3 Is cross-situational learning ability related to real-world vocabulary growth? 126
6.3 Conclusions..128

List of Tables

Table 1. Association matrices representing word–picture co-occurrence frequencies across conditions……………………………………………………………………………………..32

Table 2. Pair-wise comparisons of the likelihood selecting a particular item type (word-object co-occurrence frequency type)……………………………………………………55

Table 3. Core differences between hypothesis testing and associative learning models on three computational properties……………………………………………………..66

Table 4. Correlation results between models' and children's performance in Chapter 3………………………………………………………………………………………………………...78

Table 5. Correlation results between models' and children's performance in Chapter 4………………………………………………………………………………………………………...85

Table 6. The equipotentiality of hypothesis testing and associative learning in accounting for word learning phenomena……………………………………………………...104

List of Figures

Figure 1: A sample of a series of learning trials containing the word "hiplex" and its referent (A). A 4-Alternative-forced-choice (4AFC) task testing children's learning of the word "hiplex" (B)……………………………………………………………………………….27

Figure 2: A sample of learning trials with high contextual diversity (A). A sample of learning trials with low contextual diversity (B)……………………………………………….....29

Figure 3: Stimuli employed in the experiment: pictures of novel/altered objects used (A); orthographic representation of novel spoken words used (B)………………………..31

Figure 4: Image still from a participant depicting the experimental testing layout..…………………...……………………………………………………………………………….33

Figure 5: Sample trial structure in the experiment: a learning trial (A); a test trial (B)………………………………………………………..……..………………………………………..….35

Figure 6: Mean proportion correct across levels of contextual diversity....……………..39

Figure 7: Proportion of children performing above chance level (.25) across levels of contextual diversity…………………………………………………………………………………………41

Figure 8: Distribution of answers to objects differing in co-occurrence frequency with target word…………...………………………………………………………………………………...53

Figure 9: Toy example illustrating the learning algorithms implemented in the Hypothesis Testing model (A) and Associative Learning model (B)……….……………..69

Figure 10: HT models' performance in Simulation 1: Mean proportion correct across models varying in word knowledge threshold (ω) and across conditions of contextual diversity…………...…………………………..…………………………………………....76

Figure 11: AL models' performance in Simulation 1: Mean proportion correct across models varying in learning rates (α) and across conditions of contextual diversity.77

Figure 12: Proportion of simulated experiments of HT and AL models whose results correlated strongly, moderately, and weakly with children's contextual diversity effects……………………………………………………………………………………79

Figure 13: Performance of HT models varying in word knowledge threshold (ω) in Simulation 2: Distribution of answers to objects differing in co-occurrence frequency with the target word………………………………………………………………83

Figure 14: Performance of AL models varying in learning rates (α) in Simulation 2: Distribution of answers to objects differing in co-occurrence frequency with the target word……………………………………………………………………………………84

Figure 15: Proportion of simulated experiments of HT and AL models whose item selection effects correlated strongly, moderately, and weakly with children's item selection effects………………………………………………………………………………86

Figure 16: Contextual diversity effects in MAL models differing in learning rates (α) and familiarity bias scaling parameters (λ)………………………………………………91

Figure 17: Item selection effects in MAL models differing in learning rates (α) and familiarity bias scaling parameters (λ)……………………………………………………92

Figure 18: Proportion of simulated experiments across MAL models whose results correlated strongly, moderately, and weakly with children's contextual diversity effects……………………………………………………………………………………94

Figure 19: Comparison of human adult learner performance reported in Yu & Smith (2007) to simulations of an HT and MAL model………………………………………97

Chapter 1. Introduction

How do children learn the meanings of words? For the past 40 years, investigations into children's word learning has had both theoretical and practical significance in the cognitive, developmental, linguistic and educational sciences. From a theoretical perspective, examinations of word learning have served as test beds for accounts of children's developing conceptual systems (see Murphy, 2002), the nature of children's socio-cognitive competence (see Tomasello, 2003), the origins of children's linguistic capacities (Gleitman, 1990), and the mechanics of children's learning (Smith, Colunga, & Yoshida, 2010). Additionally, the study of word learning has been a battleground for a number of major theoretical debates in the cognitive and developmental sciences including the domain-generality versus domain-specificity debate of children's learning and cognitive capacities (e.g., Markson & Bloom, 1997; Namy & Waxman, 1998; see Namy, 2012, for a review), and the continuity versus discontinuity debate of learning and cognitive processes over development (McMurray, 2007; Nazzi & Bertonucci, 2003).

From a more applied standpoint, there is great interest in children's word learning in part due to the large inter-individual differences observed in vocabulary sizes and growth patterns, and the predictive powers of these differences. For example, whereas some two-year-olds have productive vocabularies of greater than 500 words, others have vocabularies of fewer than 50 words (Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994). The toddlers with lower vocabularies, or *late talkers* (Rescorla, 2009), are at greater risk for developing *Specific Language Impairments* (SLI), a condition marked by difficulty understanding and/or producing spoken language without any clear hearing, physical, cognitive or social deficits (Tomblin, Records, Buckwalter, Zhang, Smith, & O'Brien, 1997). Additionally, having a low vocabulary is associated with poorer pre-literacy and literacy skills (for a review, see Lee, 2011). Thus, both speech pathologists and educators are invested in better understanding the nature of early word learning in order to gain insight into the nature and sources of the variability, as well as to develop intervention strategies to remedy the negative consequences of low vocabularies.

The study of word learning involves understanding at least three inter-connected problems children that face. First, children must segment spoken speech into the correct word units and maintain phonological representations of those units (Cutler, 1994). Second, children must carve their representation of the world into the appropriate concepts and categories to which words refer (Carey, 1994). Finally, and arguably the focus of most word learning research, children must correctly link the phonological representation to the conceptual representation, a task referred to as the *mapping problem* (L. Bloom, 2000; Gleitman, 1990).

The overarching goal of the research reported herein is to extend our understanding of how children solve this mapping problem. In Chapter 2, I provide a brief literature overview of proposed solutions to the mapping problem. I first describe the most commonly proposed solution, which is that children are sensitive to a host of cues that allow them to figure out the meanings of words at the moment new words are uttered (i.e., children *fast map* words onto their correct referents). I then describe a second proposed solution, which is that children learn words by gradually accumulating the co-occurrence regularities between words and their referents over time (i.e., children learn word-referent pairings *cross-situationally*). Although this second solution to the problem is an intuitive idea, it has been neglected until recent years in empirical word learning research. Further, much of the recent work that has examined cross-situational word learning has employed adult learners as models of early word learning (e.g., Yu & Smith, 2007; Suanda & Namy, 2012). Thus across Chapters 3 and 4, I present two empirical studies testing the nature of children's (6-year-olds) ability to learn new words across a series of ambiguous naming events when the only cue to reference is the cross-situational regularities with which words and referents co-occur. Then in Chapter 5, I present a series of computational simulations that shed light on candidate mechanisms that may underlie children's cross-situational word learning. I conclude this dissertation, in Chapter 6, with a general discussion of the place of cross-situational learning in word learning theory and some suggestions for future directions that may contribute further to cross-situational learning's impact on the field.

Chapter 2. Fast Mapping and Cross-Situational Approaches to Children's Word Learning

To help conceptualize the mapping problem that children face, language acquisition researchers (e.g., Macnamara 1972; Markman, 1987; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Woodward, 2000) regularly invoke an argument initially raised by the philosopher W.V.O. Quine in reference to the problem of translating words across two languages. Quine (1960) asked us to imagine a linguist attempting to translate a new word uttered by a speaker of an unknown language (e.g., "gavagai") as a white rabbit scurried by. What is the English translation of "gavagai"? Is it a basic-level category term (e.g., "rabbit")? Is it a superordinate category term (e.g., "animal")? Or is it a property term (e.g., "white")? According to Quine, the word's translation, and thus its meaning, is simply underdetermined from physical experience. That is, even if the linguist was able to rule out translations such as "white" or "animal" through further instances of "gavagai", there still exist an infinite number of possible translations besides "rabbit". For example, Quine suggests that "gavagai" might mean "all and sundry undetached rabbit parts…[or]…brief temporal rabbit segments" (Quine, 1960, pp. 52-53). Quine's point is that for any given observation, or set of observations, of a word ("gavagai") associated with a non-verbal stimulus (a white rabbit), there exist an infinite number of hypotheses about the word's meaning consistent with that observation (Quine, 1960; 1968; 1987).

Children appear to face a similar problem in learning new words. That is, imagine a child playing with a toy train and hearing his mother say the word "train".

How does the child figure out to what the word "train" refers? Does the word refer to trains in general, that particular train, the act of playing with trains, the color of the train, etc.? Despite this under-specificity, children are exceptionally proficient word learners. According to one estimate, by the age of six, children have amassed a vocabulary of 14,000 words (Carey, 1978). Some have translated this number into the impressive statistic that children acquire on average 9-10 new words each day (Golinkoff & Hirsh-Pasek, 2000; Woodward & Markman, 1998; but see P. Bloom, 2004). Further, children apparently require very little effort or exposure to learn a word. That is, under some circumstances children learn a word after only a single exposure, a process known as *fast mapping* (Carey, 1978; Carey & Bartlett, 1978; Heibeck & Markman, 1987; Markson & Bloom, 1997).

2.1 A Fast Mapping Approach to the Mapping Problem

A central goal of lexical acquisition research has thus been to explain this paradox between the apparent difficulty of the mapping problem in word learning and the apparent facility with which children learn words. Many, if not most, empirical investigations into children's word learning have involved simple, well-controlled, artificial settings. For example, experiments commonly employ an initial *labeling phase* in which an experimenter introduces a novel word (e.g., "blicket") in the context of a particular object (e.g., a whisk), or set of objects. Importantly, despite the simple nature of this artificial learning task, it simulates the ambiguity of word learning described in Quine's *gavagai* scenario. That is, "blicket" may refer to that particular whisk, whisks in general, the color of the whisk, etc. The experimenter then tests the child's word mapping in a *testing phase*, by later presenting the labeled object (or a related object, such as a different whisk), as well as an unlabeled object(s), and asking the child to "find the blicket".

Variations on this basic paradigm have provided researchers with insights into the fast mapping process (see P. Bloom, 2000; Golinkoff et al., 2000; Woodward & Markman, 1998; for reviews). For example, manipulating the types of objects available in the testing phase has revealed children's bias to interpret novel words in specific ways. For example, if children are taught in the labeling phase that "blicket" refers to a red plastic whisk, they will tend to select in the testing phase a wooden whisk over a red plastic fork as an extension for "blicket". Findings such as this have been interpreted as a tendency to extend newly learned nouns to objects similar in shape (known as the *shape bias*, see Samuelson & Bloom, 2008, for a review), rather than objects similar in material (see Smith, 2000). Additionally, manipulating the social context of the labeling phase has highlighted children's attention to social variables when interpreting new words. For example, if the speaker is not looking at the whisk at the time of utterance, children will reject the label as referring to the whisk (e.g., Baldwin, 1991; 1993a; 1993b). This manipulation, as well as other manipulations of social contextual cues including the speaker's pointing gestures (Jaswal & Hansen, 2006), the speaker's knowledge state (Akhtar, Carpenter, & Tomasello, 1996), or the speaker's intentional state at the time of utterance (Diesendruck, Markson, Akhtar, & Reudor, 2004; Tomasello & Barton, 1994), have been interpreted as evidence that children are tuned in to the speaker's referential intentions; and, that this "mind-reading" capacity is what allows children to quickly match words to the correct referents (see Akhtar & Tomasello, 2000; P. Bloom, 2000; for reviews).

Yet other manipulations to this basic paradigm have led to the conclusion that children also attend to a range of linguistic cues in learning new words. For example, Namy and Waxman (2000) introduced 18-month-olds to novel word-object pairings either embedded within a familiar sentence frame (e.g., "Look at the blicket!") or in isolation (e.g., "Blicket!"). Eighteen-month-olds mapped the words to their referent objects more readily in the former condition, suggesting that infants may use particular sentential contexts (e.g., "look at the ____" or "this is a _____") as indexing reference (see also Fernald & Hurtado, 2006; Kedar, Casasola, & Lust, 2006). Other researchers have examined young word learners' attention to subtler linguistic cues to word meaning (e.g., Belanger & Hall, 2006; Hall, Lee, & Belanger, 2001; Katz, Baker, & Macnamara, 1974). For example, Belanger and Hall (2006) revealed that at 20, but not 16 months, infants can use the presence of a determiner (e.g., "a", "the") to distinguish possible word meanings. When an experimenter labeled an object by saying, "This is daxy", infants interpreted "daxy" as denoting only that individual object (as if it were a proper name like "John"). However, when an experimenter labeled an object by saying, "This is <u>a</u> daxy", infants extended the label to other similar objects (as if it were a count noun, like "car").

Studies such as these highlight children's impressive capacity for solving the problem of referential ambiguity *at the moment* novel words are encountered through the use of attentional (Smith, 2000), social (Akhtar & Tomasello, 2000), linguistic (Gleitman, 1990), and conceptual (Markman, 1990) cues to reference.

Although much has been learned, and continues to be learned (e.g., Borovsky, Elman, & Kutas, 2012; Gampe, Liebal, & Tomasello, 2012), about children's lexical acquisition through fast mapping, there are reasons to believe that children's real world word learning is not always a function of fast mapping within a single naming instance. First, as many scholars have noted and cautioned (e.g., Smith & Yu, 2008; Tomasello, 2003), children's real world word learning environments are far more cluttered than the prototypical word learning study. In the real world, children are exposed to full sentences containing many words (e.g., "Let's put your car and train into the box"), any of which could refer to a number of visible referents (e.g., among other things, a car and its properties, a train and its properties, and a box and its properties). Thus word learning biases, such as the *shape bias* mentioned above, would not enable children to infer which is the "train" and which is the "car". Second, a number of scholars have reported that the nature of children's input varies across cultures (see Lieven, 1994, for review) and that word learning studies do a particularly poor job of simulating the experiences of these children. That is, it has been suggested that at least in some Javanese (Smith-Hefner, 1988), Kaluli (Ochs & Schieffelin, 1984; 2008), Mayan (Pye, 1986), and Samoan (Ochs & Shieffelin, 1984; 2008) cultures, children rarely are presented with the types of direct parent-child interactions the prototypical word learning study is intended to mimic. It is likely that from the perspective of children in these cultures, their learning contexts provide less direct access to the rich social referential cues present in Western cultures, suggesting that children do not learn words exclusively through such cues.

A final reason to suspect that there is more to the word learning process than children's quick initial mappings is that although there is some evidence that even 12- to 13-month-olds can learn a new word following only a single or a few exposures (Campbell & Namy, 2003; Woodward & Hoyne, 1999; Woodward, Markman, & Fitzsimmons, 1994), a close inspection of the data for children under the age of two reveal that this ability is delicate. That is, alterations to the testing procedures (Yoshida, Fennell, Swingley, & Werker, 2009), reward regimen (Evey & Merriman, 1998), phonological properties of the to-be-learned words (Werker, Fennell, Corcoran, Stager, 2002, Yoshida et al., 2009), saliency of the to-be-learned referents (Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006), and presence of labeling frames (Namy & Waxman, 2000) result in no learning, reduced learning or patterns difficult to explain such as sporadic gender differences (Katz et al., 1974; Woodward et al., 1994). Further, a number of recent studies have demonstrated that even when words are successfully fast-mapped, children's retention of these words was weaker than initial studies suggested (Horst & Samuelson, 2008; Horst, Scott, & Pollard, 2010; Kucker & Samuelson, 2012; Vlach & Sandhofer, 2012). Thus, these findings suggest that at least early in development, the end product of fast mapping may be a fragile, fleeting initial hypothesis about a word's meaning rather than a definitive final mapping.

2.2 A Cross-Situational Approach to the Mapping Problem

The sufficiency of a fast mapping solution to the mapping problem has long been questioned. In fact, even when Carey and Bartlett initially coined the term *fast mapping*, they were quite clear that it represents only a component of children's lexical acquisition; and that only through a "long, drawn out mapping, extended over the entire period of several encounters with the word" (Carey & Bartlett, 1978, pp. 18) do children acquire a word's full meaning (see also Carey, 1978, 2010). Other prominent language acquisition scholars (Gleitman, 1990; Pinker, 1989) have also argued that word learning in part involves attending to the systematic regularities across the contexts in which words appear, a process known as *cross-situational word learning* (see also Yu & Smith, 2007). Despite its intuitive appeal and long-standing history, empirical investigations and theoretical treatments of this process are scarce. In current theories that endorse some form of cross-situational word learning (e.g., P. Bloom, 2000; Tomasello, 2003), the mechanisms are rarely discussed and the importance of this learning process takes back seat to fast mapping capacities. Recently however, interest in cross-situational word learning has experienced a revival within cognitive and developmental psychology. The goal of this section is to review the recent evidence, both behavioral and computational, for cross-situational word learning.

2.2.1 Cross-situational word learning: Behavioral evidence

In a series of recent studies, Yu and Smith (Smith & Yu, 2008, in press; Yu & Smith, 2007, 2011) investigated the extent to which human learners could learn word-to-object mappings purely through the regularities across contexts with which words and objects co-occur (i.e., co-occurrence statistics). Employing adult subjects as model word learners, Yu and Smith (2007) presented subjects with a series of learning trials, each involving ambiguous reference. In each trial, subjects viewed multiple pictures of objects (between 2 and 4) simultaneously on a computer screen and heard multiple spoken words played sequentially in a random order. In each trial, it was unclear which words referred to which objects. However, over trials, every time the subjects heard a particular word, its corresponding referent object was present. Further, across trials, word-object pairs did not always appear with the same set of accompanying word and objects. Thus, subjects could learn the words if they attended to the regularities with which particular words and objects co-occurred (but see K. Smith, Smith, & Blythe, 2009 for an alternative explanation). Yu and Smith found that adult learners were remarkably sensitive to these co-occurrence statistics. Subjects correctly mapped up to 50% of the total words tested to their referents (where chance responding would predict 25% accuracy) after encountering each word-object pairing only six times, all in ambiguous situations. This capacity is robust and the finding has been replicated across multiple labs (Ichinco, Frank, & Saxe, 2009; K. Smith et al., 2009; Suanda & Namy, 2012; Vlach & Sandhofer, 2010).

Other studies have also revealed the benefits of attending to multiple contexts in word learning although they were not specifically designed to probe cross-situational word learning. In a series of studies investigating the limits of observational learning in acquiring different classes of words, Lila Gleitman and her colleagues developed a paradigm known as the *Human Simulation Paradigm* (Gillette, Gleitman, Gleitman, & Lederer, 1999; Kako, 2005; Pappafragou, Cassidy, & Gleitman, 2007; Snedeker & Gleitman, 2004) intended to simulate children's word learning using adult humans as model participants. In this paradigm, Gleitman and colleagues constructed short 30-second video clips of parents interacting with their children. The clips depicted scenes during which the parent uttered words children commonly know (e.g., “dog”, “car”, “eat”, etc.). The entire audio track of the clip was removed and a beep was inserted at the precise moment the target word was uttered. Gleitman and colleagues presented adult subjects with multiple instances of each target word and asked them to guess the reference after each instance. The pertinent result was that subjects’ identification of the target word increased with each additional context, suggesting that subjects used the similarity across contexts in identifying word meaning (Gillette et al., 1999).

Together, these findings, and others (see Vouloumanos, 2008; Vouloumanos & Werker, 2009; Yoshida, Rhemtulla, & Vouloumanos, 2012) provide evidence for the capacity to learn words cross-situationally. Further, these findings are consistent with recent claims that human learners are remarkably sensitive to the statistical properties of their linguistic environments, and that the capacity to pick up on these properties (i.e., *statistical learning*) may play an important role in a range of language tasks including phonetic processing (Maye, Werker, & Gerken, 2002), speech segmentation (Saffran, Aslin, & Newport, 1996), and syntax acquisition (e.g., Gomez & Gerken, 1999; Thompson & Newport, 2007).

However, these initial studies of cross-situational word learning primarily serve as existence proofs of the behavior rather than specific tests of the possible underlying learning mechanism. More recently, researchers have begun to shed light on the processes and factors that make this form of learning possible. One factor that has recently been shown to have an effect on cross-situational word learning is the diversity of contexts in which a given word-referent pair appears. For example, in an extension of Yu and Smith's initial paradigm that manipulated contextual diversity, Kachergis, Yu, and Shiffrin (2009) found that the greater the diversity of other word-object pairings with which a target word-referent pair co-occurred, the more likely that word-referent pairing was to be learned (see also Suanda & Namy, 2012). Interestingly, contextual diversity effects have recently been reported in children's real world word learning as well. That is, in an analysis of a large corpus of child-directed speech transcripts and order of acquisition norms for children's early words, Hills and colleagues (2010) found that the number of different word types that co-occurred with particular words predicted the order in which children typically acquire the words. Contextual diversity, and variability more generally, has been reported to have a positive effect on learning in other studies of word learning (Bolger, Balass, Landen, & Perfetti, 2008; Perry, Samuelson, Malloy, & Shiffer, 2010), as well as phonological processing (Rost & McMurray, 2009; 2010; Singh, 2008) and artificial grammar acquisition (Gomez, 2002); though this benefit of variability is in no way universal (see e.g., Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008; Vlach & Sandhofer, 2011).

A second aspect of cross-situational word learning that has recently been examined is the type of information that cross-situational word learners "compute". For example, researchers have asked whether cross-situational learners rely on simple co-occurrence frequencies between words and their referents (i.e., joint probability) or the predictive relations between words and referents (i.e., conditional probability). In a recent study, Klein and Yu (2009) modified Yu and Smith's (2007) original design and controlled for word-object frequency but varied conditional probabilities. Klein and Yu found that adult learners readily used conditional probabilities when joint probabilities were not indicative of word-to-referent mappings, a pattern of results consistent with a number of recent findings in other domains (Aslin, Saffran, & Newport, 1998; Fiser & Aslin, 2001; 2002; but see Meyer & Baldwin, 2011).

The capacity to track predictive relations may be particularly important for young word learners given that the most frequent words in the child's input as measured by corpus speech analyses (Li & Shirai, 2000; MacWhinney, 2000; Hochmann, Endress, & Melher, 2010) consist of function words such as articles (e.g., "a", "the"), conjunctions (e.g., "and", "or", "but", "yet"), and particles (e.g., "to", "not"). If word learners strictly computed which words most frequently occurred with objects and events in the environment, they would likely mismap these function words to referents of the less frequently occurring content words that follow (e.g., "car", "eat").

A third aspect of cross-situational word learning recently investigated is the extent to which the learning involved is best characterized as implicit or explicit in nature. As a first step in assessing the automaticity of learning, Suanda and Namy (2012) asked adult participants in a replication of Yu and Smith's (2007) paradigm to estimate their performance in a post-experiment interview. The results indicated that participants vastly underestimated their actual performance, consistent with previous anecdotal evidence of participants' lack of awareness of cross-situational word learning (Ichinco et al., 2009; Yu & Smith, 2007). However, participants' verbal reports *were* positively correlated with their actual learning rates, suggesting some explicit awareness of performance. This pattern of results is consistent with a recent study that examined cross-situational word learning while learners were engaged in a primary task that distracted participants from attending to the word-referent correspondences (Kachergis, Yu, & Shiffrin, 2010). Kachergis and colleagues found that although participants continued to demonstrate cross-situational word learning under these more implicit conditions, performance was poorer compared to explicit learning conditions.

Finally, in an attempt to understand the role of cross-situational learning in the context of the broader word learning literature, researchers have begun to investigate the extent to which cross-situational word learning may interact with known learning constraints from the fast mapping tradition. For example, one constraint that has been proposed to play a role in cross-situational word learning is *mutual exclusivity* (Markman & Wachtel, 1988). Briefly, the mutual exclusivity constraint refers to a word learner's default tendency to accept only one label for each object (Markman, 1990). This assumption has recently been demonstrated to contribute to cross–situational word learning as one way to limit the word-to-referent hypothesis space, guiding learners away from entertaining many-to-one or one-to-many word-referent mappings (Ichinco et al., 2009; Yurovsky & Yu, 2008). Additionally, cross-situational learners also appear to use mutual exclusivity to rule out possible referents for unknown words at the time of referent selection (Suanda & Namy, 2012; Yoshida et al., 2012), suggesting that in-the-moment learning constraints and cross-situational learning are used conjointly to determine word-referent mappings (see also Monaghan & Mattock, 2012).

The findings discussed in this section highlight the ways in which researchers have begun to move from demonstrating *that* word learners can acquire word-referent mappings via cross-situational word learning toward understanding the nature of the learning process, its connection with other word learning processes, and its relation to other areas of learning more broadly.

2.2.2 Cross-situational word learning: Computational evidence

Computational models of word learning have served as a complementary source of evidence for the role of cross-situational learning in children's lexical acquisition. Generally speaking, the computational paradigm used to study word learning involves 1) simulating behavioral patterns characteristic of children's word learning using a computer program; and then 2) examining the learning algorithms and output patterns employed by those simulations as a window into the processes underlying children's actual word learning. Through the lens of these computational models (see Regier, 2003, for a review), lexical acquisition researchers have investigated phenomena such as growth rate of vocabulary (Li, Zhao, & MacWhinney, 2007; McMurray, 2007; Plunkett, Sinha, Moller, & Strandsby, 1992; Regier, 2005), the production-comprehension asymmetry (Plunkett et al., 1992), the development of word learning biases (Colunga & Smith, 2005; Merriman, 1999; Regier, 2005; Samuelson, 2002), the emergence of fast mapping (Mayor & Plunkett, 2010; Regier, 2005) and the prototype effect in categorization (e.g., Mayor & Plunkett, 2010; Plunkett et al., 1992).

In a sense, all of these examples employ cross-situational learning in that models are typically trained with multiple instances of a word-to-referent pairing and learning is achieved through a process of accruing information. However, these models are fed the word-referent associations. Thus the researcher, not the model, resolves the mapping problem.

A particular subset of models, however, has focused on learning algorithms designed to solve the problem of referential ambiguity (Blythe, Smith, & Smith, 2010; Caza & Knott, 2012; Fazly, Alishahi, & Stevenson, 2010; Frank, Goodman, & Tenenbaum, 2009; Siskind, 1996; Yu, 2008; Yu & Ballard, 2007). These models have thus employed cross-situational word learning more akin to the sense of the term adopted in this paper. In these studies, the models are given the task of figuring out the "meaning" of words from multiple ambiguous contexts. In any given situation the models are typically presented with a set of words (e.g., "*John*", "*took*", "*the*", "*ball*") and possible meanings (e.g., the meanings [John], [took], [the], and [ball]). At the onset of training, the models do not know which word goes with which meaning (i.e., the word "*John*" could equally likely mean [John], [took], [the], or [ball]). The model's task is to figure out across situations, the meanings of the words.

What makes these cross-situational models interesting and relevant for children's word learning is not the fact that an artificial learning algorithm can identify word-referent mappings from artificially generated scenarios. Instead, what make these cross-situational models compelling are the following. First, although some cross-situational models utilize artificially constructed input (e.g., Siskind, 1996; Blythe et al., 2010; see Vogt, 2012), more recent models employ as their inputs transcriptions of child-directed speech (Fazly et al., 2010) as well as coded input representing the surrounding visual context (Frank et al., 2009; Yu & Ballard, 2007; Yu, 2008). These models suggest that, at least in principle, a cross-situational learning mechanism could acquire words from ecologically valid word-learning environments characteristic of children's word learning. Second, some models show learning rates that parallel those of child word learners (Blythe et al., 2010; Siskind, 1996). For example, Siskind's model is comparable to child word learners in the amount of input (i.e., number of utterances) needed to acquire a lexicon of 10,000 words, based on estimations from observational studies (Snow, 1977 as cited in Siskind, 1996). Third, these models demonstrate a number of signature characteristics of children's word learning. These include the *vocabulary spurt* (a relatively flat learning curve followed by a steeper one; see Siskind, 1996), *mutual exclusivity behavior* (preference to attach a novel label onto a novel object as opposed to a familiar one; see Frank et al., 2009), and *synonym learning* (learning a second word for a meaning already associated with another word; see Fazly et al., 2010, Siskind, 1996).

Together, the behavioral and computational findings discussed in the above sections provide some recent evidence for a cross-situational learning approach to the mapping problem. However, the studies discussed are recent and limited in scope relative to the number of investigations devoted to children's fast mapping capabilities. As a result, cross-situational word learning has played a relatively small role in contemporary theories of children's word learning. In the remaining sections of this chapter, I raise two limitations of the recent studies of cross-situational word learning, which will be the focus of the experimental studies that follow.

2.3. Outstanding Questions

2.3.1 Scaling from adult word learning to children's word learning

The vast majority of the studies of cross-situational word learning have employed adult learners. This reliance on adult subjects in artificial language learning studies has a rich history in language acquisition research (see Gomez & Gerken, 2000, for discussion). However, given that the primary phenomena of interest with respect to cross-situational word learning occurs in infancy and childhood, it is also critical for researchers to begin to extend their empirical investigations to developmental populations.

Towards this goal, Smith and Yu (2008) recently designed a version of their adult cross-situational learning paradigm suitable for testing infant learners. Employing a simplified looking-based version of the task, Smith and Yu found that 12- to 14-month-old infants successfully associated words and their corresponding objects in a task that, like its adult precursor, required them to attend to the co-occurrence statistics across situations (Smith & Yu, 2008, in press; Yu & Smith, 2011). Clearly, studies such as Smith and Yu's (see also Vouloumanos & Werker, 2009) help bridge the findings based on adult samples and developmental populations. However to date, most of the work that probes into the underlying learning mechanism of cross-situational learning (e.g., Ichinco et al., 2009; Kachergis et al., 2010; Klein & Yu, 2009; K. Smith et al., 2009; Suanda & Namy, 2011) have relied solely on adult models.

In part this is due to the fact that such studies involve complex paradigms that are beyond what current infant research methodologies would allow, given the small stimulus sets required to accommodate infants' limited attention spans. I suggest that one possible way to bridge the gap between the findings based on adult samples and the generalization to infant learners is to supplement the infant studies with artificial learning tasks with older children (i.e., 5- to 7-year-olds). These children are old enough to complete artificial language learning tasks that are closer in complexity to those used with adult learners. At the same time, these children are obviously closer in age to infant language learners than adults and thus may serve as better models of early language learning.

A handful of studies across various aspects of language acquisition have adopted this approach. In a number of cases, researchers have found very little difference between adults' and children's performance (Saffran, 2001; 2002; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Others, however, have identified conditions under which adults and children perform differently. For example, in a series of studies on artificial grammar learning, Hudson Kam and Newport (2005; 2009) found that 6- to 7-year-olds tend to over-generalize a probabilistically prevalent grammatical pattern, whereas adults tend to distribute their patterns in a manner that matches the probability with which each grammatical patterns occurred in the input. Finally, some researchers have proposed that the differences between adults and children are mainly quantitative, rather than qualitative, in nature, likely reflecting simply increases in information processing capacity over development (e.g., Braine et al., 1990; Ferman & Karni, 2010; Janacsek, Fiser, & Nemeth, 2012; Piccin & Waxman, 2007).

Thus, although early school-aged children are far from equivalent to novice language learners, their inherently more limited memory and attentional capacities may nonetheless lead to different patterns of performance compared to adult learners. This approach has the potential to assess whether some of the cross-situational learning findings gleaned from adult learning tasks are operating in development.

2.3.2 The nature of the underlying learning mechanisms

A second issue that warrants further investigation is the nature of the mechanism that underlies cross-situational word learning. As described above (see *Section 2.2.2*), numerous formal and computational models have been proposed that suggest that word learning involves aggregating information across situations. These extant models vary greatly in the detailed mechanics of their operations and thus comparing between them can be difficult. At a high level of abstraction however, most of them can be classified as falling into two broad classes of learning processes: *hypothesis testing* models and *associative learning* models. In hypothesis testing models (e.g., Siskind, 1996), cross-situational word learning is conceptualized as a process through which the learner selects specific word-to-referent mappings from a pre-defined set of possible mappings. Over the course of learning, mappings are either confirmed or rejected and replaced based on additional input, depending on the consistency across naming events. The outcome of learning in these accounts is conceptualized as a hypothesis list, a set of definitive mappings between a specific word and a specific referent. A separate class of models (e.g., Yu, 2008) suggests a more basic associative learning account of cross-situational word learning. According to this account, a word is linked to multiple candidate referents each time that word is uttered. Over time, the learner develops a large associative network consisting of connections between multiple words and multiple referents. The strength of each connection is proportional to the regularity with which the relevant word and referent co-occur. Thus, in associative learning accounts, the outcome of learning is not a definitive mapping between a word and its referent but rather multiple probabilistic word-to-referent links that vary in their associative strength.

Recently, there has been great debate over which of these models best account for the current data (Nicol Medina, Snedeker, Trueswell, & Gleitman, 2011; K. Smith, Smith, & Blythe, 2010; Yu & Smith, 2012). One limitation to this current debate, and to many extant process-models of cross-situational word learning more generally, is that it is based largely on analysis of adult learner behavioral data (for one notable exception, see Yu & Smith, 2011). As a result, its relevance to the mechanisms of children's cross-situational word learning is an open question. Indeed some scholars have hypothesized that relative to adult learners, young children may exhibit more associative characteristics whereas later learning may take more of a hypothesis testing form (e.g., Smith & Yu, 2008). Thus, in addition to elucidating the general mechanics of children's cross-situational word learning, computational analyses of developmental data may also shed light onto the associative learning – hypothesis testing debate.

In the chapters that follow, I present a series of studies with the goal of addressing these two outstanding issues. In Chapters 3 and 4, I ask whether young school-aged children possess the cross-situational word learning prowess previously observed in mature adult learners. The study in Chapter 3 provides an initial test of children's cross-situational learning capacities. The study also represents a first step in understanding the constellation of factors that impact children's learning by examining the extent to which the contextual diversity of the learning environment, a factor known to influence adults' learning (Kachergis et al., 2009; Suanda & Namy, 2012), influences children's learning. The study reported in Chapter 4 probes further children's learning by testing the precision of the word-to-referent mappings made across ambiguous naming events. In Chapter 5, I present a series of simulation studies that explore the candidate mechanisms underlying children's cross-situational word learning. Specifically, I construct two computational models that instantiate the principles of hypothesis testing and associative learning respectively, and test simulations of these models performing artificial versions of the tasks reported in Chapters 3 and 4. Of interest is the extent to which, and under what conditions, each model predicts children's observed learning patterns. Finally, in Chapter 6, I end this dissertation with a broad discussion of the role of cross-situational learning in accounts of children's word learning.

Chapter 3. The Effect of Contextual Diversity on Children's Cross-Situational Word Learning

3.1 Background

In this chapter, I report the results of an experiment on cross-situational word learning in school-aged children. Although cross-situational statistical word learning has received much attention recently as a potentially important component of children's vocabulary growth, much of the current research on this type of learning has been conducted with adult learners (Fitneva & Christiansen, 2011; Ichinco et al., 2009; Kachergis et al., 2009; Kachergis, Shiffrin, & Yu, 2012a; Klein & Yu, 2009; Klein & Yu, 2009; K. Smith et al., 2009; K. Smith et al., 2010; Suanda & Namy, 2012; Vlach & Sandhofer, 2011; Vouloumanos, 2008; Yu & Smith, 2007; Yu, Zhong, & Fricker, 2012; Yurovsky & Yu, 2008; Yurovsky, Fricker, Yu, & Smith, 2010). Important exceptions do exist (Akhtar & Montague, 1999; Smith & Yu, 2008, in press; Vouloumanos & Werker, 2009; Yu & Smith, 2011). For example, in one recent study, Smith and Yu (2008) demonstrated that in a simplified version of the adult learning paradigm, 12- and 14-month-old infants were able to map words onto their referents when the only cue to reference was the word-to-referent cross-situational co-occurrence statistics. The finding that even young word learners possess the capacity for cross-situational word learning is important because it is an existence proof for the claim that a process such as cross-situational word learning can get early lexical acquisition off the ground.

In the current experiment, I examine cross-situational word learning in an older population of children, between the ages of 5 and 7. There are two reasons why this age group is an interesting population for studying lexical development. First, this is a period of development in which children are very much in the process of building their vocabulary. In fact, the rate of vocabulary growth during middle childhood is greater than during late infancy and toddlerhood, the period typicaly emphasized in word learning theory (for discussion, see Anglin, 1993; P. Bloom, 2000; Snedeker, 2009). Thus an understanding of cross-situational word learning during this period of vocabulary development has the potential to inform not only the role of cross-situational learning in later stages of lexical development, but also the constellation of learning processes that support prolific word learning more generally.

A second reason to study cross-situational word learning in older children is that older children may be an effective population to test the extent to which accounts of the nature of cross-situational word learning developed from research with adult learners (e.g., K. Smith et al., 2010; Yu & Smith, 2007) are applicable to developmental populations. That is, school-aged children are mature enough to complete a range of tasks that are commonly used with adults but, for methodological reasons, are difficult to implement in infant populations. At the same time, school-aged children still possess more limited attentional, memory, and general cognitive capacities, and thus can address whether more limited learners exhibit similar learning patterns as mature learners. Indeed, although a number of researchers have found similarities between adult and child learners within artificial language learning tasks (Meuleumans, Van der Linden, & Perruchet, 1998; Saffran, 2001; 2002; Saffran et al., 1997), other researchers have found both quantitative and qualitative developmental differences (e.g., Braine et al., 1990; Ferman & Karni, 2010; Hudon Kam & Newport, 2005; 2009; Piccin & Waxman, 2007).

To test cross-situational word learning capacities in school-aged children, I have adapted Yu and Smith's adult cross-situational word learning paradigm (Yu & Smith, 2007) to render the task suitable for young children. As in the adult paradigm, children encounter ambiguous naming events in which they see multiple pictures of objects and hear multiple words with no disambiguating information regarding which word refers to which picture. Across situations, words and their referents always co-occur together while the accompanying word-referent pairings vary. Thus the logic behind this paradigm is that children can figure out word reference only if they are able to utilize the cross-situational co-occurrence information.

Given that there is referential ambiguity present on every single trial, success in Yu and Smith's paradigm has been interpreted as evidence for cross-situational word learning. However, K. Smith and colleagues have recently proposed an alternative explanation for success in the Yu and Smith paradigm (K. Smith et al., 2009). They argue that a learner who simply keeps track of the set of words and objects present during a *single* learning trial could successfully demonstrate learning. That is, imagine a hypothetical single-trial learner who is presented with the three ambiguous naming events in Figure 1A. Further, imagine that this learner only encodes the final trial in the sequence. From this trial alone, the learner would know that the words "hiplex" and "bemkin" go with either the object on the left (yellow tool) or the object on the right (chandelier). As K. Smith and colleagues pointed out, this knowledge alone would be sufficient to respond correctly on the test trial presented in Figure 1B because only one of the candidate objects is present at test. K. Smith and colleagues supplemented their critique with a formal analysis revealing that the levels of adult learning in Yu and Smith's original task were not above the levels that would be predicted through single-trial learning, suggesting that Yu and Smith may have overestimated learners' abilities to track cross-situational co-occurrence information (K. Smith et al., 2009).

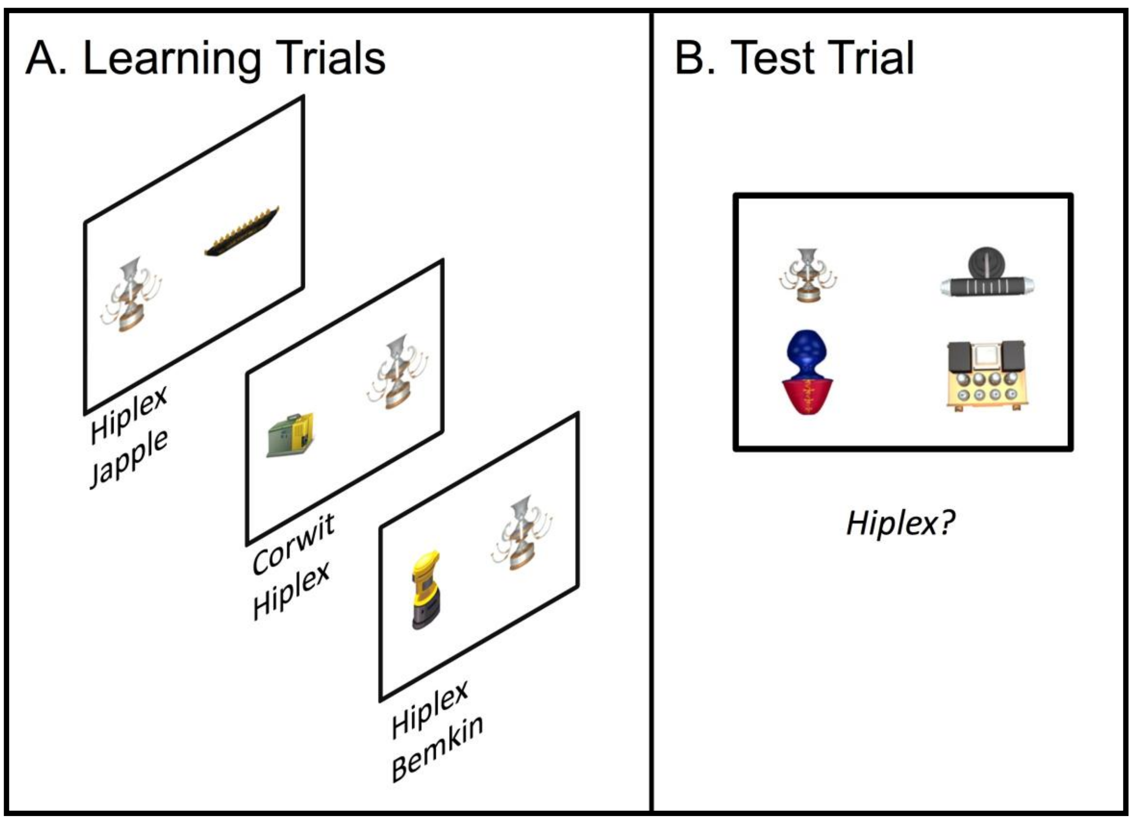

Figure 1. *A sample of a series of learning trials containing the word "hiplex" and its referent (A). A 4-Alternative-forced-choice (4AFC) task testing learning of the word "hiplex" (B).*

Given K. Smith and colleagues' critique, coupled with evidence suggesting that relative to adults, children may be less likely to aggregate information across trials (e.g., Piccin & Waxman, 2007), the current paradigm includes a manipulation that sheds light on the extent to which children employ a single-trial learning strategy or a truly cross-situational one. Specifically, children participated in one of three learning conditions that differed in the contextual diversity of the learning environment. Contextual diversity here refers to the different word-object pairings with which a particular word-object pair co-occurs; the difference between a high and low diversity context is illustrated in Figure 2. To illustrate how manipulating contextual diversity sheds light on the underlying learning strategy, consider single-trial learners in both high and low contextual diversity conditions, who randomly select a single observation of the word "hiplex" (e.g., the last trial). Across both conditions of contextual diversity, the single trial learner should perform identically. That is, nothing about a single encoding from a high diversity context makes learning more or less difficult than a single encoding from a low diversity context. In contrast, imagine a cross-situational learner who attempts to figure out the referent of the word "hiplex" by tracking information across trials. This learner may find conditions of high contextual diversity more conducive to learning because there are more opportunities to disambiguate "hiplex"'s referent. Indeed research with adult learners in this paradigm (Kachergis et al., 2009; Suanda & Namy, 2012) has found that more diversity leads to better learning. To the extent that children also demonstrate an effect of contextual diversity on learning this would suggest a strategy that involves combining co-occurrence information across situations.

In the following experiment, I randomly assigned children to one of three conditions of contextual diversity: high contextual diversity, moderate contextual diversity, and low contextual diversity. To ensure that any difference between conditions would be attributable to contextual diversity, I tested all children's word learning under identical conditions. To the extent that I find an effect of contextual diversity on learning, this would rule out a single-trial learning explanation of performance and would suggest that learning at this age is driven by a cross-situational learning strategy.

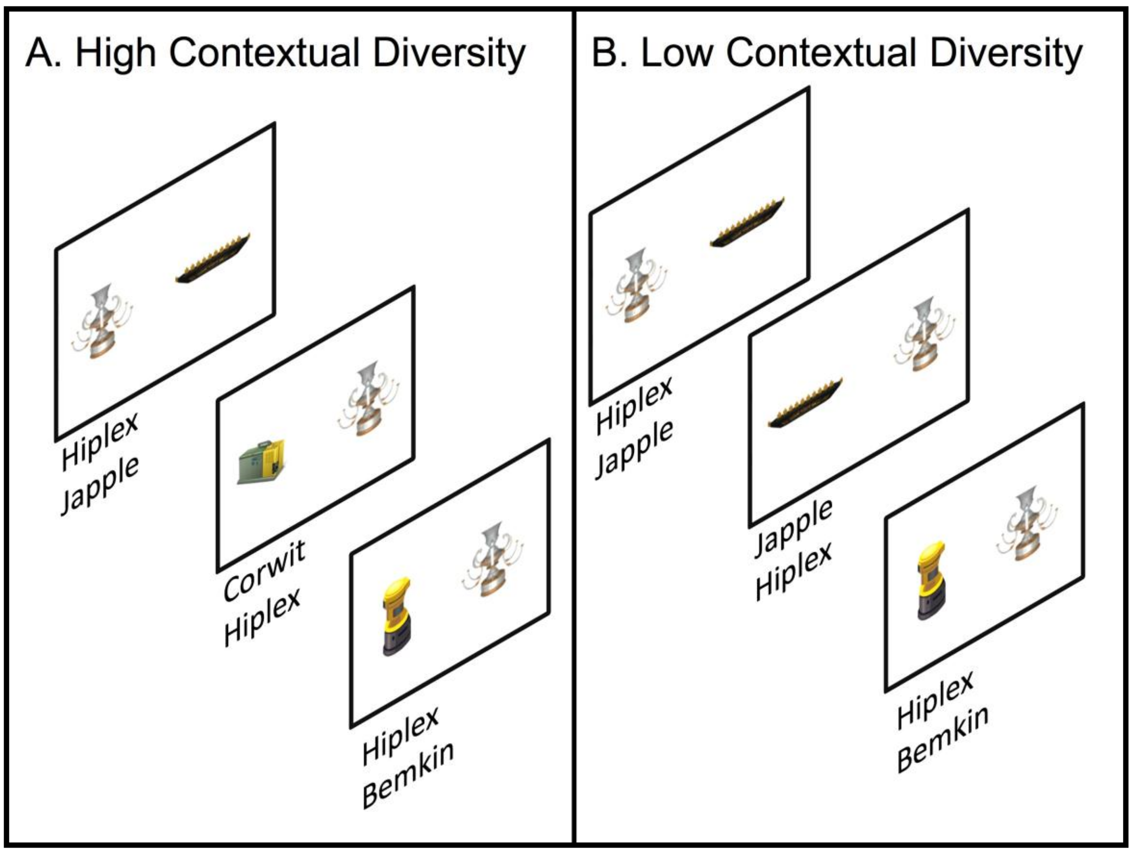

Figure 2. *A sample of learning trials under conditions of high contextual diversity (A). A sample of learning trials under conditions of low contextual diversity (B).*

3.2 Method

3.2.1 Participants

Eighty-four 5- to 7-year-olds (*Mean* age = 73.6 mos, *Range* = 57.3 - 94.9) from the greater Atlanta area participated. Forty-nine children were female. Participants were children of families who had volunteered to participate in studies at Emory University's Child Studies Center. 77% of children were Caucasian, 19% were African-American, 2% were Asian, and 1% were of other racial categories. 9% of families identified as Hispanic or Latino. An additional 13 participants were excluded from data analysis due to exhibiting a position bias. See *Coding* (section 3.2.5) below for details on exclusion criteria.

3.2.2 Stimuli

Stimuli were eight recorded bi-syllabic novel words (e.g., "*blicket*") and eight pictures of uncommon or artificially altered objects (e.g., a phototube). The same female speaker recorded all the words using adult-directed speech and neutral prosody. Each word was paired with a picture to create eight to-be-learned word-object pairings. Initial pilot data revealed that children exhibited no bias towards learning any particular word-object mapping. Four additional novel word-object pairings were used for task familiarization. The full stimulus set used in the experiment proper is displayed in Figure 3 below. Stimuli were incorporated into a computer application that was created in-house using Real Basic software (REALbasic, 2008) and used to control stimulus presentation on a 17-in. monitor connected to a Power Mac G5. An add-on touch screen (Magic Touch, 2009) was mounted onto the monitor to allow children to advance trials and make selections at test.

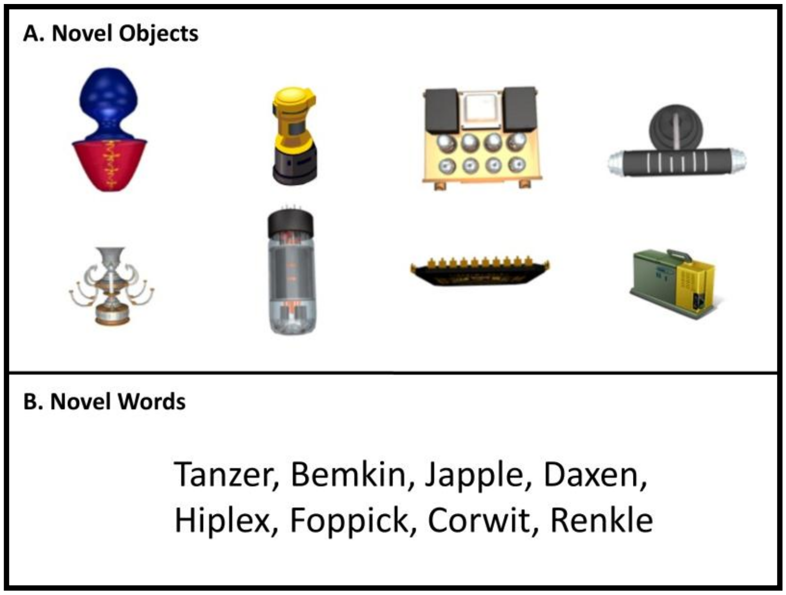

Figure 3. *Stimuli employed in the experiment: pictures of novel/altered objects used (A); orthographic representation of novel spoken words used (B).*

3.2.3 Design

Children were randomly assigned to one of three conditions varying in the contextual diversity of the learning environment: High Contextual Diversity (High CD), Moderate Contextual Diversity (Moderate CD), and Low Contextual Diversity (Low CD). The association matrices in Table 1 illustrate the total frequencies with which words (columns) co-occur with different pictures (rows) in each condition. In all three conditions, a word co-occurred with its referent in a total of four trials. In each learning trial, a second word and picture was also presented, with the correspondence between words and pictures ambiguous on any given trial. Conditions differed in the number of *different* distractors with which a word co-occurred across the four trials, as well as the *frequency* with which a given distractor co-occurred with a word. To illustrate, in Table 1, Word 1 (W1) co-occurs with its referent (Picture 1-P1) on four trials throughout learning. W1-P1 is accompanied by

Table 1. *Association matrices representing word – picture co-occurrence frequencies across conditions.*

A. High CD

Pictures \ Words	1	2	3	4	5	6	7	8
1	4	1	1	1	1			
2	1	4	1	1		1		
3	1	1	4	1			1	
4	1	1	1	4				1
5	1				4	1	1	1
6		1			1	4	1	1
7			1		1	1	4	1
8				1	1	1	1	4

B. Moderate CD

Pictures \ Words	1	2	3	4	5	6	7	8
1	4	2	1	1				
2	2	4	1	1				
3	1	1	4	2				
4	1	1	2	4				
5					4	2	1	1
6					2	4	1	1
7					1	1	4	2
8					1	1	2	4

C. Low CD

Pictures \ Words	1	2	3	4	5	6	7	8
1	4	3	1					
2	3	4		1				
3	1		4	3				
4		1	3	4				
5					4	3	1	
6					3	4		1
7					1		4	3
8						1	3	4

W2-P2 on one of those trials, W3-P3 on a different trial, W4-P4 on another trial, and W5-P5 on yet another trial, resulting in maximal contextual diversity. In the Moderate CD condition, word-picture pairings will co-occur with one word-picture pairing on two trials, and two other word-picture pairings on the other two trials, resulting in less diversity across trials. Finally, in the Low CD condition, word-picture pairings co-occurred with one word-picture pairing three times, and another word-picture pairing once, so the diversity across trials is low.

3.2.4 Procedure

Children sat in front of the touch-screen computer next to the experimenter with a video camera positioned over children's shoulders (see Figure 4). The experimenter employed a ladybug puppet named "Lulu the Ladybug" and introduced the experiment as a game with the goal of learning Lulu's names for her favorite toys. Children completed a familiarization phase followed by a learning and test phase.

Figure 4. *Image still from a participant depicting the experimental session layout.*

3.2.4.1 Familiarization phase

The goal of the familiarization phase was to introduce children to the experimental setting and to the general goal of learning Lulu's names for her toys. This procedure has been thoroughly piloted and adapted to optimize task comprehensibility for children. There are three parts to the familiarization phase. First, the twelve pictures (the 8 to-be learned pictures and 4 additional pictures) were displayed simultaneously on the computer screen. Then, the twelve novel words that corresponded to each of the pictures were played in a random order. The experimenter then told children, "We are going to learn which name goes with which picture". In the second step of the familiarization phase, the experimenter explicitly taught children the name of two of the 12 pictures. One novel picture was presented on the computer screen and its corresponding word was played. This was repeated for a second novel picture. Then, children's learning of those two words was tested in two four-alternative forced-choice (4AFC) trials. In these trials, the same four pictures (the two labeled pictures and two unlabeled pictures) appeared simultaneously on the computer screen. In the first trial, the first novel word was played and children were asked to make a choice by touching the picture they thought went with the word. A second 4AFC trial tested children's learning of the referent of the second novel word. Correct selections were reinforced by the experimenter's clapping and a rewarding audio clip (applause and cheering). For any incorrect selections, children were asked to make a different selection until they were correct. The two novel word-picture pairings taught during the familiarization phase as well as the two distractors present in the 4AFC test trials of the familiarization phase did not appear during the experiment proper. The goal of the familiarization phase was simply to familiarize children with the game of learning words. Given that there was no referential ambiguity in the learning of words during this phase, the familiarization phase was not likely to "train" children on how to learn words cross-situationally. However, it did familiarize children with the experimental setting, the touch screen, the 4AFC task and the goal of the task, to learn words.

3.2.4.2 Learning phase

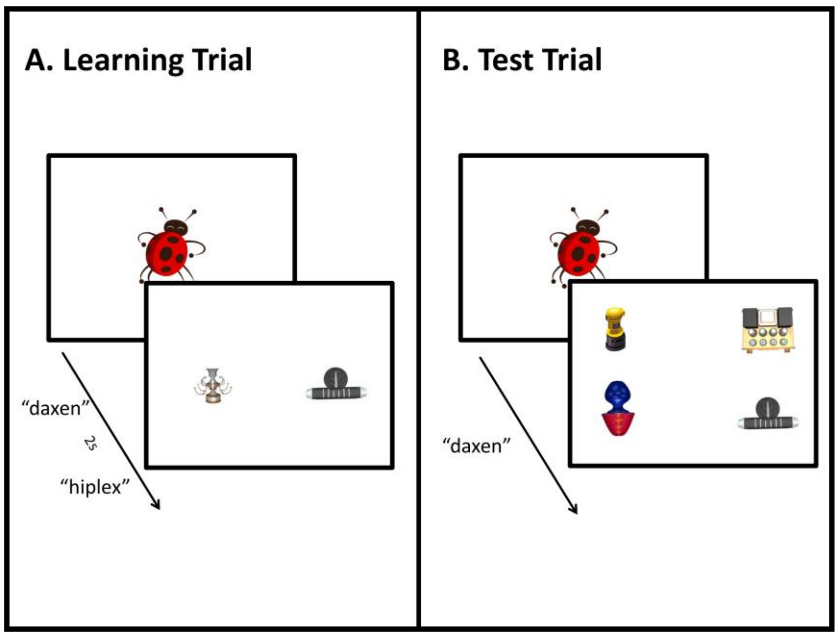

Figure 5. *Sample trial structure in the experiment: a learning trial (A); a test trial (B).*

Following the familiarization phase, children proceeded immediately to the learning phase. At this time, the experimenter said, "Now, we are going to learn all of the names of Lulu's other toys". In each trial of the learning phase, children saw two pictures, one on each side of the monitor. Children also heard two spoken words, played sequentially in random order, corresponding to the two pictures (see Figure 5A). The use of two picture-word pairings on each learning trial, which is the simplest (i.e., lowest amount of referential uncertainty) version employed in Yu and Smith's adult paradigm (Yu & Smith, 2007), was selected to minimize the task demands for children.

Each of the eight to-be-learned word-object pairings occurred on four trials throughout the learning phase. Given that these 32 instances of word-object pairings were presented two at a time on each trial, the learning phase consisted of 16 total learning trials. Two training lists were created with the order of the trials pseudo-randomized such that each of the eight word-object pairings appeared once before any given pairing was repeated and no word-object pairing appeared in back-to-back trials, consistent with Yu and Smith's original adult paradigm. Which training list was used was counterbalanced across participants.

3.2.4.3 Test phase.

The test phase immediately followed the learning phase and consisted of eight 4AFC test trials, one per target word. In each trial, four pictures appeared simultaneously, one in each quadrant, followed by the presentation of the target word (see Figure 5B). The child indicated which picture she thought went with the target word by touching the picture on the screen. Children received no feedback during the test phase. All test trials were constructed by selecting the target word's corresponding picture and 3 pseudo-randomly selected foils that had never co-occurred with the target word during the learning phase. All pictures served as foils an equal number of times. Two test lists were created and were identical across conditions. Which test list was employed was counterbalanced across participants.

Since none of the foils had co-occurred with the target word during the learning phase, this testing regimen was not designed to detect whether children definitively mapped the target word to its correct referent. Instead, this procedure simply tested children's sensitivity to whether the word had co-occurred with the target picture at all during the learning phase. Although this testing regimen did not provide the most rigorous test of word mapping *within* each condition, it did provide a straightforward way to assess the effect of contextual diversity *across* conditions. A follow-up experiment (see Chapter 4) was designed more specifically to address the definitiveness of children's word mappings.

3.2.5 Coding

For each test trial, children's choices were automatically registered as correct or incorrect. Children were considered as exhibiting a position bias and excluded from the analysis if they selected the object located in the same quadrant on five or more of the eight trials. This cut-off was chosen because it is the point at which the probability of selecting a single quadrant across 8 independent trials is statistically greater, $p < .05$, than would be predicted by chance responding.

Because the experimental procedure involved some interaction between experimenter and child, it was important to ensure that the experimenter (a) did not inadvertently cue children to the correct answer, and (b) conducted sessions identically across conditions. To ensure that the experimenter was not providing any cues at test, two coders blind to experimental condition and blind to the correct answer watched each test trial of each child and judged whether they believed the child had gotten the trial correct or incorrect. The logic behind this coding is that if the experimenter was providing some cue or feedback to the child, the coders should be able to reliably discriminate a child's correct from incorrect trials. To test whether coders were able to detect correct from incorrect trials, I calculated a Cohen's Kappa coefficient (K) to measure the agreement between each coder's conjecture and children's actual performance[1]. Coders' guesses showed very little agreement with children's actual performance ($K_{coder\ 1}$ = .07, $K_{coder\ 2}$ = .07). There was also little agreement between the two coders' conjectures (K = .12). This suggests that there were no obvious observable cues that may have influenced children's testing performance.

To ensure cross-condition consistency in experimenter protocol, a coder blind to experimental condition watched the entire learning phase for each child and rated the experimenter's enthusiasm during the session (from 1 to 7). Mean enthusiasm rating was similar across all three conditions (M_{High}= 4.26, M_{Mod}= 4.29, M_{Low}= 4.21). A second coder also blind to experimental condition watched 25% of the sessions. Agreement between the two coders on enthusiasm ratings was high, K = .78[2].

3.3 Results

For each child, I computed the proportion of test trials answered correctly. I then derived a mean proportion correct for each condition of contextual diversity. Initial analyses within each condition revealed no effects of training list, testing list,

[1] The Kappa coefficient was chosen over other measures (e.g., percent correct) to take into consideration baseline differences between how often coders believed children answered correctly and how often children actually answered correctly.
[2] For reliability calculations, enthusiasm ratings were reduced to a three-point scale. On a seven-point scale, 83% of the two coders' ratings were within one point of each other.

or interaction, on mean proportion correct, smallest $p = .10$. Thus all subsequent analyses were collapsed across training and testing lists. To explore any effects of age on performance, I examined, within each condition, the correlation between age and mean proportion correct. I found no correlation between age and performance in any condition ($r_{\text{High}} = .12$, $r_{\text{Mod}} = .15$, $r_{\text{Low}} = .006$; smallest $p = .43$). There were no sex differences in performance, $p = .30$.

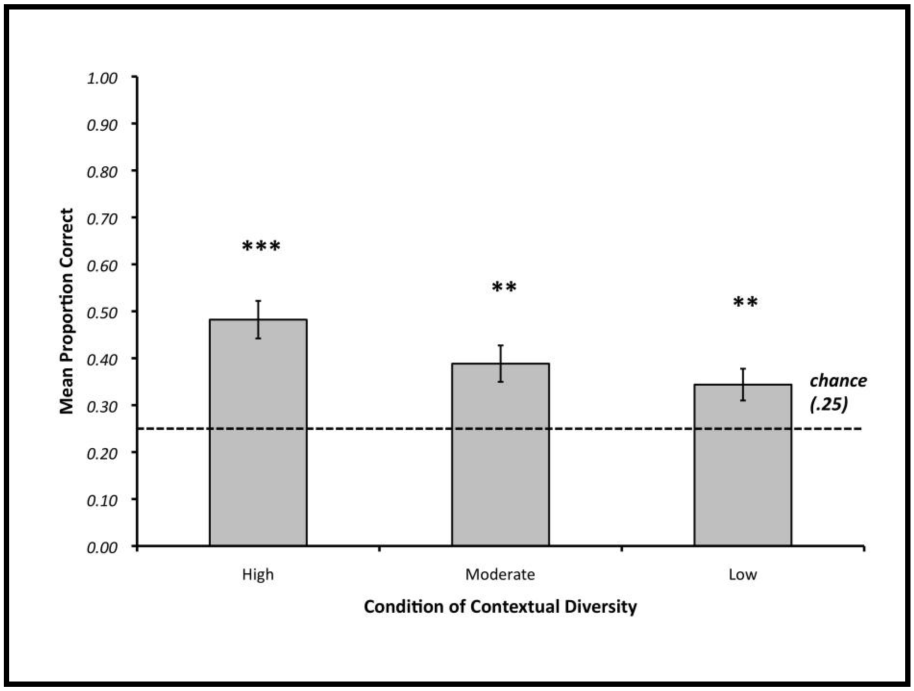

Figure 6. *Mean proportion correct across levels of contextual diversity. Note: *** p <.001, ** p < .01*

Figure 6 shows the mean proportion correct across conditions of contextual diversity ($M_{High} = .48$, $SD_{High} = .21$; $M_{Mod} = .39$, $SD_{Mod} = .20$; $M_{Low} = .34$, $SD_{Low} = .18$). To test whether children demonstrated word learning, for each condition I conducted a series of single-sample t-tests to examine whether mean proportion correct was above the proportion that would be expected from chance performance (i.e., random guessing in a 4AFC trial, .25) in each condition. These tests revealed that learning in each condition was significantly above chance performance ($t_{High}(27)$ = 5.80, d_{High} = 1.10; $t_{Mod}(27)$ = 3.57, d_{Mod} = .67; $t_{Low}(27)$ = 2.77, $d_{Low}(27)$ = .52, all p's <.01). This finding underscores the power of children's cross-situational learning: from only a handful of ambiguous naming events, children mapped words to their referents even when confronted with low levels of contextual diversity.

To investigate the effect of contextual diversity on learning, I conducted a one-way analysis of variance (ANOVA) on mean proportion correct with condition as a between-subjects factor. As depicted by the downward trend in mean proportion correct across conditions in Figure 6, contextual diversity had a significant effect on performance, $F(2, 81) = 3.53$, $p = .03$, $\eta^2 = .08$. Planned comparisons with Bonferroni correction revealed that the only statistically significant pair-wise difference, was between the High CD and Low CD conditions, p = .03. The difference between the High CD and Moderate CD conditions, p = .23, or Moderate CD and Low CD conditions, p = .28, did not reach statistical significance. The finding that increased contextual diversity improves cross-situational word learning is consistent with research with adult learners in a similar paradigm (Kachergis et al., 2009; Suanda & Namy, 2012). This finding is particularly informative because it rules out a single-trial learning explanation of children's task performance and underscores that children's learning must have emerged from a process of tracking word-picture associations across situations.

To investigate how representative these group-level results were of individual children's performance, I examined individual patterns of performance, dichotomizing children in each condition as either performing above chance or performing at/below chance. Figure 7 illustrates the proportion of children across conditions that performed above chance. The main patterns of the group level analyses were upheld at the individual level. That is, the proportion of children performing above chance in each condition ($Prop_{High}$ = .786, $Prop_{Mod}$ = .643, $Prop_{Low}$ = .500) was statistically greater than the proportion that would be predicted from chance performance (.321)[3], $p_{High} < .001$, $p_{Mod} < .001$, $p_{Low} = .04$. Further, a chi-square test of independence revealed a marginally significant effect of contextual diversity on individual level performance, $\chi^2 = 4.97$, $p = .08$.

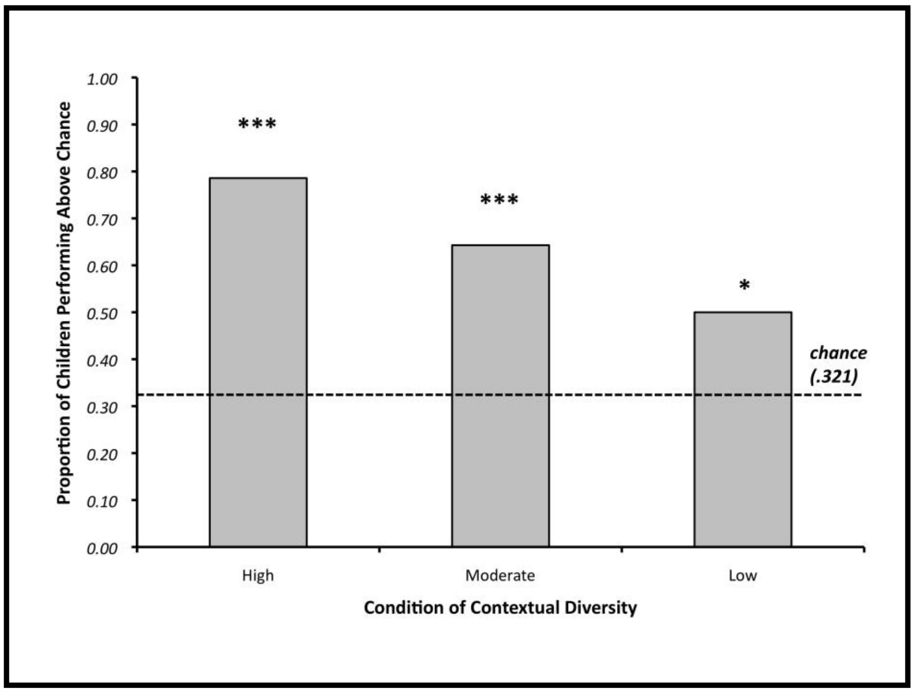

Figure 7. *Proportion of children performing above chance level (.25) across levels of contextual diversity. Note: ***p < .001, *p < .05.*

---

[3] Chance for this analysis was defined as the proportion of times one would expect to see a participant perform above chance (.25) if participants randomly guessed on each trial. This chance value was derived from multiplying the likelihood of each response pattern given random guessing (i.e., the likelihood of getting exactly 3 trials correct, 4 trials correct, etc.) and the number of permutations of each response pattern that is above chance (e.g., there are exactly 56 different ways in which a random performer would get exactly 3 trials correct = trials 1,2,3 are correct, trials 1,2,4 are correct, trials 1, 2, 5 are correct, etc.).

As Figure 7 illustrates, the marginal difference revealed by the individual-level analysis is consistent with that revealed by the group-level data, namely that the largest difference is between proportion of children performing above chance in the High CD condition and the proportion of children performing above chance in the Low CD condition.

3.4 Discussion

In the current experiment, 6-year-old children learned word-to-referent mappings from just a handful of ambiguous naming events. Although children demonstrated learning across all conditions of contextual diversity, children's learning patterns clearly indicated that the more diverse the learning contexts, the better the learning. These findings are consistent with previous findings suggesting that infants, toddlers and adults are prodigious cross-situational word learners (Scott & Fisher, 2012; Smith & Yu, 2008; Yu & Smith, 2007) and that contextual diversity influences performance in adult learners (Kachergis et al., 2009; Suanda & Namy, 2012). This latter finding begins to shed light on the underlying learning strategy children employed in this task. Although the specific mechanism remains unclear, that contextual diversity impacts learning rules out a single-trial learning strategy as a candidate process, and suggests that the process is one that involves combining co-occurrence information across situations. In what follows, I discuss three implications and issues raised by the present results: (a) the role of cross-situational word learning in school children's vocabulary growth; (b) similarities and differences in adult and child statistical learning; and (c) contextual diversity effects in learning

3.4.1 Cross-situational word learning in children's vocabulary growth

Most developmental research on cross-situational word learning has focused on infant and toddler populations (Akhtar & Montague, 1999; Scott & Fisher, 2012; Smith & Yu, 2008; in press; Yu & Smith, 2011; but see Piccin & Waxman, 2007; Werner & Kaplan, 1950, for notable exceptions). This focus on young word learners is warranted: to the extent that cross-situational word learning is a viable candidate process that gets word learning off the ground, it is important to demonstrate the availability of this learning process in the youngest of word learners. But what about the role of cross-situational word learning in later vocabulary development? According to some theorists (e.g., Golinkoff & Hirsh-Pasek, 2006; Nazzi & Bertoncini, 2003) early word learning is qualitatively different from later learning. Whereas early learning (i.e., prior to 18- to 24-months) is driven by associative learning mechanisms that are characteristically slow and effortful, later learning is driven by sophisticated social and cognitive learning processes that are characteristically fast and effortless. Thus, perhaps cross-situational word learning plays its primary role in early lexical acquisition. Alternatively, cross-situational word learning may continue to play an important role in later word learning as well. Two features of vocabulary acquisition during middle childhood support this latter view.

First, the contexts in which most words are learned during middle childhood appear to be particularly conducive to a cross-situational word learning strategy. That is, many words learned during this phase are acquired unintentionally through reading (Gordon, Schumm, Coffland, & Doucette, 1992; Nagy, Herman, & Anderson, 1987; Nagy, Herman, & Anderson 1985; Shu, Anderson, & Zhang, 1995) or listening

(Elley, 1989; Robbins & Ehri, 1994) contexts. This type of word learning is known as *incidental word learning* in educational and reading studies (see Swanborn & de Glopper, 1999, for review). Evidence suggests that incidental word learning exhibits two patterns suggestive of a cross-situational learning process: (A) the type of word knowledge children acquire from these contexts are often only fragments of children's eventual word knowledge (e.g., Nagy, et al., 1985, 1987; Schwanenflugel, Stahl, & McFalls, 1997); (B) researchers have found that children benefit from multiple exposures to words, and that a single exposure is rarely sufficient for learning (Horst, Parsons, & Bryan, 2011; Jenkins, Stein, & Wysocki, 1984; Robbins & Ehri, 1994). Thus, the ability to track probabilistic relations between words and potential meanings across multiple contexts, like the ability demonstrated by the children in the current study, seem well suited for incidental word learning.

A second finding about the nature of vocabulary development during middle childhood that suggests a role for cross-situational word learning in later development is the *type* of words acquired during this phase in development. In an in-depth analysis of the vocabularies of 6-, 8-, and 10-year-olds, Anglin (1993) found that much of the increase in vocabulary size during this period is accounted for by a large increase of derivative words (e.g., "advisable", "competitive"). Specifically, Anglin found that whereas derived words account for only 16% of a 6-year-olds' vocabulary, they account for 39% of a 10-year-olds'. Anglin's findings are consistent with other findings suggesting that derivational morphology is a relatively late development (e.g., Freyd & Baron, 1982; Wysocki & Jenkins, 1987) and one that plays an important role in children's vocabulary growth (e.g., McBride-Chang, Wagner, Muse, Chow, & Shu, 2005; McBride-Chang et al., 2008). The type of cross-situational learning demonstrated by the children in the current study may help explain how children could come to use morphological patterns to derive the meaning of newly encountered derivative words. That is, perhaps as some have argued (e.g., Goldberg, 1999; Tomasello, 2003), children initially learn words with common affixes (e.g., doable, fixable, understandable) individually. As children continue to hear other similarly structured words (e.g., drinkable), they begin to notice consistent relations between the affix and its underlying semantic and syntactic effects. Once children have learned the general rule, they can then implement these patterns when confronted with new instances (e.g., cur-*able*).

Further research is needed to investigate the extent to which the findings from the study of incidental word learning and *morphological problem solving* (Anglin, 1993) is linked to the cross-situational learning work reported in the current study. Such research would not only shed light on the role of cross-situational word learning beyond the earliest stages of vocabulary development, it would also bring together approaches to children's vocabulary acquisition that are currently investigated separately.

3.4.2 Statistical learning in adult and child learners

In the current study, I find that children, like adults (e.g., Yu & Smith, 2007), can rapidly learn words despite referential uncertainty within naming events. Additionally, similar to adults' learning (Kachergis et al., 2009; Suanda & Namy, 2012), children's learning is positively impacted by contextually diverse learning conditions. Thus, in conjunction with the research on adult cross-situational word learning, the current findings are in accord with other statistical language learning studies in other domains (Meulemans et al., 1998; Saffran et al., 1997) that find similarities between adults' and children's statistical learning.

The extent to which adult and child learners show similar or different learning patterns in the context of artificial language learning studies is of great interest given the well-known finding that children are often more successful at acquiring a second language relative to adults (Newport, 1990). Thus, that children show similar patterns to adult learners in the context of statistical learning studies may raise the question of whether learning in these artificial settings has anything to do with real-world language learning. Although whether and how performance in artificial language learning studies is related to real-world language learning abilities is an important and under-studied area of inquiry (but see Misyak & Christiansen, 2012, for a recent exception), it is important to point out that children do not outperform adult learners in all aspects of real-world language learning. Additionally, in statistical learning studies of syntactic acquisition, a domain of language learning where children often do outperform adult learners (e.g., Johnson & Newport, 1989), researchers have found that children and adults demonstrate different learning patterns (Hudson Kam & Newport, 2005; 2009). Further, even when researchers do find similarities between adult and child learners, as is the case in word learning, the extent to which similar performance is driven by the same processes is always an open question. To further probe the similarities and differences between adult and child cross-situational word learning, future should explore whether children's learning is similarly constrained by in-the-moment cues to reference (Ichinco et al., 2009; Kachergis et al., 2012a; Yoshida et al., 2012; Yurovsky & Yu, 2008) and whether children's learning exhibits similar computational signatures (Klein & Yu, 2009) and memory effects (Vlach & Sandhofer, 2010) as adult learning.

3.4.3 Why does contextual diversity aid learning?

In the current experiment, greater diversity in learning contexts aided children's word-to-referent mapping. This finding is consistent with previous studies of adult learners in a similar paradigm (Kachergis et al., 2009; Suanda & Namy, 2012), other work on learning new words from written texts (Bolger et al., 2008), and observational and corpus analyses that connect early language environments to acquisition outcomes (Hills et al., 2010; Hoff & Naigles, 2002). Further, this finding is in accord with a broader body of evidence suggesting that increasing variability of learning environments improves learning (Gomez, 2002; Hintzman & Stern, 1978; Postman & Knecht, 1983; Rost & McMurray, 2009, 2010; S. Smith, Glenberg, & Bjork, 1978; Verkoijen, Pikers, & Schmidt, 2004).

Although there is abundant evidence demonstrating *that* contextual diversity helps learning, the precise reason for *why* it helps is unclear, though a number of hypotheses have been put forward. First, some have argued that increasing variability of learning instances allows for more decontextualized representations (e.g., Apfelbaum & McMurray, 2011). Second, based on earlier memory research, some scholars have argued that contextually diverse learning environments allow for a greater number of potential cues at time of memory retrieval (Bower, 1972, Glenberg, 1979). Finally, Bjork and colleagues have offered an explanation based on the notion of "desirable difficulties" in learning. That is, contextually diverse learning opportunities initially create more difficult individual learning instances due to the mismatch between learning instances. This initial difficulty boosts the strength of learning in the long run, so long as the encoding of individual instances is successful (Bjork, 2011). Thus, a number of potential explanations exist to explain the current findings and determining which is at play in the current study is an interesting direction for future work.

Adding to the puzzle of the nature of contextual diversity effects in learning is the large number of findings across the memory, learning, and language literatures that fail to find a benefit of contextual diversity on learning (Dempster, 1987; Postman & Knecht, 1983; Young & Bellezza, 1982), as well as those that find a benefit for context *redundancy* across learning contexts (e.g., Haryu, Imai, & Okada, 2011; Maguire et al., 2008; Vlach & Sandhofer, 2011). For example in one study of early verb learning, Maguire and colleagues presented two-year-olds with an actor performing a novel action coupled with a verb-naming event, "wow, watch her *blicking*!" Two-year-olds either saw four instances of the same actor *blicking*, or four instances of four different actors *blicking*. Maguire and colleagues found that children who were presented with the same actor on all learning instances were better able to extend the label to a novel instance of *blicking* (i.e., with a novel actor). These results in the domain of verb learning (see also Haryu et al., 2011) are reminiscent of other findings from the infant and child categorization literatures suggesting that children are better able to detect higher-order relational categories when presented with instances that share similar surface characteristics (e.g., Casasola, 2005; Cohen & Oakes, 1993; Gentner & Namy, 1999; Namy, Clepper & Gentner, 2007; Namy & Gentner, 2002).

Thus, although the current findings, along with others (Kachergis et al., 2009; Suanda & Namy, 2012) clearly demonstrate a positive effect of increased contextual diversity in cross-situational word learning, the nature of these contextual diversity effects, and how they relate to other findings of contextual diversity, is a topic for future research.

3.4.4 Conclusion

In the current study, children readily discovered word meaning through a process of tracking word-to-referent co-occurrence information across situations. This work extends previous work on cross-situational word learning in both infant and adult learners, as well as previous work on children's statistical learning in other areas of language acquisition. The current work begins to shed light on the constellation of factors that influence learning, demonstrating the positive effect of contextual diversity on children's learning. Future work is needed to uncover other factors that influence learning, whether they are the same as those that influence adult learning, and whether cross-situational word learning shares similar patterns to other types of learning. Finally, although the current work highlights the availability of a cross-situational learning strategy in children during a critical period of vocabulary growth, future research should more directly assess whether these children employ their cross-situational learning in real-world word learning.

Chapter 4. Probing the Precision of Children's Mappings in Cross-Situational Word Learning

4.1 Background

How precise were the mappings children made from the few ambiguous naming events presented to them in the above experiment? Recall that the structure of the test trials in the experiment in Chapter 3 involved the presentation of a target word, a target object, and three foil objects that had *never* occurred with the target word during the learning phase. Although this structure allowed for equating the structure of the test trials across conditions and thus served as a strict test of the effects of contextual diversity on learning, it did not allow for a strict test of the specificity of mappings between words and their referents. That is, based on the test results in the experiment above, we do not know if children created a definitive mapping between words and their most frequently co-occurring objects, or whether children simply discriminated between objects that had versus those that had not co-occurred with the target words during learning.

The goal of the current experiment is to probe the precision of children's cross-situational mappings following a small number of ambiguous naming events. In this experiment, children completed a task identical to the experiment presented in Chapter 3 with the exception that at test, children were presented with 4AFC trials that contained foils that also had co-occurred with the target word, with varying degrees of frequency. Of particular interest in this experiment is the nature of children's selection patterns. If what children gained from these ambiguous naming events is a broad sensitivity to objects that had versus had not co-occurred with the target word, then I should find that children distribute their answers equally among objects that had co-occurred with the target object. Alternatively, if children's learning was sensitive to the relative frequencies with which objects had co-occurred with the target word, then children's selection patterns should reflect word-to-object co-occurrence frequencies, resulting in a preference for selecting the objects that occurred more consistently with the word over those that co-occurred with the word less frequently.

4.2 Method

4.2.1 Participants

Twenty-eight 5- to 7-year-olds (*Mean* age = 74.7 mos, *Range* = 62.5 – 96.5) from the greater Atlanta area participated. Fourteen children were female. As with Experiment 1, participants were children of families who had volunteered to participate in studies at Emory University's Child Studies Center. 78% of children were Caucasian, 13% were African-American, 3% were Asian, and 6% identified as members of other racial categories. 6% of families identified as Hispanic or Latino.

4.2.2 Stimuli, Design & Procedure

The stimuli, design, procedure, and coding were identical to the experiment reported in Chapter 3 except in the following ways. There was only one learning condition in the experiment. The contextual diversity of the learning environment was identical to that of the Moderate CD condition of Chapter 3's experiment (see Table 1, page 32). The critical difference between the current experiment and Chapter 3's experiment was the structure of the test trials. In Chapter 3's experiment, the foils presented on test trials had never co-occurred with the target word during learning. In the current experiment, the foils *had* co-occurred with the target word during learning. Specifically, each of the eight 4AFC test trials was constructed such that the target word was paired with its corresponding referent and 3 pseudo-randomly selected foils that had co-occurred at different rates with the target word during the learning phase. One foil had co-occurred with the target word on two of the four learning trials in which that word occurred, another foil had co-occurred with the target word once during the learning phase, and the other foil had never co-occurred with the target word during the learning phase. Of interest is whether children's selection patterns reflect the co-occurrence frequency between the item selected and the target word.

As in Chapter 3's experiment, two coders, blind to the experimental hypothesis, tested for experimenter bias. Enthusiasm ratings of the current experiment (M = 4.25) were comparable to the ratings reported in Chapter 3 (M = 4.25). Additionally, there was very little agreement between coders' judgments of children's performance and children's actual performance ($K_{Coder\ 1}$ <.01; $K_{Coder\ 2}$ <.01), or between the two coders' judgments (K = .08).

4.3 Results

As in the previous experiment, I computed the proportion of trials each child answered correctly. Preliminary analyses revealed a significant effect of training list, qualified by a significant interaction between training list and test list on mean proportion correct, p = .043, suggesting that performance in one training-test list combination was significantly higher than the others. An inspection of the data revealed that this effect was driven primarily by a single child who performed well above mean performance for the group. When this child was removed from the analysis, the training and testing list effect was no longer statistically significant, $p > .05$. Importantly, the statistical significance of the primary analyses presented below were not markedly altered when this participant's data were removed from analyses. As in the previous experiment, I found no correlation between age and performance, $r = -.01$, $p = .95$, and no sex differences in performance, $p = .71$.

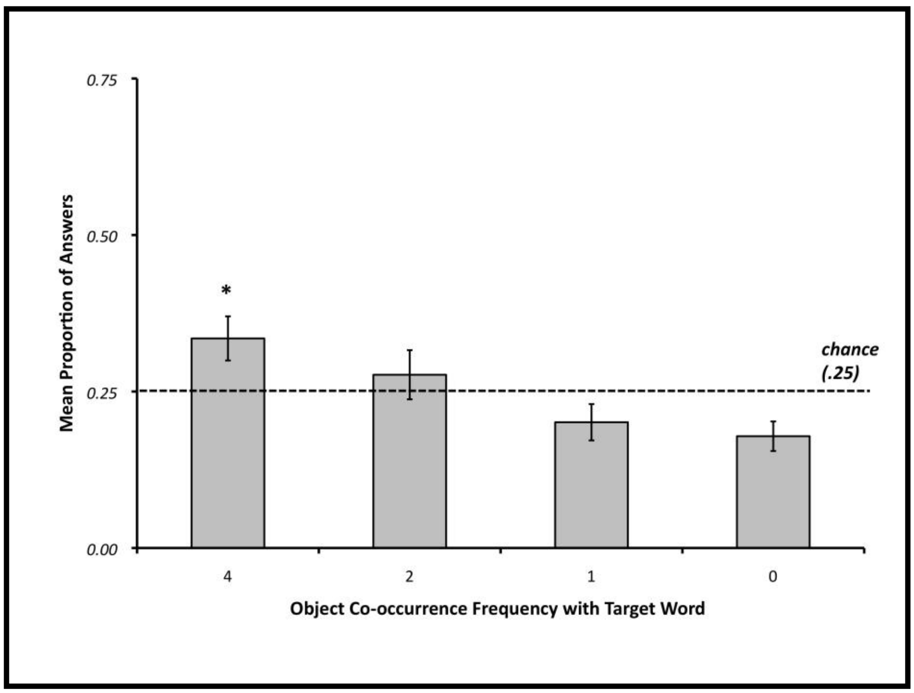

Figure 8. *Distribution of answers to objects differing in co-occurrence frequency with target word*

To examine the extent to which participants learned word-to-referent pairings in this more challenging testing regimen, I performed a single-sample *t*-test on mean proportion correct against the learning rate that would be expected by chance performance (.25). Results revealed that the mean proportion correct ($M = .335$; $SD = .19$) was significantly higher than chance levels, $t(27) = 2.41$, $p = .02$, suggesting that children successfully mapped words onto their referents. This group-level analysis was supported by individual-level analysis of performance. A chi-square goodness-of-fit test revealed that the proportion of children performing above chance (*Prop.* = .50) was significantly higher than what would be predicted by random performance, $\chi^2 = 4.09$, $p = .04$.

To investigate the precision of children's mappings, the primary question of interest in the current experiment, I examined the relation between the likelihood of selecting a particular picture and the word-to-picture co-occurrence frequencies. As Figure 8 illustrates, there is a relation between the word-to-picture co-occurrence frequency and picture selection probability: the more frequently a word and picture co-occurred during learning, the more likely that picture was selected at test. To statistically investigate this relation, I conducted a series of pair-wise comparisons, given that the responses across the four response types were non-independent. I compared the relative likelihood of selecting one item (e.g., the item that co-occurred 4 times with the target word) over a different item (e.g., the item that co-occurred 0 times with the target word). To do this, for each child, I tallied the number of trials the child selected either item (e.g., the 4-item or the 0-item). I then examined the proportion of these trials (i.e., trials in which the child selected either the 4-item or the 0-item) that the child selected the item with greater frequency (in this case, the 4-item). A mean proportion across all children was derived and compared to the proportion that would be expected if the child selected the greater frequency item just as much as the lower frequency item (i.e., .50). Comparisons were made for all possible answer pairs (see Table 2). As the table illustrates, children were more likely to select the 4-item more than the 0- and 1-item. That children discriminated between the 4-item and 1-item, suggests that children's mappings were more than just an acknowledgement that a word-picture pairing had co-occurred, children's mappings reflected sensitivity to the *relative* frequency with which a word-picture had co-occurred. However, the lack of statistical significance in all other pair-wise comparisons highlights the imprecision of children's mappings. For example, there is no evidence to suggest that children discriminated between the two items that co-occurred most frequently with the target word (i.e., the 4- and 2-item).

Table 2. *Pair-wise comparisons of the likelihood selecting a particular item type (word-object co-occurrence frequency type).*

Pair-wise Comparison (word-object co-occurrence frequency)	*Mean Proportion Selection of More Frequent Item (SD)*	*One Sample t-test*		
		t-value	*p-value*	*Cohen's d*
4 vs. 0	.663 (.24)	3.66	.001	.69
4 vs. 1	.633 (.23)	3.08	.005	.58
4 vs. 2	.564 (.28)	1.19	.24	.23
2 vs. 0	.553 (.32)	.86	.39	.17
2 vs. 1	.535 (.32)	.56	.58	.11
1 vs. 0	.521 (.33)	.34	.73	.06

4.4 Discussion

In the current experiment, I investigated children's ability to distinguish between candidate referents differing in their frequency of co-occurrence with the target word. Results from the study suggested that children were sensitive not only to whether a word and picture had co-occurred but, at least in part, to the relative frequencies with which pictures and target word had co-occurred. At the same time, there was imprecision in children's mappings. That is, although children picked out the picture with a high co-occurrence frequency relatively more often than the object with a low co-occurrence frequency, they did not select the picture with a high co-occurrence frequency more than the picture with a moderate co-occurrence frequency.

There are at least two learning processes that can readily explain the pattern of performance children exhibited. First, children may have kept track of multiple candidate referents for each word via an associative learning process (see Yu, 2008). Consequently, the observed selection pattern is a result of the relative strength of internalized associative links between words and objects that reflect the relative co-occurrence frequencies of the learning environment. Alternatively, children may have adopted a hypothesis testing strategy (see Nicol Medina et al., 2011) by which they generated and tracked a single hypothesized mapping for each target word. Under this account, the selection pattern is a result of a simple likelihood analysis: the likelihood of landing on a particular hypothesized referent object is proportional to the frequency with which that object had co-occurred with the target word. There has been recent debate over which of these two learning mechanisms best explains cross-situational word learning (Nicol Medina et al., 2011; Scott & Fisher, 2012; Yu & Smith, 2012). Although the current findings cannot distinguish between the two, through a series of computational simulations conducted in the following chapter, I hope to shed some light onto this issue.

Regardless of whether the current findings are the result of an associative learning or hypothesis testing mechanism, the finding that children's learning fell short of definitive mappings between words and their correct referents highlights an important fact about real-world word learning: learning word meaning is not an all-or-none process. The notion that children go through intermediate levels of learning and may possess only partial knowledge of a word's meaning might seem like a trivial notion. However, partial learning is a topic that has received very little attention in the early word learning literature, though it is a topic that is studied in both reading research (Schwanenflugel et al., 1997; Wagovich & Newhoff, 2004) and adult learning literature (Durso & Shore, 1991; Frishkoff, Perfetti, Collins-Thompson, 2011; Shore & Durso, 1990; Whitmore, Shore, & Smith, 2004). This lack of attention to partial learning is likely the result of a number of factors, including (1) a theoretical focus of word learning theories on fast mapping and the nature of children's *initial* word mappings (see Carey, 2010; Swingley, 2010 for discussion), (2) methodological limitations in the measurement instruments of children's vocabulary that tend to be binary (e.g., in parent report measures, see Fenson et al., 1994, parents commonly simply indicate whether children do or do not know a word), and (3) methodological limitations in how partial word knowledge is operationalized (i.e., partial word knowledge is often operationalized in part using metalinguistic judgments at which young children are notoriously weak, see

Marazita & Merriman, 2004). The growing interest in cross-situational word learning and the development of cross-situational learning paradigms, such as the one presented here, provides the opportunity to study the progression of learning through different degrees of word knowledge, as well as study the role partial word knowledge may play on subsequent word learning (see Yu, 2008; Yurovsky et al., 2010).

Chapter 5. Children's Cross-Situational Word Learning: Hypothesis Testing or Associative Learning?

5.1 Background

In Chapters 3 and 4, I reported two behavioral experiments on children's cross-situational word learning. The experiment in Chapter 3 served as both a first test of school-aged children's cross-situational learning capacities and an initial investigation into the factors that influence learning. The findings of that experiment revealed that children could readily learn new words by tracking word-to-referent co-occurrence statistics across ambiguous naming events. Furthermore, the results pointed to one factor that appears to influence cross-situational word learning, namely the diversity of contexts in which words and their candidate referents co-occur. The experiment in Chapter 4 probed further the nature of cross-situational word learning by testing the precision of children's word-to-referent mappings. The findings of that experiment revealed that although children displayed some sensitivity to the relative frequencies of word-to-referent co-occurrences, they also exhibited some imprecision in their mappings, namely they were unable to discriminate between highly frequent and moderately frequent co-occurring referents.

Documentations of learning patterns and investigations into the factors that influence learning are necessary steps to understanding cross-situational word learning. However, a deeper understanding of the phenomenon involves investigating the underlying mechanisms that underlie behavior. The experimental manipulation in Chapter 3 was a first step in this direction. That is, that manipulations of contextual diversity impacted children's learning ruled out one potential mechanism, namely a one-trial learning strategy. Thus, children must have recruited the cross-situational consistencies in word-to-referent co-occurrences in the service of word learning. However, multiple mechanisms, as discussed in Chapter 4, are consistent with learning via sensitivity to cross-situational co-occurrence statistics. For example, it is possible that when children encounter a new word in an ambiguous naming event, they pick out a specific object as a hypothesized referent of the word. As children encounter this word in a subsequent naming event, children either confirm the hypothesis if it is consistent with the event (i.e., the initially hypothesized referent is also present in the subsequent event) or reject and replace the hypothesis if it is inconsistent with the event (i.e., the initially hypothesized referent is not present in the subsequent event). This process of confirming or replacing hypothesized word-to-referent mappings should yield successful cross-situational word learning based on the logic of probability: more frequently co-occurring word-to-object pairs will be more likely to be selected as word-to-referent hypotheses than infrequently co-occurring word-to-object pairs.

Alternatively, it is also possible that when children encounter a new word in an ambiguous naming event, they encode the connection between this word and multiple, if not all, co-occurring objects. As children encounter this word in subsequent naming events, they strengthen previous connections as well as create new connections. This process of aggregating cross-situational co-occurrence statistics also yields successful learning because the connections between frequently co-occurring word-to-object pairs will be stronger than infrequently co-occurring word-to-object pairs.

Thus, at least two learning mechanisms are consistent with cross-situational word learning. These two learning processes, better known as *hypothesis testing* and *associative learning* (see Yu & Smith, 2007, 2012; Xu & Tennenbaum, 2007a) have long been central to many discussions of children's word learning (e.g., Markman, 1992; Smith, 2000). They have received increased attention in recent years in the context of cross-situational word learning as there has been much debate over which of the two mechanisms best accounts for the empirical data (e.g., Nicol Medina et al., 2011; Scott & Fisher, 2012; K. Smith et al., 2010; Yu & Smith, 2011). Thus, the goal of this chapter is to investigate the mechanisms underlying children's cross-situational word learning and to shed light onto the hypothesis testing vs. associative learning debate. I do so by comparing how readily computer simulations of the two learning algorithms model and predict the children's data collected in Chapters 3 and 4. In what follows, I briefly discuss the two accounts and summarize important differences between them. I then describe the current computational instantiations of these accounts and present a series of simulations that test models of these processes against the children's data, testing the extent to which, and the conditions under which, they can account for the behavioral findings. I then end by discussing the implications of these findings for the learning mechanisms underlying cross-situational word learning.

5.1.1 Hypothesis Testing

Under hypothesis testing accounts (e.g., Frank et al., 2009; Nicol Medina et al., 2011; Siskind, 1996; Xu & Tennenbaum, 2007a), learning involves a process of picking out specific hypothesized word-to-referent mappings that are consistent with a given naming event and then testing these hypotheses against future naming events. Imagine a simple example of a child hearing the word "ball" in the context of a ball and a bat. If the child has no experience with this word and these objects, according to a basic hypothesis testing account, the child would randomly select one of the objects (e.g., the bat) as the referent of "ball". Further, imagine the same child encountering "ball" again, but now in the context of a ball and a basket. According to this account, that the hypothesized referent (i.e., bat) is not present in the environment would lead the child to discard this hypothesis and select a new hypothesized referent (e.g., ball). This process continues until the child lands on a word-to-referent hypothesis that has received sufficient cross-situational confirmation.

Computational models that reflect variations of this general process have been proposed as candidate mechanisms of many empirical findings in children's word learning, including cross-situational word learning. In one of the earliest computational investigations of cross-situational learning, Siskind (1996) examined the extent to which a computer algorithm could learn word-meaning mappings from a large corpus of artificially generated sentences (e.g., "John walked to school") each paired with the sentence's corresponding meaning (e.g., [John], [walked (John, to school)], [school])[4]. Learning in the model involved reducing the hypothesis space for candidate word meanings by testing hypotheses generated via a series of inference rules (e.g., that each word must map onto only a single meaning). Not only could Siskind's model eventually converge on the right word-meaning mappings, the model demonstrated a number of findings that parallel children's vocabulary development. For example, the model showed a pattern of slow word learning in the early stages (i.e., the model required a large number of instances to learn a single word), followed by a pattern of fast learning later in the process, consistent with the typical trajectory of vocabulary development in children (see Fenson et al., 1994). More recent computational models that employ a hypothesis testing algorithm have accounted for other aspects of children's learning such as children's bias to generalize novel nouns based on shape (Kemp, Perfors, & Tenenbaum, 2007), the effect of labeling contexts in novel noun generalization (Xu & Tennenbaum, 2007a), the effect of socio-pragmatic cues in word learning (Frank et al., 2009; Xu & Tennenbaum, 2007b), and children's mutual exclusivity bias (i.e., children's preference to map a novel label onto a novel object over a familiar one; Frank et al., 2009; Regier, 2003).

5.1.2 Associative Learning

In contrast to hypothesis testing, associative learning accounts of word learning typically involve a learning process that builds a network of associations

---

[4] In this example, predicates are represented in a [P(x,y)] format where *P* is the predicate, and *x,y* represents the predicate's arguments. In Siskind's actual model, there was a broader range of meanings from the sentence "John walked to School": [**John**], [Go(John, school)], [Go(John,y)], [Go(,school)], [To(school)], [To(x)], [**school**]

between many words and many referents. Under these accounts, definitive word-to-referent mappings do not necessarily exist. Instead, there may simply be strongly associated word-referent pairings and weakly associated ones. Returning to the example mentioned above, if a child encounters the word "ball" in the context of a ball and a bat, according to this account, the child would link the word "ball" with both the ball and the bat. The child would also link the word "ball" with the objects present in the second naming event, thus increasing the strength of the link between the word "ball" and the ball and adding an associative link between the word "ball" and the basket. According to this account, as children develop their vocabulary, they acquire a large associative network between many words and many referents; with certain associative links being stronger and others weaker. Word mappings in any given moment are resolved probabilistically based on the relative strengths of the associations with all candidate referents.

A number of recent computational models of cross-situational word learning employ associative principles. For example, Yu (2008) conducted one computational study to examine the extent to which an associative learning model could learn word-to-referent mappings from input patterns derived from parent-child semi-naturalistic interactions. Parents narrated text-free picture books to their children. The words parents produced while viewing each page coupled with the pictures appearing on that page were then fed to the associative model as input. The model learned words by tracking the associative probabilities between words and objects in the pictures[5]. After presenting the model with multiple exposures to each page of the book (multiple parent-child dyads contributed to the input), the model was tested for whether it correctly associated words with their referents. Yu found that the simple learning model could reliably discover correct word-object mappings from a set of inputs that approximates the types of input to which child word learners are exposed. Over the years, associative models have accounted for a number of other empirical findings in children's word learning, including the shape (Colunga & Smith, 2005; Samuelson, 2002) and taxonomic biases (Mayor & Plunkett, 2010) in children's novel noun generalizations, the vocabulary spurt (Plunkett et al., 1992; McMurray, 2007) and the increasing sensitivity to linguistic forms in word learning that children exhibit with development (Apfelbaum & McMurray, 2011; Regier, 2005).

5.1.3 Current Endeavor

Clearly, both hypothesis testing and associative learning accounts have established records in accounting for a wide range of word learning phenomena. The goal of the current endeavor is to test these two accounts against the behavioral findings reported in Chapters 3 and 4. Given the success of both accounts, it is plausible, and in fact likely, that both models could account for the findings. Thus, the purpose of this endeavor is not necessarily to address whether some hypothesis testing or associative learning model *can* account for children's learning patterns but rather to understand the conditions under which each account does so successfully. To do this, I purposefully start this endeavor by creating and testing

[5] Yu (2008)'s model also included additional processes such as a function-word filter as well as a mutual exclusivity component.

versions of each model that are maximally different from one another along three properties of learning: (1) the learning algorithm of the models, (2) the nature of information intake, and (3) the type of information the models retain. Table 3 illustrates how the two models differ along these properties. Although many recent instantiations of hypothesis testing and associative learning models stray from these standards, and thus blur the line between the two models, these differences are generally considered to be the core canonical properties on which the two accounts differ (for discussion, see Nicol Medina et al., 2011; Yu & Smith, 2012). The complete architecture of each model is presented in the following section.

Table 3. *Core differences between hypothesis testing and associative learning models on three computational properties.*

	Model	
Model Property	*Hypothesis Testing*	*Associative Learning*
Cross-situational learning algorithm	Updating of word-to-referent hypotheses via a process of hypothesis confirmation, rejection, and/or replacement based on a defined confidence threshold	Accrual of associative links between co-occurring words and objects, the strength of which can vary continuously
Information selection within a naming event	Mutually exclusive word-to-referent mappings	Associations between multiple words and multiple objects
Information retention from previous naming events	Alternative hypotheses, information from past learning situations are not retained	Previously associated word-object pairs are represented in the associative network

5.2 Models

5.2.1 Hypothesis Testing (HT) Model

A Hypothesis Testing (HT) model was developed to instantiate the learning components described above in Table 3. In the HT model, the learners' lexicon is represented as a hypothesis list of *mutually exclusive* word-to-referent hypotheses. That is, each word is mapped to one and only one candidate referent. Learning in the HT model occurs through a process of updating the hypothesis list based on the matches and mismatches between the existing hypothesis and incoming evidence from naming events. Below is a step-by-step description of the current model's operation, followed by a toy example of the model in action.

At the onset of learning, the learner's hypothesis list is blank. On each naming event, the model randomly selects mutually exclusive word-to-referent hypotheses from the candidate mappings that are present in the naming event. For example, if a naming event consisted of two words {W1, W2} and two objects {O1, O2}, the model might select hypotheses W1-O2 and W2-O1 from the naming event's set of candidates {W1-O1, W1-O2, W2-O1, W2-O2}. Early in the learning phase, when there are no existing hypotheses to guide learning, selected hypotheses are simply added to the hypothesis list. As learning progresses, the learner updates the hypothesis list by *confirming*, *rejecting*, or *replacing* hypotheses. That is, if a selected hypothesized word-referent pair from a naming event is consistent with a learners' existing hypothesis, then that hypothesis is *confirmed*, and the learner adds confidence in the hypothesis. Alternatively, if the hypothesis selected is inconsistent with any existing hypotheses (i.e., the learner possesses a different hypothesis for either the word, the object, or both), then what happens depends on the learners' confidence in the existing hypothesis relative to a set threshold value. The threshold value is a model parameter that is free to vary and determines whether hypotheses are rejected or maintained. If the strength of the existing hypothesis is above the threshold, then the hypothesis is retained, even in the face of conflicting evidence, and the selected hypothesis generated by that naming event is *rejected.* Alternatively, if the strength of any existing hypothesis is below the threshold, then the existing hypothesis is *replaced* in favor of the newly selected hypothesis generated on that trial.

When tested for word knowledge (i.e., presented with a 4AFC test trial), there are two potential outcomes for the simulated hypothesis tester. If the learners' hypothesized referent for the target word is among the objects available to select, then the learner will select that object. Alternatively, if the learners' hypothesized referent for the target word is not among the objects available to select, then the learner selects randomly among available objects.

To illustrate how the HT model works, consider the toy example in Figure 9A. In this example we will assume that the simulated learners' confidence threshold ω (i.e., the threshold that determines whether selected hypotheses are rejected or existing hypotheses replaced) is set at 2. That is, the simulated learner will be sufficiently confident in a hypothesis if it has been selected from two naming events. When the simulated learner is presented with the words *a* and *b*, and objects A and B on the first trial, the learner randomly selects two mutually exclusive hypotheses {*a*-A, *b*-B}, and these hypotheses are added to the hypothesis list. On the second learning trial, two additional randomly selected hypotheses {*c*-D, *d*-C} are added to the hypothesis list. On the final trial, the model selects the hypotheses *a*-A and *c*-C.

Because one hypothesis {*a*-A} is consistent with the existing hypothesis list, this hypothesis is confirmed, and confidence in this hypothesis is increased. Since the second selected hypothesis in this example {*c*-C} is inconsistent with existing hypotheses in the list {*c*-D, *d*-C}, and these existing hypotheses are below the hypothesis threshold, the existing hypotheses are replaced in favor of the newly selected hypothesis. On the test trial illustrated in the figure, this simulated learner would select object A because it is the existing hypothesized referent of the word tested (*a*).

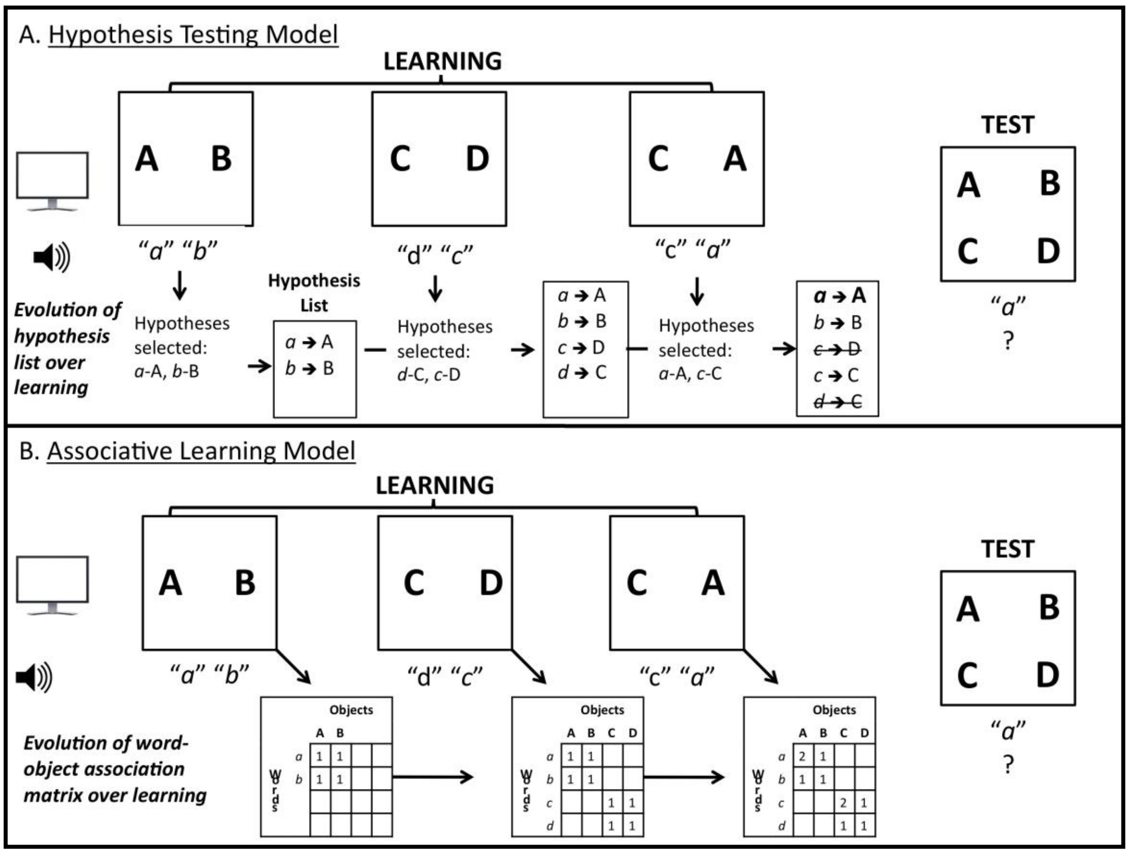

Figure 9. *Toy example illustrating the learning algorithms implemented in the Hypothesis Testing model (A) and Associative Learning model (B).*

This toy example illustrates how a HT learner could succeed at learning amidst referential ambiguity. Because words and their referents always co-occur, simulated learners are more likely to have the opportunity to select, and build confidence in, correct hypotheses over incorrect hypotheses. The toy example also underscores the importance of the confidence threshold parameter in hypothesis testing. For example, had the confidence threshold in this example been set lower (ω = 1), then on the third learning trial, the learner would not have replaced the incorrect existing hypotheses *c*-D and *d*-C in favor of the correct hypothesis *c*-C. In the reported simulations below, the confidence threshold ω is a model parameter that is varied across simulations to examine its effects on cross-situational word learning.

5.2.2 Associative Learning (AL) Model

The current Associative Learning (AL) model was constructed to instantiate the learning principles laid out in the right-hand column of Table 3. The lexicon of the simulated associative learner is represented as a large word-object association matrix, with cells representing the associative strength between specific word-object pairings. Over learning, cells are updated using a simple learning algorithm: the associative strength between a word-object pairing is increased whenever that word and object co-occur. Importantly, this updating is applied to all word-object pairs present on a given trial. There is no lateral inhibition built in to this associative learning model. That is, the association strength between a word and objects that are *not* present on a given trial is not diminished by virtue of those objects' absence. Instead, the association strength established on previous trials is maintained. Below is a step-by-step description of the model's operation, followed by a toy example of the learning algorithm employed.

At the onset of learning, the association matrix consists of empty cells. On each learning trial, the model links all presented words with all presented objects with equal weighting or association strength. For example, if words W1 and W2 were presented with objects O1 and O2, then the four possible associations {W1-O1, W1-O2, W2-O1, W2-O2} would be added to the learner's association matrix. As learning progresses, the simulated learner updates its cells using the following simple function:

$$M^{t}(w,o) = M^{t-1}(w,o) + \alpha \qquad \text{(Eq. 1)}$$

In this equation, $M^t(w,o)$ represents a particular word-object association on the t-th trial, and α represents a parameter controlling the gain in associative strength.

At test, the simulated learner computes a probability value for each object present, which is equal to the association strength between the target word and that object normalized by the sum of the association strengths between the target word and all objects present in the test trial. Thus, if the learner is presented with target word W1, and objects O1, O2, O3, and O4, the learner computes the probability of selecting O1, p(O1), which is equal to:

$$p(o_1) = \frac{M(w_1, o_1)}{M(w_1,o_1) + M(w_1,o_2) + M(w_1,o_3) + M(w_1,o_4)} \qquad \text{(Eq. 2)}$$

The learner then selects one object with a likelihood equal to the computed probability value for that object.

To illustrate how the AL model works, consider the toy example presented in Figure 9B. In this example, the learning parameter α is set to 1 (multiple values of α will be tested in the actual simulations). On Trial 1, the simulated learner associates all presented words with all presented objects {*a*-A, *a*-B, *b*-A, *b*-B}. These four associations are added to the learner's lexicon and the same association strength value is added to the relevant association cells. On Trial 2, four additional association cells {*c*-C, *c*-D, *d*-C, *d*-D} are added and receive activation. Finally, on Trial 3, the four cells {*a*-A, *a*-C, *c*-A, and *c*-C} receive activation. On the test trial, when the learner is presented with the target word *a* and the objects A, B, C, and D, the learner has a .5 probability of selecting the "correct" referent A which is higher than the probability of selecting any other individual object:

$$p(A) = \frac{M(a,A)}{M(a,A) + M(a,B) + M(a,C) + M(a,D)} = \frac{2}{2+1+1+0} = \frac{2}{4} = .5$$

This learner has only a .25 probability of selecting objects B and likewise object C. This learner will never select object D. The model implements a choice in the current simulations via random selection from a weightedpool. This can be conceptualized as analogous to picking a lottery ball from a box with half of the balls marked A, a quarter of them marked B and a quarter marked C. Although on any given trial the simulated learner might select an object with a lower probability, over repeated testing the average of selecting each object should converge on that object's probability value.

This toy example illustrates how an associative learner, who links a single word to multiple objects and a single object to multiple words, nonetheless succeeds at a test of cross-situational word learning: when tested for the referent of a particular target word, the probability with which the simulated learner selects the most frequently co-occurring object is always greater than the probability with which the simulated learner selects less frequently co-occurring objects.

In the following simulations, I examine the extent to which hypothesis testing and associative learning mechanisms can model and predict children's learning patterns reported in Chapters 3 and 4. I will first present data addressing the extent to which HT and AL models are impacted by the contextual diversity of the learning environment in a way similar to children as reported in Chapter 3. I then test whether the HT and AL models predict similar item-selection patterns to children's reported in Chapter 4. I then end by considering the implications of the results for potential mechanisms underlying children's cross-situational word learning.

5.3 Simulation 1: The role of contextual diversity in cross-situational word learning

5.3.1 Method

As with children in the experiment in Chapter 3, simulated learners completed one of three learning phases, differing in the contextual diversity of the learning environment. Following the learning phase, simulated learners completed a series of 4AFC trials that included the target word, the target word's referent and three foils that had not occurred with the target word during learning. To compare model performance with children's performance, the data from simulated learners were processed the same way that children's data were. That is, a proportion correct score was derived from each simulated learner. Then, a mean proportion correct across simulated learners was computed for each condition of contextual diversity. Of interest is the extent to which simulated hypothesis testers and associative learners produce the same *contextual diversity effect* observed in children. That is, does accuracy at test increase as contextual diversity during learning increases?

Each models had one free parameter, word knowledge threshold in the HT model and learning rate in the AL model. Four variants of each model (varying only in the value of the free parameter) were tested. HT models varied in the word knowledge threshold (ω = 1, 2, 3, and 4). In HT models with a high word knowledge threshold, simulated learners would need to observe evidence consistent a word-to-referent hypothesis more times before they commit to that hypothesis. AL models varied in the learning rate (α = .25, .5, .75, and 1). In AL model variants with high learning rates, there is larger gain in associative strength added each time a word-object pair co-occurs. As a consequence, models with high learning rates will have a larger absolute difference in the associative strength between frequently co-occurring word-to-object pairs and infrequently co-occurring word-to-object pairs.

For each model variant, I conducted 30,000 runs (each run representing 1 simulated learner). Because there is variation in each simulated learner, the large number of runs ensures that averages across runs converge on the learning algorithms' true predictions. The large number of runs is particularly needed in the HT model where the simulated learner randomly selects the hypotheses, which leads to great variation between learners depending on which hypotheses are selected on each learning trial. Multiple runs are also needed in AL simulations because the testing algorithm, as described earlier, makes a semi-random choice on each test-trial.

To assess the fit between simulated learners and children, I correlated simulated learners' average performance in the three conditions of contextual diversity to children's average performance in those same conditions. I used a simple Pearson's correlation coefficient with a sample size of 3 (one data point per condition) as the metric of comparison. This metric was selected because my primary interest was not necessarily whether the models would show the same learning *rate* as children but rather whether models would show the same *pattern* of learning (i.e., increased learning as levels of contextual diversity increased). By evaluating models using this method (as opposed to using a metric such as *mean squared error*), fit in this case speaks only to the relative difference between conditions, rather than on absolute fit in accuracy. One limitation to employing a correlation coefficient in the current context is the small sample size, which will lead to the possibility that large correlation values could be observed by chance alone as well as the increased possibility of a Type II error, a failure to reject the null hypothesis.

5.3.2 Results

5.3.2.1 HT model results

HT models' performance is depicted in Figure 10. As the figure illustrates, there is variability in the extent to which HT models are affected by contextual diversity. Specifically, only the model with a moderately low threshold (ω = 2) shows a clear downward trend in performance, exhibiting lower accuracy under conditions of lower contextual diversity.

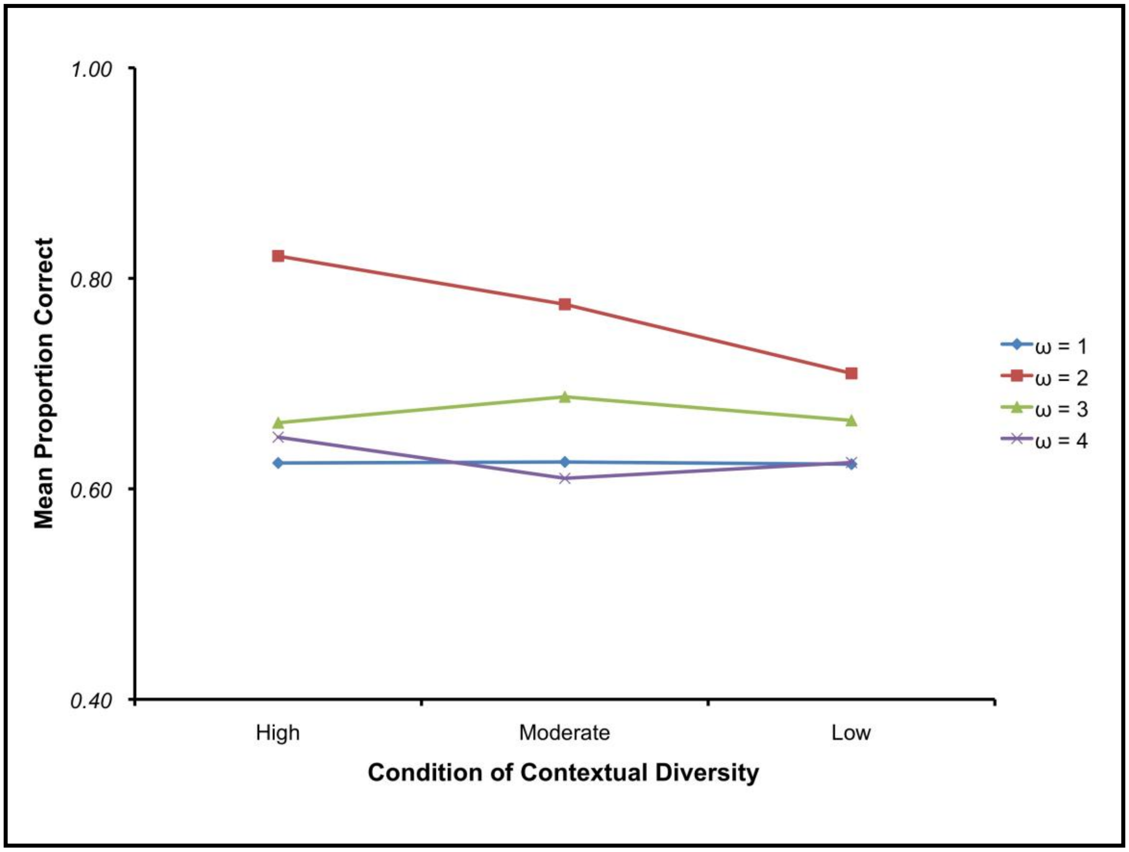

Figure 10. *HT models performance in Simulation 1: Mean proportion correct across models varying in word knowledge threshold (ω) and across conditions of contextual diversity.*

5.3.2.2 AL model results

As Figure 11 illustrates, and in contrast to the HT models' results, there is no evidence that any of the AL models' performance is impacted neither by the contextual diversity of the learning environment nor by associative learning rates. All AL model variants exhibit a ceiling effect across all conditions of contextual diversity in this simulation.

5.3.2.3 Correlations between models' and children's performance

Results of correlation analyses between children's performance variants of each model are reported in Table 4. These analyses confirm that only the HT model with a moderately low threshold (ω = 2) shows a marginally significant correlation with the contextual diversity effect observed in children. In contrast to the qualified success of the HT model in accounting for children's learning patterns, none of the AL model variants' performance patterns correlate with children's contextual diversity effect.

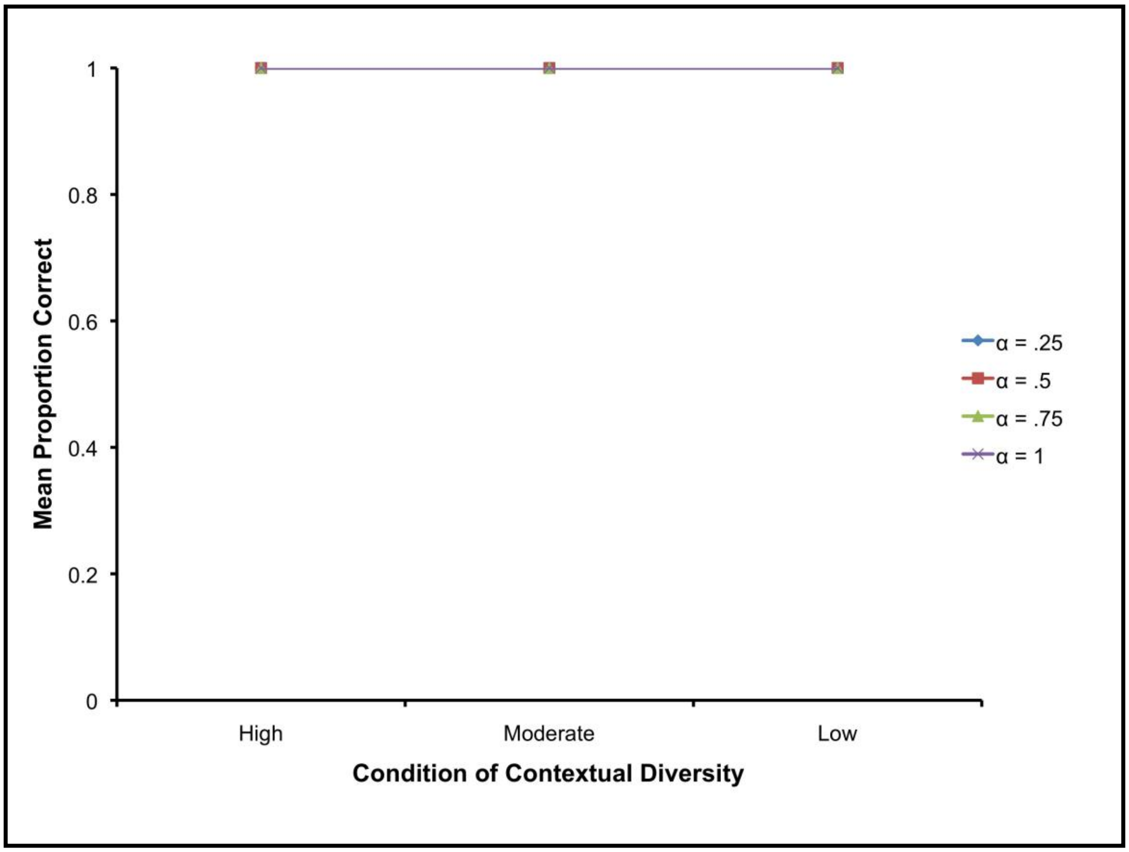

Figure 11. *AL models performance in Simulation 1: Mean proportion correct across models varying in learning rates (α) and across conditions of contextual diversity.*

Although these correlation analyses revealed little evidence that either HT or AL models fit the contextual diversity effects observed in children, the power of these analyses was quite low given the small number of observations per analysis.

One potential issue with interpreting correlations with a low number of observations is that there is an increased chance of a Type II error.

Table 4. *Correlation results between models' and children's performance in Chapter 3. Note: all tests of significance in this chapter were one-tailed, ^p < .10*

Models Tested	*Correlation Statistics*	
	r	*p*
HT Models varying in hypothesis threshold (ω)		
HT Model 1 ($\omega = 1$)	.361	.38
HT Model 2 ($\omega = 2$)	.954	.09^
HT Model 3 ($\omega = 3$)	-.278	.41
HT Model 4 ($\omega = 4$)	.758	.22
AL Models varying in learning rate (α)		
AL Model 1 ($\alpha = .25$)	-[6]	-
AL Model 2 ($\alpha = .5$)	-	-
AL Model 3 ($\alpha = .75$)	-	-
AL Model 4 ($\alpha = 1$)	-	-

To address this concern, I conducted a separate type of analysis on the models. Rather than running a large number of simulations and averaging across simulated learners, I instead conducted a series of simulated "experiments", each with a sample size equivalent to that collected with children. In these simulated experiments, 84 simulated learners were randomly assigned to one of the three conditions of contextual diversity. In each experiment, mean proportion correct across conditions was correlated with children's condition means in Chapter 3. One

[6] Pearson values of all AL models could not be calculated because there was no variability in the simulated learners' averages across conditions of contextual diversity.

thousand simulated experiments were conducted, and each experiment was classified as either correlating strongly ($r > .50$), moderately ($.30 > r > .50$), or weakly ($r < .30$) with children's performance[7]. Of interest in these simulated experiments is the proportion of experiments that correlate strongly with children's performance. Thus, this analysis addresses the question: given a particular population of simulated learners, what is the likelihood of sampling a group of learners that mimic children's performance?

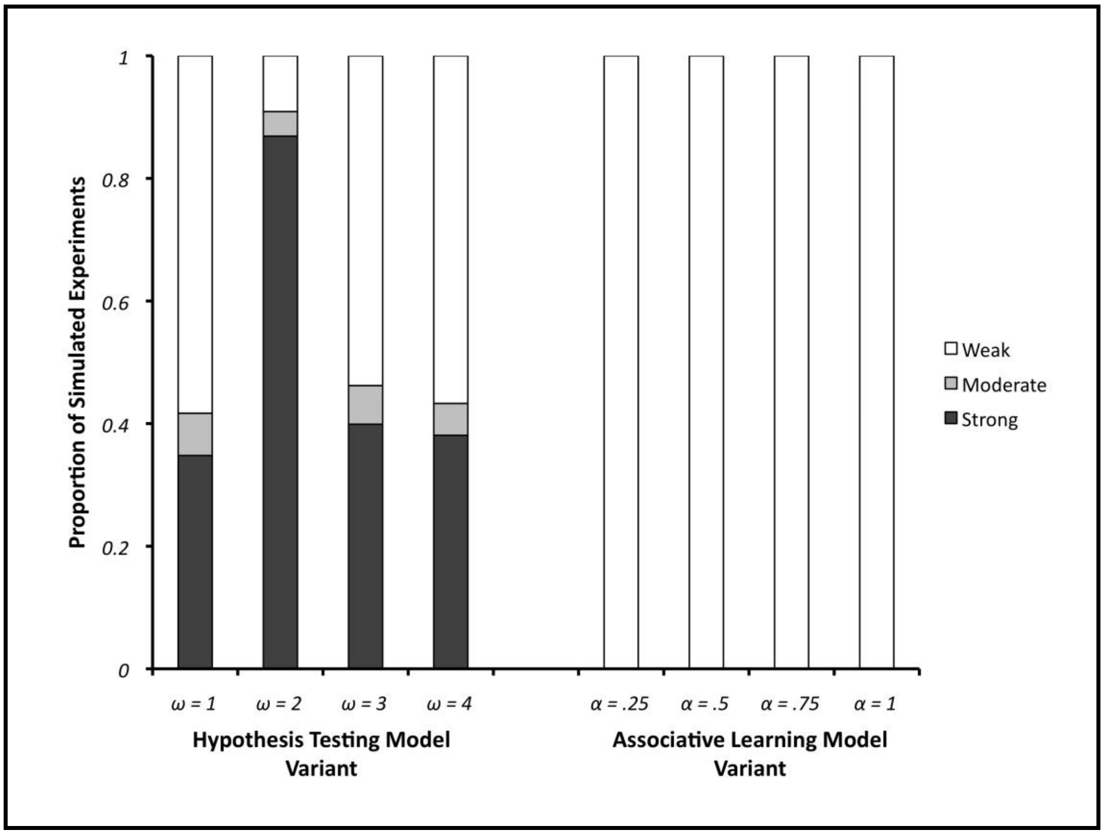

Figure 12. *Proportion of simulated experiments of HT and AL models whose results correlated strongly, moderately, and weakly with children's contextual diversity effects.*

The results of these simulated experiments are displayed in Figure 12. As the figure demonstrates, the results are consistent with, and clarify, the first analysis.

[7] Cut-off values were chosen based on thresholds established by Cohen and colleagues (Cohen, 1992a, 1992b; Hemphill, 2003)

Specifically, there is a much higher likelihood of sampling learners whose results strongly correlate with children's performance in the second HT model (ω = 2) than in the other HT models tested. None of the AL models correlated strongly, or even moderately, with children's performance.

5.3.3 Discussion

The results from these simulations suggest that a simple hypothesis testing model can account for children's contextual diversity effects. However, only one version of this model consistently predicts the contextual diversity effect. Specifically, only when the learner's hypothesis threshold, ω, is set neither too low nor too high will the model display the patterns observed in children. When ω is set too low (i.e., ω = 1), simulated learners commit too quickly to randomly selected word-to-referent hypotheses, which are just as likely to be incorrect as correct hypotheses. In contrast, when ω is set too high (e.g., ω = 4), simulated learners commit the opposite error. These learners rarely achieve confidence in their hypotheses and thus are too quick to replace existing hypotheses with newer ones, which are again equally likely to be correct and incorrect. Important in the context of the current study is the fact that the consequence of both too liberal and too conservative of a hypothesis threshold is a lack of contextual diversity effects on performance. However, when the threshold is set neither too low nor too high, simulated learners in the High CD condition only ever achieve confidence in correct hypotheses because the frequency of co-occurrence between a word and any given distractor is low enough that the simulated learner will never wrongly commit to an incorrect hypothesis. In contrast, simulated learners in the lower contextual diversity conditions will sometimes commit to incorrect hypotheses because words and distractors in these conditions occur multiple times. These mismappings are what account for the model's observed contextual diversity effect.

In contrast to the success of at least one variant hypothesis testing model in explaining the contextual diversity effect, all variants of the associative learning model failed to account for the effect. The most obvious answer to why associative learners did not show the contextual diversity effect is that on test trials, associative learners considered only the relative associative strength between words and the objects present on the test trial. Because the test trials contained only foils that had never co-occurred with the target word, associative learners always accurately selected the target object, resulting in a ceiling effect. Later in this chapter I return to this issue, considering whether there are modifications to the associative model that would yield the contextual diversity effect. First, however, I examine whether this basic associative model, as well as the hypothesis testing model can account for a second pattern exhibited in children's performance, namely, the item selection effects observed in children (Chapter 4).

5.4 Simulation 2: Item selection as a function of cross-situational co-occurrence statistics

5.4.1 Method

In this simulation, learners completed the same learning phase as children did in the experiment reported in Chapter 4 (learning with moderate contextual diversity), and then completed a series of 4AFC trials that included the target word, that word's referent (or the "4-item", the item that co-occurred four times with the target word during learning) and three foils, two of which had co-occurred with the target word during learning (2-item, which co-occurred twice with the target word; 1-item, which co-occurred once with the target word; 0-item, which never co-occurred with the target word). The proportion of trials on which simulated learners selected each answer type (i.e., 4-item, 2-item, 1-item, and 0-item) was computed. Then, a mean proportion for answer type was derived across simulated learners. Of interest is the extent to which simulated learners show the same *item selection effect* observed in children: the finding that the higher the item's co-occurrence frequency with the target word, the higher the probability of selecting that item in the 4AFC trial.

As in Simulation 1, four variants of HT models (ω = 1, 2, 3, 4) and four variants of AL models (α = .25, .5, .75, 1) were conducted at 30,000 simulation runs per variant. Pearson's correlation coefficients were again used for comparing model performance to children's performance.

5.4.2 Results

5.4.2.1 HT models results

As Figure 13 illustrates, the pattern of findings across hypothesis testing variants in this simulation is reminiscent of Simulation 1. That is, although all model variants show an overall decline in likelihood of selecting objects with lower co-occurrence frequencies with the target word, only the HT model with a moderately low threshold (ω = 2) shows the graded downward trend that is most similar to that of children.

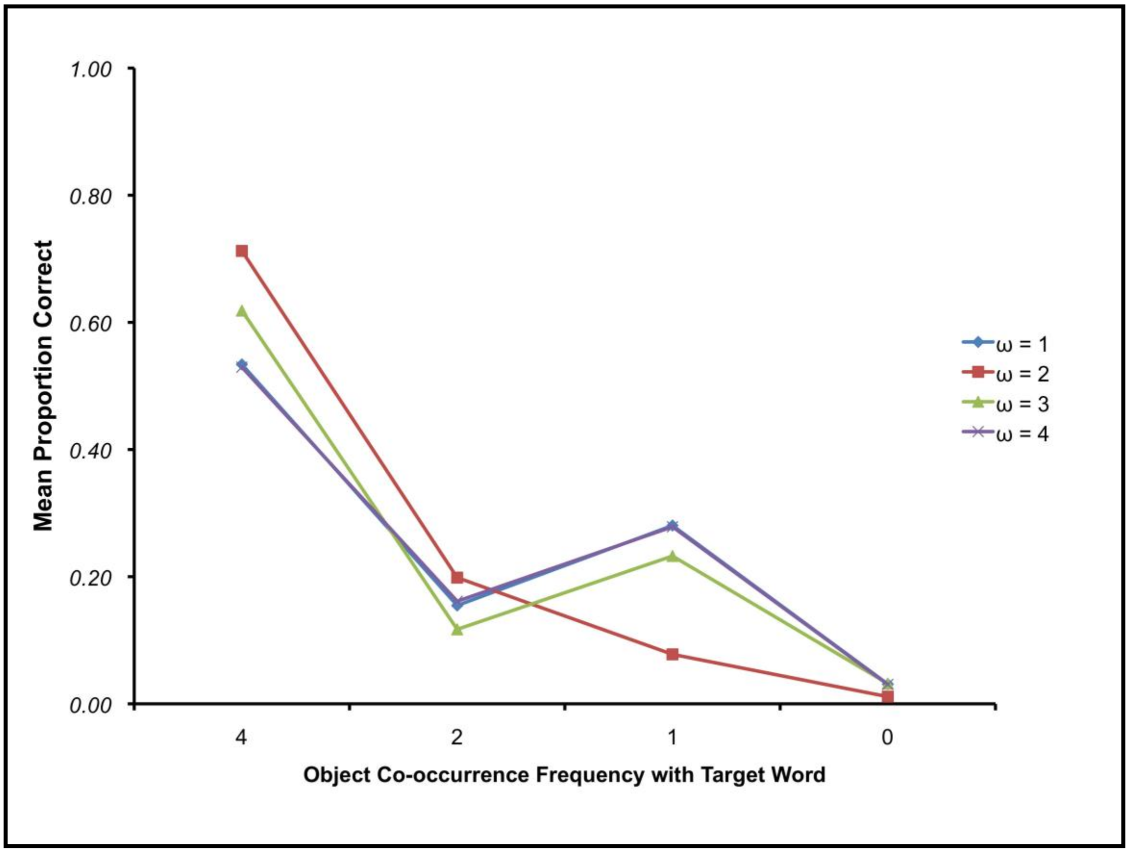

Figure 13. *Performance of HT models varying in word knowledge* (ω) *threshold in Simulation 2: Distribution of answers to objects differing in co-occurrence frequency with target word.*

5.4.2.2 AL model results

All AL model variants tested revealed an identical graded selection pattern as a function of object co-occurrence frequency with the target word (see Figure 14). Selection patterns of these models were similar to the patterns exhibited by children, as an object's co-occurrence frequency decreases, the likelihood of selecting that object also decreases.

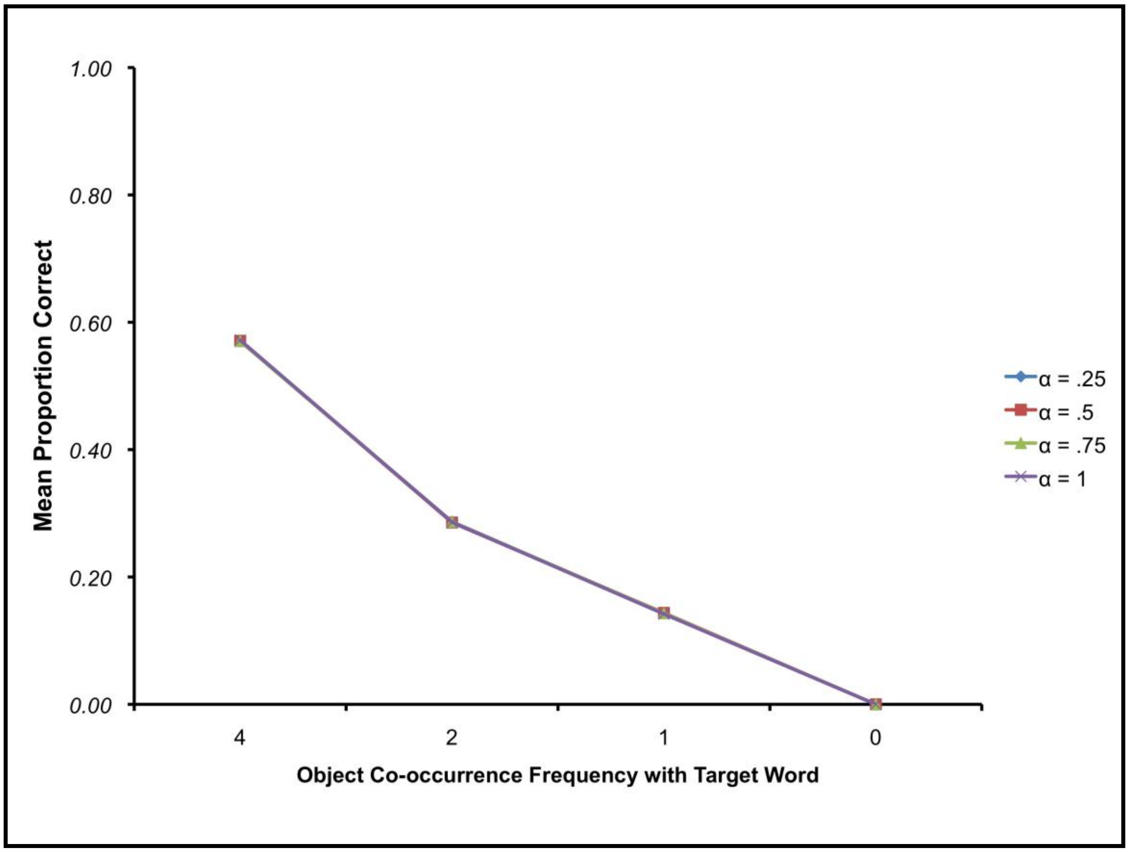

Figure 14. *Performance of AL models varying in learning rates (α) in Simulation 2: Distribution of answers to objects differing in co-occurrence frequency with the target word.*

5.4.2.3 Correlations between models and children's performance

Correlations between simulated learners' performance and human children's performance is displayed in Table 5. In contrast to the results of Simulation 1, the results of Simulation 2 suggest that both models, and especially the AL models, can account moderately well for the children's selection patterns. Consistent with what can be gleaned from Figure 13, HT model success appears dependent on the hypothesis threshold. Only the HT model with the moderately low threshold ($\omega = 2$) correlates significantly with children's performance, though the other model variants do show a trend towards displaying children's performance. These trends are likely the result of all HT model variants showing an overall downward slope despite differences across model variants in the predicted outcomes for the two middle frequency items (see Figure 13).

Table 5. *Correlation results between models and children's performance in Chapter 4. Note: *p < .05, ^p < .10*

Models Tested	*Correlation Statistics*	
	r	*p*
HT Models varying in hypothesis threshold (ω)		
HT Model 1 ($\omega = 1$)	.772	.11
HT Model 2 ($\omega = 2$)	.927	.03*
HT Model 3 ($\omega = 3$)	.793	.10^
HT Model 4 ($\omega = 4$)	.781	.11
AL Models varying in learning rate (α)		
AL Model 1 ($\alpha = .25$)	.979	.01*
AL Model 2 ($\alpha = .5$)	.978	.01*
AL Model 3 ($\alpha = .75$)	.979	.01*
AL Model 4 ($\alpha = 1$)	.979	.01*

As done in Simulation 1, I also conducted a series of simulated experiments to examine the likelihood of obtaining results similar to that of children under different assumptions (i.e., the model variants) of the population. For each model variant, correlations between children's item selection effects and item selection effects observed in 1,000 simulated experiments (28 simulated learners each) were classified as strong, moderate, or weak. Using these criteria, all variants of both models displayed strong correlations with children's results (see Figure 15).

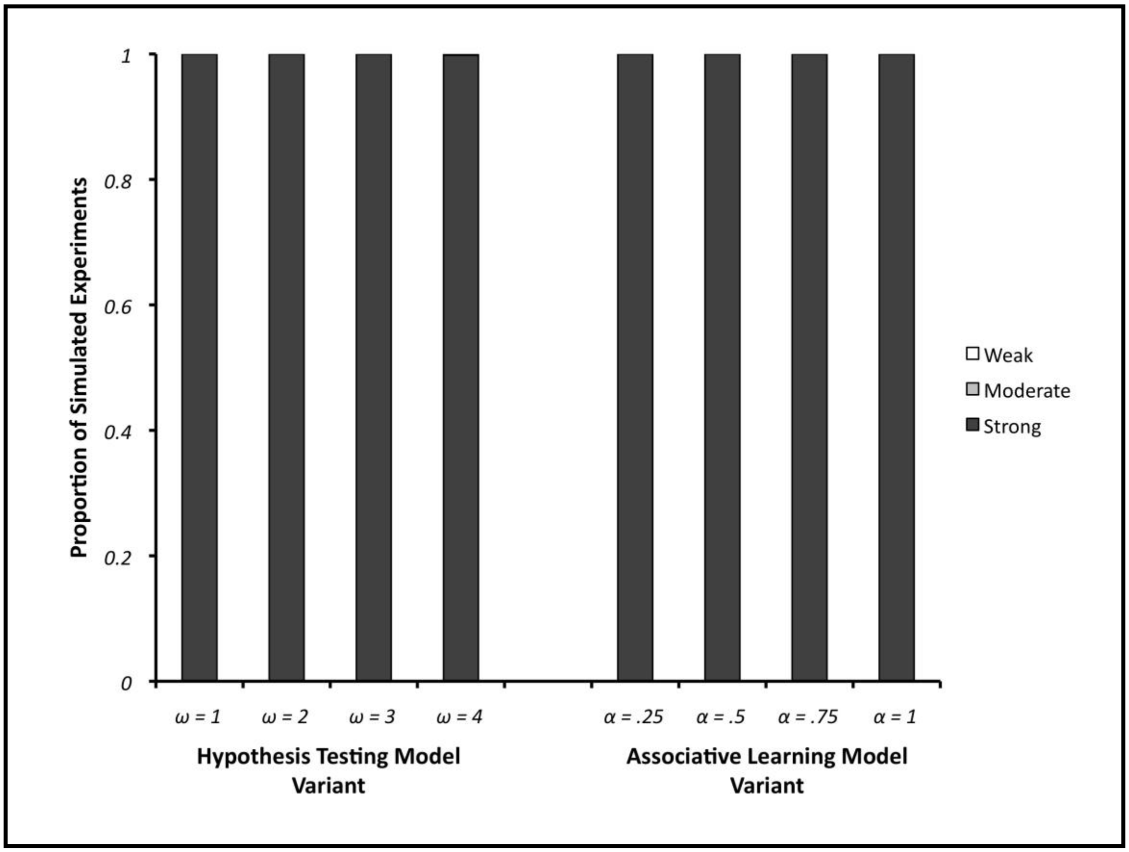

Figure 15. *Proportion of simulated experiments of HT and AL models whose item selection effects correlated strongly, moderately, and weakly with children's item selection effects.*

These results suggest that the lack of statistical significance in some variants of the HT models in the first set of simulations may indeed have been due to the lack of statistical power. These results also point to the limitation of the correlation analyses, particularly the simulated experiments, which used more liberal criteria of fit. A strong correlation between models and children can be driven by the overall gestalt of performance despite some differences in the learning pattern specifics. Further, even in cases where models fit using the more strict overall correlation analysis, fit in the current simulation is approximate. Because the data points in this task were dependent on one another, that the simulated learners across all variants of both models (M_{HT} = .60, M_{AL} = .57) outperformed children (M = .33), meant that the proportion of simulated learner's answers that included the less-frequently co-occurring objects is necessarily smaller than that of children. As a result, the slopes of models' item selection functions are naturally steeper than children's functions. Thus, relative fit between models and children's performance in this task strictly refers to the fact that both models and children exhibit a negative slope, rather than that they show a similar degree of steepness in the slope.

5.4.3 Discussion

The results of Simulation 2 suggest that both learning accounts predict the item selection effects observed in children. Graded item selection patterns, observed both in models and in children, are often considered to be a marker of learners keeping track of multiple word-to-object links per word (Vouloumanos, 2008). Further, since links between one word and multiple objects is a key feature that distinguishes associative learning accounts from hypothesis testing accounts, graded item selection patterns could be interpreted as evidence for associative learning accounts of cross-situational word learning. However, the simulations of the Hypothesis Testing models here illustrate that one need not maintain representations of both highly frequent and less frequent candidate referents. Instead, the graded item-selection effect, under a hypothesis testing account, emerges from the fact that the probability with which a word-object mapping is hypothesized is proportional to the frequency of that word-object mapping in the environment.

That the associative learning model succeeds at accounting for children's behavior in Simulation 2 is important in part because it demonstrates that the simple associative model developed in this chapter can account for at least *some* patterns of behavior. That is, the failure of the associative model at accounting for the contextual diversity effects in Simulation 1 is not the result of an overly simple learning algorithm unable to predict any human learning patterns. The question remains however as to whether there are any conditions under which an associative learning model can account for both contextual diversity and item selection effects.

5.5 Simulations 3 and 4: A Modified Associative Learning Model

In this section, I describe two modifications to the above AL model's algorithm hypothesized to improve its fit with the contextual diversity patterns of children's learning. Then, I describe and discuss the results of simulating this Modified Associative Learning (MAL) model in the experiments of Chapters 3 and 4.

As alluded to in the discussion of Simulation 1, one major reason that the basic associative model fails to account for the contextual diversity effect is that the models' selection strategy at test considers only the relative association strength between the test options and the target word. Given that across conditions of contextual diversity only one item present at test had co-occurred with the target word during learning, all simulated learners, regardless of condition, selected the same target referent 100% of the time. Thus, the first modification to the basic associative model is that at test, the model considers more than just the associations between the target word and other objects present on the test trial. That is, the model is affected by associations between the target word and all objects during learning. In other words, the MAL model incorporates a greater amount of competition in its learning algorithm (see also the models in Merriman, 1999; Regier, 2005). The decision process of the MAL model can be written as follows:

$$p(o_1) = \frac{M(w_1,o_1)}{\sum_{i=1}^{N} M(w_1,o_i)} \qquad \text{(Eq. 3)}$$

Where $p(O_1)$ is the probability of selecting Object 1 on a test trial, $M(w_j,o_i)$ is the strength of the association between word *j* and object *i*, and *N* refers to the number of objects in the learner's lexicon.

Although the greater amount of competition in this modified model is likely to bring performance off ceiling, the addition of the competition term by itself is unlikely to explain children's contextual diversity effect. That is, although different conditions of contextual diversity differ in the strength of the highest co-occurring foil, they do not differ in the total number of target word to foil co-occurrences. Thus, the denominator in Equation 3 will be equivalent across conditions. An additional process then must be added for the MAL model to induce the contextual diversity effect.

I propose to introduce one such process, a familiarity bias in updating associative strengths. With a familiarity effect, which has been proposed in other models of word learning (e.g., Kachergis et al., 2012a; Yurovsky et al., 2010), the model adds extra strength to co-occurrences in learning trials that have been observed on previous learning trials. In the present model, this familiarization bias is set to the negative natural logarithm of that word-object's conditional probability of co-occurrence[8]. The updating rule employed when the model observes a previously co-occurring word-object pair can be expressed as:

$$M^t(w_1,o_1) = M^{t-1}(w_1,o_1) + \alpha - \lambda \log\left(\frac{M^{t-1}(w_1,o_1)}{\sum_{i=1}^{N} M^{t-1}(w_1,o_i)}\right) \quad \text{(Eq. 4)}$$

Under this formulation, α is the constant learning rate, and λ is a scaling parameter, determining the strength of model's familiarity bias in adding to its association strength.

5.5.1 Method

As with the previous models, I conducted simulations of the MAL model performing the experiments of Chapters 3 and 4. Because the MAL model possesses two free parameters (α and λ), I conducted a total of 16 variants of the MAL model, testing the same four variants of learning rate tested in the basic associative models (α = .25, .5, .75, 1), crossed with four variants of the familiarity scaling parameter (λ = .25, .5, .75, 1). Model variants with a greater scaling parameter show a greater familiarity bias. Thus, the larger the scaling parameter, the greater the difference in the associative strength added to familiar relative to unfamiliar word-object pairs in a given trial. All other aspects of the simulation details (i.e., task procedures, data

[8] The result of employing the logarithmic function is that the familiarity bias early in learning is greater than the familiarity bias later in learning. Although algorithms using this function showed the best fit, other functions (e.g., linear, exponential) also fit children's learning patterns, but to a somewhat lesser extent.

processing, and metrics for comparison) were the same as those in the simulations outlined above.

5.5.2 Results

5.5.2.1 Contextual diversity effects in MAL models

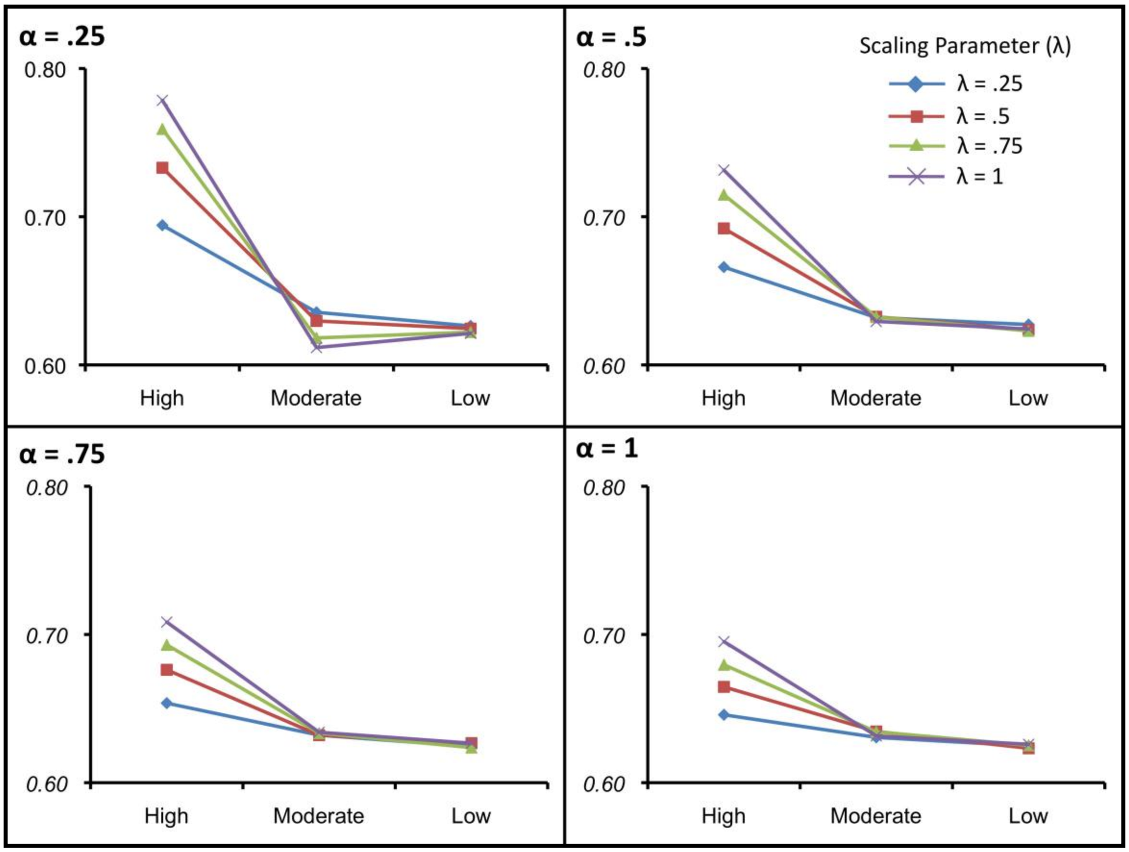

Figure 16. *Contextual diversity effects in MAL models differing in learning rates (α) and familiarity bias scaling parameters (λ). Note: x-axes are conditions of contextual diversity, y-axes represent mean proportion correct.*

Figure 16 depicts the performance of different MAL model variants across conditions of contextual diversity. As the figure illustrates, unlike the basic AL model tested earlier, the MAL models demonstrate a contextual diversity effect on performance: the greater the diversity of learning contexts, the better the performance. An inspection of these graphs also underscores the importance of the familiarity bias in accounting for contextual diversity effects. That is, across learning rate levels, the greater the familiarity bias, the greater the contextual diversity effect the model exhibits.

5.5.2.2 Item selection effects in MAL models

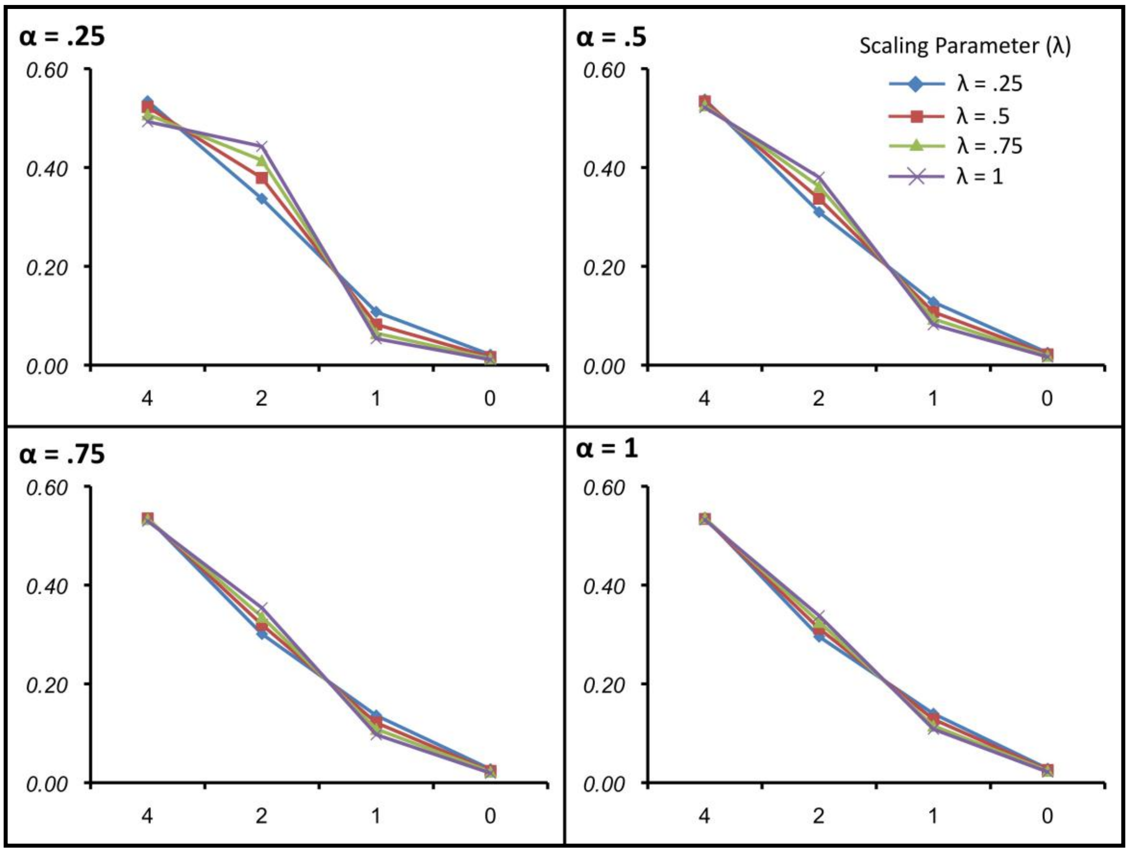

Figure 17. *Item selection effects in MAL models differing in learning rates (α) and familiarity bias scaling parameters (λ). Note: x-axes are object co-occurrence frequency with the target object, y-axes represent mean proportion selections.*

Figure 17 illustrates that the MAL models also readily account for graded item selection effects. Across all MAL variants, the greater the co-occurrence frequency between an object and the target word, the greater the likelihood the simulated learner selects that object. Relative to the effect the scaling parameter had on contextual diversity effects in learning, the scaling parameter and the learning rate had less of an effect on whether models exhibited the item selection effect. These item selection effects appear to be a property shared across all MAL models tested.

5.5.2.3 Correlations between MAL models' and children's performance

Correlation analyses between MAL models' and children's performance point to similar conclusions. That is, all MAL model variants appear to predict the patterns of performance observed in children, correlating significantly, marginally significant or strongly trending with children's performance (Pearson correlation coefficients between MAL models and children's contextual diversity effects ranged from .93 to .99, *p*s between .02 and .12; Pearson correlation coefficients for item selection effects ranged from .96 to .99, smallest p =.02).

Further, simulated experiments of the MAL model also revealed similar patterns. Specifically, the majority of all variants demonstrated either a strong or a moderately strong correlation with children's contextual diversity effect (see Figure 18). Additionally, as suggested by the patterns in Figure 16, across MAL models varying in learning rate, models with a greater familiarity bias had a greater likelihood of correlating strongly with children's contextual diversity effects. All variants of the MAL exhibited a strong correlation with children's item selection effects (all MAL variants correlated strongly 100% of the time with children's item selection effects).

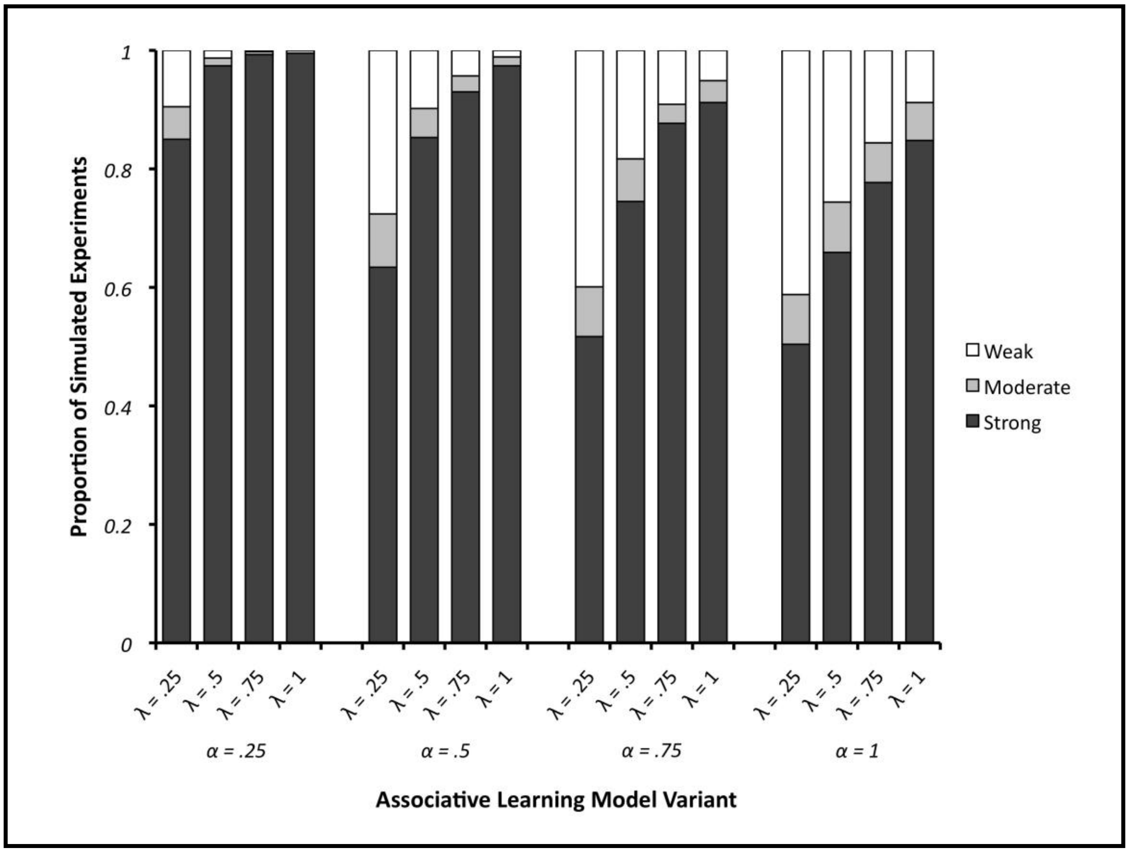

Figure 18. *Proportion of simulated experiments across MAL models whose results correlated strongly, moderately, and weakly with children's contextual diversity effects.*

5.5.3 Discussion

Thus, as hypothesized, a variant of associative learning can account for both children's contextual diversity and item selection effects. The variant proposed here involves incorporating greater competition (Merriman, 1999; Regier, 2005) and a familiarity bias (Kachergis et al., 2012a; Yurovsky et al., 2010) into the associative learning framework. Together with the results from the Hypothesis Testing model, this finding demonstrates that two different learning mechanisms that differ in learning algorithm, information intake, and information retained, can both readily fit behavior. The results of this endeavor demonstrate, however, that models constructed based on these properties alone are insufficient to account for children's behavior. Instead, additional parameters (hypothesis threshold in hypothesis testing models, familiarity and competitive processes in associative learning models) are needed to make the models' learning resemble children's learning. I argue that the role additional parameters like these play are often ignored in debates of hypothesis testing and associative learning accounts but they are nonetheless critical and must be incorporated in debates and discussions of cross-situational word learning mechanisms. I return to this issue in the General Discussion for this chapter.

5.6 Simulation 5: Accounting for other empirical data

The models described above were specifically constructed and selected to fit the learning patterns of human children in Chapters 3 and 4. To what extent do these same model architectures, properties and parameters also predict other empirical findings in the study of cross-situational word learning? In this section, I investigate the ability of one successful variant of each HT and MAL model to account for a learning pattern observed in Yu and Smith (2007)'s initial study of cross-situational word learning (see also K. Smith, et al., 2009; Vlach & Sandhofer, 2010).

In addition to providing the initial demonstration that adult learners were capable of rapid word learning through tracking of cross-situational statistics, Yu and Smith observed that learning varied depending on the level of referential uncertainty present on each learning trial. Each subject in Yu and Smith's study completed three learning conditions differing in within-trial referential uncertainty. In each condition, subjects attempted to learn 18 word-to-referent pairs from ambiguous naming events. In one condition, each naming event involved two words and two pictures (low ambiguity). In a second condition, each naming event involved three words and three pictures (moderate ambiguity). In a third condition, each naming event involved four words and four pictures (high ambiguity). Across all conditions, word knowledge was tested in a series of 4AFC trials. Yu and Smith found, perhaps not surprisingly, that as within-trial referential ambiguity increased, learning decreased. Of interest in the present set of simulations is whether the successful hypothesis testing models and associative learning models that accounted for the contextual diversity effects and item-selection effects in children, also exhibit Yu and Smith's effects of ambiguity on learning.

5.6.1 Method

Following the constraints employed by Yu and Smith (see Yu & Smith, 2007, Experiment 1), I first created two randomized training lists for each condition of ambiguity, differing on which word and picture pairings appeared together on each trial. I also created two randomized testing lists, differing on the order in which words were tested and which pictures served as foils on each trial. There were no alterations made to the HT and MAL models' learning algorithms. However, slight adjustments were made to ensure the models could accommodate task differences between Yu and Smith's study, and the studies in Chapters 3 and 4 (i.e., larger hypothesis lists and association matrices to accommodate the larger to-be-learned lexicons, and increased within-trial information intake capacities to account for the greater number of words and objects present on each trial in some of Yu and Smith's conditions).

In the current simulations, I tested the single variant of each of the HT Model (ω = 2) and the MAL Model ($\alpha = .5, \lambda = 1$) that most successfully accounted for children's contextual diversity and item selection effects. As with all previous simulations, 30,000 simulation runs were completed for each condition, with an equal number of simulation runs per training-testing list permutation. For each simulation, the proportion of trials learners answered correctly was tallied and a mean proportion correct for each condition was derived.

5.6.2 Results and Discussion

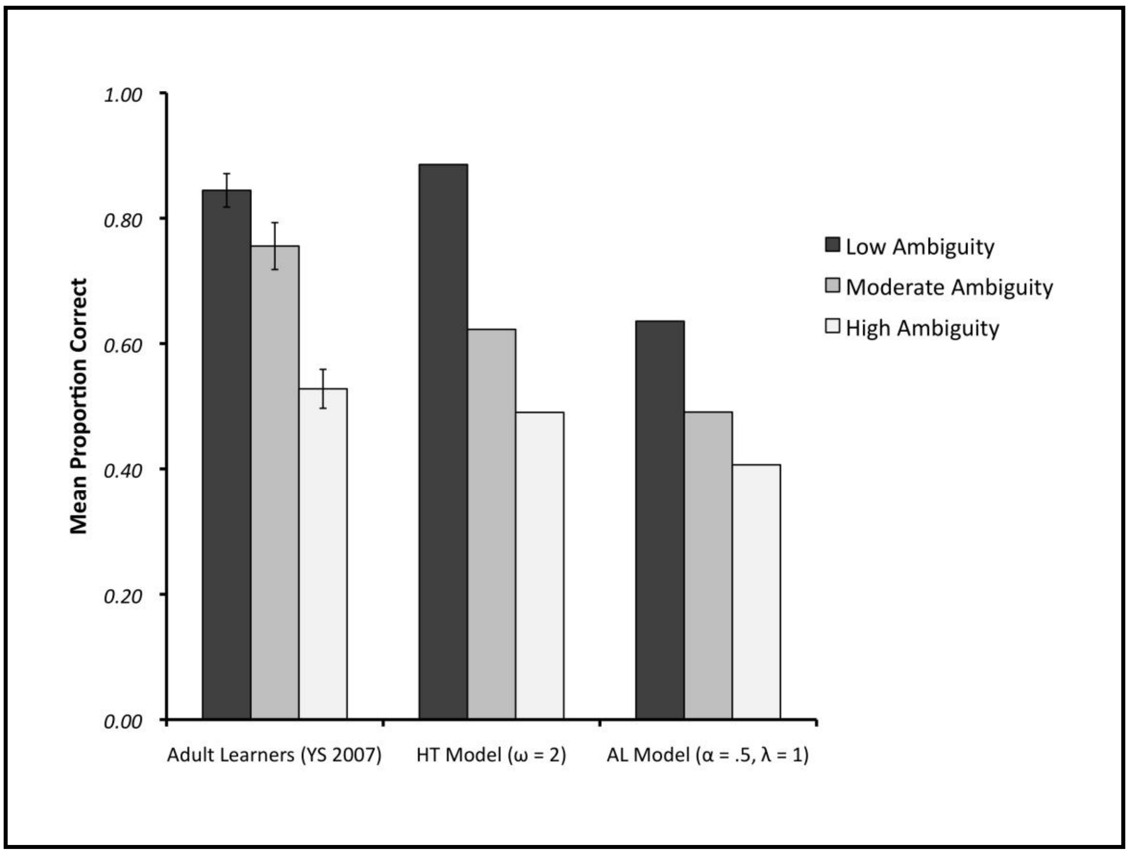

Figure 19. *Comparison of human adult learner performance reported in Yu & Smith (2007) to simulations of an HT and MAL model.*

Figure 19 depicts a side-by-side comparison of the performance patterns of adult learners, the HT model, and the MAL model. As the figure illustrates, although the precise learning rates differed between humans and models, both models show the effect of referential ambiguity on learning: as within-trial ambiguity increased, performance decreased. These findings suggest that the particular model assumptions and components that accounted well for children's data are not simply a set of processes that have been tweaked to uniquely fit a single data set comprised of children's performance within a particular experimental context but instead are processes that might simulate a wider range of behavioral phenomena.

5.7 General Discussion

The goal of the simulations presented in this chapter was to investigate the nature of children's cross-situational learning mechanisms, testing the efficacy of both hypothesis testing and associative learning frameworks. To distinguish between the accounts, the hypothesis testing and associative learning models were constructed to differ on three core learning properties: learning algorithm, information selection, and information retention. Consistent with prior research, comparisons between model performance and children's performance suggests that the core learning properties of both hypothesis testing and associative learning frameworks can be consistent with empirical evidence. However, the results also suggest that although the learning dynamics instantiated in these models were sufficient to mimic some aspects of children's learning patterns (item-selection effects), only when these core properties were supplemented with additional parameters were they sufficient in accounting for the full set of findings (item-selection effects + contextual diversity effects) as well as other published results (Yu & Smith, 2007). In this section, I first describe some empirical evidence in support of the plausibility of the current models. I then end with a discussion of the implications of the current findings for the hypothesis testing – associative learning debate.

5.7.1 Empirical evidence for the proposed parameters that improve model fit

The proposed parameters that improve fit in the current instantiations of hypothesis testing and associative learning models are probably not the only ones that could explain children's learning patterns. There are likely other processes that could have accounted for the data as well. However, the benefit of conceptualizing the processes as competition effects, familiarity biases (in the case of associative learning), and threshold effects (in the case of hypothesis testing) is that there is some prior support for them. Both competitive and familiarity processes have been invoked in a number of previous models of word learning (Fazly et al., 2010; Kachergis et al., 2012a; Merriman, 1999; Regier, 2005; Yurovsky et al., 2010). Further, the notion that competition and familiarity play a role in word learning has been observed empirically. For example, in one study of early word learning, Horst and colleagues (2010) presented toddlers with an array of 2, 3, or 4 objects, labeling one of these objects with a novel name. Following a short delay, toddlers in each condition were tested for their ability to select the referent of that novel name from a new array of objects, which included the labeled object and 4 previously unseen objects. Horst and colleagues found that only children in the 2-object condition remembered the word-object mapping, a finding consistent with the notion that learning was influenced by the greater competition present in the 3-object and 4-

object conditions. In a different study, Kucker and Samuelson (2012) found that 24-month-old infants' ability to learn novel word – novel object mappings was boosted when those infants were first familiarized with the novel object (without labeling it). Although this effect is different from the familiarity effect observed in the current study, it nonetheless is consistent with the notion that there are benefits of familiarity in word learning.

The key component to the success of the current hypothesis testing models was the model's hypothesis threshold, the point at which a learner accepts a hypothesized referent of a word as the word's "true" referent. That a hypothesis threshold existed in the current models meant that in this instantiation of hypothesis testing, learners tracked the relative strength of a hypothesis. This assumption may be considered a departure from the purist notions of hypothesis testing that does not accommodate any sensitivity to the relative strengths of hypotheses. Thus, there is less empirical work validating the existence of the hypothesis threshold, though the notion that one exists is intuitive. In the early stages of learning a particular word, exposure to the word and its referent may be important evidence for children's learning. However, once a child has observed that word-referent pairing numerous times, any additional naming event might contribute very little additional word knowledge (for a brief discussion on this point, see also Merriman, 1999). Some behavioral evidence from adult learning is consistent with a threshold effect in word learning. In a recent investigation of cross-situational word learning, Suanda and Namy (2012) examined adult subjects' error patterns in a 4AFC testing regimen. Specifically, we examined the objects participants erroneously selected as a function of the co-occurrence frequency between the object and the target word. Error patterns suggested that the likelihood of learners erroneously selecting a high frequency co-occurring foil was significantly greater than the likelihood of selecting either a moderate, a moderately low, or a low frequency foil. However, there was no difference in the participants' likelihood of selecting the moderate, the moderately low, and the low frequency foil (Suanda & Namy, 2012), consistent with the notion that only foils that surpassed a high threshold of co-occurrence lured participants.

Despite some empirical support for these processes, the validity of the proposed models as possible mechanistic accounts of children's cross-situational word learning in this experimental paradigm still awaits future empirical substantiation, an important next step in the current research endeavor. A number of possible experimental adjustments could potentially be suitable to investigate the extent to which familiarity, competitive and threshold effects are operating in children's learning. For example, eye-tracking measures that provide moment-by-moment measurements of where children are looking while listening to a particular word might reveal whether children show a familiar bias during learning, fixating longer to an object that had previously co-occurred with that word than to one that had not. Further, perhaps the use of reaction time as a dependent measure might shed light onto the amount of competition that occurs during referent selection, though establishing a reliable reaction time paradigm can be challenging in children. Finally, a learning paradigm that intermittently tested children during learning could potentially reveal whether indeed there is a learning threshold, and what that threshold is, in this task.

5.7.2 Implications for the hypothesis testing vs. associative learning debate

The current findings suggest that the learning dynamics created by the core learning properties of both hypothesis testing and associative learning accounts can be consistent with the empirical evidence. The critical finding from the current set of simulations is that what appears central to the current models' success in accounting for human behavior are parameters that are added to those core learning properties of learning algorithm, information selection, and information retention, that have received the bulk of attention in debates on hypothesis testing and associative learning accounts (e.g., see Nicol Medina et al., 2011). These parameters need to be considered part of the theoretical debate as well.

One could argue that the particular parameters proposed here are only valid if we accept the current assumptions of the core learning properties. That is, if we relax the differences between hypothesis testing and associative learning accounts, parameters such as competition, familiarity, or hypothesis threshold might not be needed. For example, we could create associative learners that can maintain associations between multiple words and multiple objects but only select a single word-object association to strengthen on any given learning trial. Likewise, we could create a hypothesis tester that can maintain two hypothesized referents per word (as opposed to only one). What might performance in these learners look like? Would such models exhibit contextual diversity effects and item selection effects in learning? These are interesting questions worthy of future investigation, though such endeavors blur the lines between what counts as a hypothesis testing model and what counts as an associative learning model. Of course, such blurring may be one way in which the hypothesis testing – associative learning debate could be advanced, namely by accepting that the two mechanisms are in fact polar extremes of a single learning process (for discussion see Yu & Smith, 2012, K. Smith et al. 2010).

The implication of the current findings is in part that we do not need to relax our definitions of hypothesis testing and associative learning in order to account for children's learning patterns. Even under conditions in which the two accounts are maximally different on the core learning properties, they can both still be consistent with the data. That both models of learning readily account for the current data, as well as a wealth of other findings (see Table 6), despite their disparate properties, raises a second, and non-mutually exclusive, way in which the debate may be advanced. That is, perhaps the two mechanisms are in fact separate but human learners have access to both. One possible manifestation of this view is a developmental one; which mechanism a learner employs depends on developmental stage. A number of developmental scholars of children's word learning have argued for important differences between the underlying processes of early word learning and that of later word learning (e.g., Namy, 2009; 2012; Nazzi & Bertoncini, 2003). Consistent with these previous accounts, perhaps early cross-situational word learning is best characterized by associative processes, whereas later word learning is best described as hypothesis testing (see also Smith & Yu, 2008; Yu & Smith, 2007; 2011).

Table 6. *The equipotentiality of hypothesis testing and associative learning in accounting for word learning phenomena.*

Word Learning Phenomenon	Model Reference	
	Hypothesis Testing	Associative Learning
Cross-Situational Word Learning I: learning words across referentially ambiguous naming events (observational findings)	*Siskind (1996); Frank et al. (2009)*	*Fazly et al. (2010); Yu (2008); Yu & Ballard (2007)*
Cross-Situational Word Learning II: learning words across referentially ambiguous naming events (experimental findings)	*Frank et al. (2009); Ichinco et al. (2009); Yu et al. (2007); Yu & Smith (2012)*	*Kachergis et al. (2012a); Yu et al. (2007); Yu & Smith (2012); Yurovsky et al. (2010)*
Disambiguation Effect: Selecting a novel object as opposed to a familiar object as a referent for a novel word	*Frank et al. (2009); Regier (2003)*	*Merriman (1999)*
Vocabulary Spurt: a slow period of vocabulary growth followed by an accelerated one	*Siskind (1996)*	*McMurray (2007)*
Shape Bias: a tendency to generalize a novel noun based on the referent's shape	*Kemp et al. (2007)*	*Samuelson (2002); Colunga & Smith (2005)*

It is also possible that rather than, or in addition to, learners relying on one mechanism early in development and another one later in development, learners might readily utilize both. A number of scholars have recently argued that word learning consists of processes occurring over at least two different time scales, one that involves referent selection and one that involves long-term word retention (McMurray, Horst, Toscano, & Samuelson, 2009; see also Carey, 2010; Swingley, 2010). Thus, another possibility is that hypothesis testing may underlie the in-the-moment referent selection aspect of learning whereas associative learning underlies the long-term consolidation of word knowledge in memory. Finally, that both models can be consistent with actual human learning patterns may also suggest that learners can deploy both learning mechanisms at any given point in development, depending on learning context. For example, perhaps when there are relatively few candidate referents, learners employ an associative learning strategy, linking multiple referents to each word. However, when there are many candidate referents, given the attentional and memory demands of encoding and maintaining associative links between words and all potential candidate referents, learners might employ a hypothesis testing strategy (see K. Smith et al., 2010, for some evidence that adult learners may indeed behave in this way).

5.7.3 Conclusion

Many mechanistic discussions of word learning more generally, and cross-situational word learning specifically, center on the distinction between associative learning and hypothesis testing accounts. The attention to these two accounts as potential learning mechanisms is well placed. Both models have accounted for a wide range of word learning phenomena, and the findings in Chapters 3 and 4 are no exception. The main conclusion from the current simulation study is that when the workings of each model were examined in detail, the features of these models that were critical to simulating the data are not the processes typically central to hypothesis testing – associative learning debates. As a result, advancing our mechanistic understanding of cross-situational word learning will involve going beyond asking whether hypothesis testing or associative learning accounts better model cross-situational word learning. Instead, it will involve addressing the many components that make each account work as well as assessing their psychological validity.

Chapter 6. General Discussion

How children acquire such impressively large vocabularies in such a short amount of time has long captured the interests of scholars across a range of fields. As reviewed in Chapter 2, a large body of literature suggests that part of the answer to this question is fast mapping -that to learn a word, children only need a single or a few exposures to it. A single exposure is sufficient because children can use a range of cues that allow them to infer reference at the moment a new word is encountered. As many have recently argued (Carey, 2010; Horst & Samuelson, 2008; Swingley, 2010; Yu & Smith, 2007), this fast mapping solution is likely only part of the answer to how children acquire their vocabularies. In addition to learning words from a single naming event via impressive inferential capabilities, children also learn words across multiple naming events via tracking the co-occurrence patterns between words and their referents.

Although the notion of cross-situational word learning is neither novel nor unintuitive, research into this type of word learning is a relatively recent development in lexical acquisition research. The studies presented herein are thus part of a larger endeavor (e.g., Blythe et al., 2010; Fitneva & Christiansen, 2011; Frank et al., 2009; Kachergis et al., 2012a; Monaghan & Mattock, 2012; Nicol Medina et al., 2011; Scott & Fisher, 2012; K. Smith et al., 2010; Smith & Yu, 2008, in press; Suanda & Namy, 2012; Vogt, 2012; Vouloumanos, 2008; Vouloumanos & Werker, 2009; Yoshida et al., 2012; Yu & Smith, 2007, 2011, 2012) to better document human learners' cross-situational word learning capacities, investigate their underlying mechanisms, and understand their role in vocabulary growth. I conducted the studies described here with two specific goals in mind. The first was to extend the research on cross-situational word learning, most of which has been conducted with adults as model word learners, to a developmental population, and to examine the extent to which children demonstrate some of the behavioral signatures previously observed in adult populations. The results of the studies presented in Chapters 3 and 4 demonstrated that like adults, children can rapidly acquire word-to-referent mappings across ambiguous naming events, when the only clue to reference was the cross-situational word-to-referent co-occurrence statistics. The findings also revealed that children's learning, like adults' learning (see Kachergis et al., 2009; Suanda & Namy, 2012), is affected by the contextual diversity of the learning environment. Further, children, also like adults (see Suanda & Namy, 2012; Vouloumanos, 2008), exhibited learning patterns that reflected the co-occurrence statistics of the learning environment.

The second goal of these studies was to go beyond examinations of children's learning behavior and to begin investigations of the candidate mechanisms underlying children's learning. The results of the computational simulations presented in Chapter 5 suggest that both a hypothesis testing and associative learning account of cross-situational word learning are consistent with children's learning patterns. Although these studies do not definitively arbitrate between the two accounts as a mechanistic account of children's data, the simulations do offer some conditions under which each account is compatible with the data.

In this last chapter, I reflect on some of the contributions the broad cross-situational word learning endeavor and the studies conducted for this thesis make in accounts of word learning. I then raise some limitations of current approaches to understanding cross-situational word learning research and suggest some logical next steps in advancing cross-situational word learning research.

6.1 Contributions of a Cross-Situational Approach

There are at least four important contributions a cross-situational learning approach makes to the study of word learning, each of which I discuss in turn. These include, first, that research on cross-situational word learning offers an additional cue to which children are attuned in learning words. Second, a cross-situational approach to word learning may not only help explain how children solve the problem of referential ambiguity, it may also help explain how children discover other cues to word learning. Third, a cross-situational learning approach may help researchers make a connection between early word learning and later vocabulary acquisition. And finally, a cross-situational learning approach is an important contribution because it meshes well with findings across other language learning tasks and thus may contribute to a unified account of children's language acquisition.

6.1.1 Cross-situational co-occurrence statistics as an additional cue to learning

The predominant approach to studying children's word learning has been to document the range of cues children use to map words onto their referents. Previously well-documented cues to reference include a range of perceptual-attentional cues (e.g., Landau, Smith, & Jones, 1988; Samuelson & Smith, 1998; Pruden et al., 2006), social cues (Akhtar et al., 1996; Baldwin, 1993b; Diesendruck et al., 2004), linguistic cues (Brown, 1957; Katz et al., 1974; Yuan & Fisher, 2009) and conceptual cues (Markman & Hutchinson, 1984; Markman & Wachtel, 1988). In recent years, this list has continued to grow. For example in one recent study, Kidd and colleagues revealed that two-year-olds used a speaker's disfluencies (i.e., filled pauses such as 'uh' and 'um') to disambiguate reference (Kidd, White, & Aslin, 2011). That is, when presented with two pictures, one novel and the other familiar, hearing a novel label embedded in a disfluent sentence ("Look at thee…uh…dax!") led two-year-olds to gaze to the novel object more quickly than when the novel label was embedded in a fluent sentence ("Look at the dax!"). In another recent study, Herold and colleagues reported that children were able to use speaker prosody to infer the meaning of a novel word (Herold, Nygaard, Chicos, & Namy, 2011). That is, Herold and colleagues found that 5-year-old children were able to infer that a novel word (e.g., "blicket") spoken in a deep, slow, and loud voice referred to a large object but the same novel word spoken in a high, fast, and quiet voice referred to a small object.

The recent findings suggesting that infants (e.g., Smith & Yu, 2008), children (Chapters 3 and 4), and adults (e.g., Yu & Smith, 2007) learn words cross-situationally indicate that cross-situational statistics are yet another cue that word learners can exploit to infer meaning. Thus, when a young learner hears his parent say "can you get the *train*", in addition to using the social cues parents may provide (e.g., what the mother is pointing at), the linguistic properties of the target word (e.g., that the word "train" is followed by the determiner "the"), the conceptual biases children may possess (e.g., the word likely refers to a whole object rather than an object part), the young learner might also weight the word's co-occurrence statistics: the range of objects that have previously co-occurred whenever the word "train" has been uttered.

That cross-situational statistics is only one of many cues children employ in the service of word learning raises the question of the weighting children place on these statistics relative to other cues to word learning, and whether this might change over development. To address this question, future studies could pit cross-situational statistics against other learning cues (e.g., eye gaze) to examine which cue children of different ages favor. One potential outcome is that co-occurrence statistics may be favored earlier in development whereas non-statistical cues may be preferred later in development (see Thiessen & Saffran, 2003 for an example of this pattern in the domain of speech segmentation). Such a shift in cue-preference could contribute to the observed slow pace of early word learning (co-occurrence statistics after all require multiple exposures to accumulate), followed by the more rapid pace of later learning (other cues such as eye-gaze could allow learners to map new words at the moment those words are first encountered). If this shift in emphasis on statistical cues does indeed exist in word learning, as it does in speech segmentation, it in turn raises the question through what process does the shift occur? Importantly, it underscores the likely possibility that children do not begin word learning sensitive to the many fast mapping cues; instead, they acquire these cues over the course of word learning. I consider this issue in more detail in the following section.

6.1.2 The role of cross-situational learning in the discovery of fast mapping cues

A second contribution of the work on cross-situational word learning is that it may add to our accounts of how fast mapping cues are acquired. Although the idea that word learners use cues to fast map words to their referents provides an elegant solution to the problem of referential ambiguity, it introduces a different, more rudimentary problem that developing word learners must solve, namely how to determine which of the host of possible cues is relevant to or diagnostic of word meaning. For example, it is logical and appropriate for word learners to utilize the familiar sentential frame, "this is a ____", to figure out that the speaker is referring to an object. However, learners could not use such a cue at the *onset* of lexical development because the mapping between the cue to meaning (sentential frame) and the item to-be-learned (mapping between word and object) must also be acquired. That the reliability of many cues to mapping is language- or culturally-specific, further underscores the importance of the discovery process of the diagnostic cues to word meaning. Surprisingly little work has focused on understanding cue discovery in the context of word learning.

Important exceptions to the neglect of this question do exist. For example, in her seminal work on the shape bias in children's word learning, Linda Smith has argued and provided empirical support that word learning constraints such as the shape bias emerge through the dynamics of basic cognitive processes such as attention, memory, and associative learning. In support of this account, Smith and her colleagues have shown the following. First, young word learners (i.e., children under the age of two) do not show the shape bias (Samuelson & Smith, 1999), highlighting that the bias is not present at the onset of word learning. Second, children's adherence to the shape bias is positively correlated with their vocabulary size (Smith, 2001). Third, there exist statistical regularities in how the early nouns children typically learn are organized; namely, that the majority of them label shape-based categories (Samuelson & Smith, 1999). Fourth, a training regimen that highlights shape-based regularities for young word learners can both speed up the emergence of the shape bias in artificial lab tasks, and boost children's object label acquisition outside the lab (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). Taken together, these four pieces of evidence persuasively imply that children discover the shape bias over the course of development as a result of acquiring a set of labels for categories of objects well organized by shape.

Although Smith and colleagues have convincingly demonstrated that the shape bias develops over the course of word learning, Smith herself has suggested that attention to shape may be present pre-linguistically (Smith, 2003). Thus, rather than being a bias learned entirely over the course of word learning, the shape bias may instead reflect a small initial attentional bias that is carried over from pre-linguistic categorization processes and further tuned through the process of learning words. Thus, whether Smith's attentional learning theory also accounts for the acquisition of the social or linguistic cues to word learning that are likely to be more cultural- and language-specific than a human infant's attention to shape is an open question.

A separate series of studies that has begun to address the issue of how children determine the reliability of a wider range of cues to word meaning comes from the work of Namy and her colleagues (Namy, Knight-Schwartz, & Smith, 2011; Namy & Waxman, 2000; see Namy, 2012, for review). For example, Namy and Waxman (2000) asked whether 18-month-olds could be taught a novel cue to word learning. Namy and Waxman paired a novel sentence frame (i.e., "Shalem bosher ______") with a familiar word and its referent. For example, an experimenter would hold up a spoon while saying, "Shalem bosher spoon!" Following a brief familiarization period with the novel sentential cue, 18-month-olds were able to use this naming phrase to infer the referent of novel words. For example, after hearing "Shalem bosher blicket" paired with an object, infants interpreted the word "blicket" as that object's label. This finding suggests that word learners discovered the reliability of this novel sentence frame as a cue to learning, which was, in turn, recruited in the service of subsequent learning a novel word.

Namy and colleagues' work thus demonstrates how children can use the predictive nature of a novel cue to determine its reliability in indicating word meaning. Although their work suggests how word learners can detect the reliability of a *single* cue to meaning (i.e., a particular sentential context), the question still remains how it is that word learners determine which among *many* candidate cues are reliable predictors of word learning. That is, the task children face consists not only of deciding whether a specific cue to reference, such as a speaker's eye gaze is a reliable one to attend. Rather, children must somehow also discover that a speaker's eye gaze is a more reliable cue than a speaker's touching of a referent (Booth, McGregor, & Rohlfing, 2008), and that a speaker's eye gaze is a less reliable cue than a speaker's pointing behavior (Frank, Tenenbaum, & Fernald, in press).

Thus, how it is that children discover which among many cues is diagnostic of word meaning is an important extension of Smith and Namy's previous work. Building off of their work, one possibility is that children solve this problem of cue ambiguity much the same way they solve the problem of referential ambiguity, by tracking cross-situational statistics. That is, much like solving referential ambiguity is argued here to involve a process of detecting predictable patterns between words and their referents, solving the problem of cue ambiguity involves a process of detecting predictable patterns between cues to naming and reliable word-to-referent mappings.

Whether this is indeed the right characterization of cue discovery and what the precise nature of the process awaits empirical investigations. For example, would learners have to learn a stock of word-to-referent mappings before they begin to track the cues to mapping, or might learning word-to-referent mappings and cues to mapping take place simultaneously? Further, are all cues to naming acquired in roughly the same way or are some cues acquired through a different (e.g., more strategic) process? Finally, how might the process of acquiring cues to word learning be similar to or different than acquiring the cues to other language learning tasks such as speech segmentation which have also been shown to rely in part on statistical regularities? Answers to these questions will help fill an important gap in the work on children's word learning.

6.1.3 Cross-situational learning as a common theme underlying vocabulary acquisition research programs

Unlike many components of language that are thought to be fully developed by the end of childhood (e.g., speech perception, grammatical development), vocabulary acquisition continues into adulthood (e.g., Zechmeister, Chronis, Cull, D'Anna, & Healy, 1995). Thus, the study of how words are learned has been of interest not only to scholars of early language development, but also to education and reading scientists (Wagner, Muse, & Tannenbaum, 2007), scholars of second language acquisition (e.g., Laufer, 2009), and even educational gerontologists (e.g., Laumann, Long, & Shaw, 2000). Despite this common topic of interest, there has been little work connecting the findings across these fields. This may, in part, be due to the fact that studies of early fast mapping capacities are not applicable to the study of later vocabulary growth. That is, that toddlers can fast-map words by attending to his parent's social cues has little to do with how third-graders learn the name of a scientific phenomenon from a text book.

Interestingly, although research on cross-situational word learning is relatively new in the study of early word learning, the proposal that words are learned incrementally over multiple exposures is a relatively old topic of study in educational research, commonly studied under the term *incidental word learning* (e.g., Nagy & Herman, 1987). In fact, a comparison of the recent cross-situational word learning work in the cognitive and developmental sciences and the incidental word learning work in educational research reveals a number of shared topics of inquiry. For example, researchers across both fields are interested in the nature of partial word meaning and its role in learning (see Yu, 2008; Yurovsky et al., 2010; and Frishkoff, Collins-Thompson, Perfetti, & Callan, 2008; Schwanenflugel et al., 1997). Researchers in the two fields are also interested in the relative success of active attempts to acquire word meaning compared to more passive experiences of accumulating word knowledge (see Akhtar, 2004; Kachergis, Yu, & Shiffrin, 2012b; and Swanborn & De Glopper, 2002). Further, researchers across fields have examined the effects of frequency of exposure on word learning (see Kachergis et al., 2009; and Rott, 1999). Finally, researchers in both fields have also examined how the diversity of learning contexts impacts learning (see Kachergis et al., 2009; Suanda & Namy, 2012; and Bolger et al., 2008; Wilkinson & Houston-Price, in press).

Of course the extent to which these similarities are a manifestation of a deeper common learning mechanism is a matter to be determined through future investigations. Nonetheless, these similarities open the door for a potentially productive dialogue between researchers of early word learning and those of later vocabulary acquisition. For example, cognitive and developmental researchers could draw on the larger set of findings in the incidental vocabulary acquisition literature to help guide hypotheses in studying the nature of cross-situational word learning. In turn, the findings from early cross-situational learning research may have implications for educational research on vocabulary acquisition.

6.1.4 Cross-situational learning as statistical learning: Toward a unified account of language acquisition

The finding that infants, children and adults can learn words by tracking cross-situational co-occurrence statistics strikes a chord with the growing body of evidence on statistical learning in other areas of language acquisition research (for recent reviews, see Aslin & Newport, 2008, 2012; Romberg & Saffran, 2010; Saffran, 2009, 2010; Thiessen, 2009). That is, researchers in the fields of speech perception (Maye et al., 2002; Maye, Weiss, & Aslin, 2008), speech segmentation (Aslin et al., 1998; Saffran et al., 1996), and grammatical acquisition (Gomez & Gerken, 1999; Thompson & Newport, 2007) have also discovered that human learners show a striking sensitivity to the statistical properties of their language environment and that they can use this sensitivity in the service of learning.

One of the upshots of this common statistical learning theme across aspects of acquisition is that it provides a common lens through which phenomena from disparate fields can be viewed. For example, in the study of early word learning, Baldwin (1991; 1993a) has revealed that 18-month-olds can learn object names even when there is a lag in the timing between the infants' attention to the object and the speakers' production of the word. Further, Tomasello and Krueger (1992) have found that the majority of verbs parents utter occurs before an action is performed, rather than during the event. That word learning occurs when words and their referents occur in a temporally non-contiguous fashion is often interpreted as evidence against statistical or associative accounts of word learning (see P. Bloom, 2000; Sabbagh & Baldwin, 2005; Tomasello & Akhtar, 2000). However, such learning could also be interpreted as a case of non-adjacent dependency learning, a form of statistical learning that has been demonstrated in speech segmentation (Newport & Aslin, 2004), tone segmentation (Creel, Newport, & Aslin, 2004) and visual sequence learning (Turk-Browne, Junge, & Scholl, 2005).

Interestingly, findings in these domains suggest that non-adjacent dependency learning is more difficult than adjacent dependencies (Creel et al., 2004; Newport & Aslin, 2004), which is consistent with findings in the word learning domain that the learning of object names is more difficult when object and word presentation are non-contiguous (Whitehurst, Kedesdy, & White, 1982) and that verb learning generally lags behind noun learning (see Snedeker & Gleitman, 2004). These observations suggest the possibility that more findings across disparate aspects of language acquisition may also be understood through a unified statistical learning approach.

Studies of statistical language learning have received a great deal of attention in recent years in part because it is considered to be a theoretical departure from traditional notions that the child's linguistic input is too impoverished and that general learning processes are insufficient to account for the linguistic competence language users display (see Bates & Elman, 1996, for discussion). That is, the many observations of statistical learning across areas of language research suggest that perhaps general learning processes play a greater role in language acquisition than held by the prevailing view. Indeed, statistical learning is not limited to the domain of language but instead is demonstrated in tasks such as tone sequence learning (Saffran, Johnson, Aslin, & Newport, 1999; Creel et al. 2004), visual sequence learning (Kirkham, Slemmer, & Johnson, 2002; Turk-Browne et al., 2005), visual scene parsing (Fiser & Aslin, 2001, 2002), tactile sequence learning (Conway & Christiansen, 2005) and event segmentation (Baldwin, Anderson, Saffran, & Meyer, 2008; Meyer & Baldwin, 2011). Also consistent with the domain-generality of statistical learning is the observation that individuals with certain language disorders are impaired not only in linguistic statistical learning tasks, but in non-linguistic ones as well (Christiansen, Louise Kelly, Shillcock, & Greenfield, 2010; Evans, Saffran, & Tore-Robes, 2009; Tomblin, Mainela-Arnold, & Zhang, 2007).

Although the findings of statistical learning research are generally consistent with the notion that domain-general learning capacities play an important role in language acquisition, the findings should not be taken as evidence that there is no role for domain-specific processes. In fact, a number of findings have indicated that both general and specific mechanisms play a role in the learning process. For example, Saffran and Thiessen (2003) found that statistical learning in speech segmentation is constrained by phonotactic knowledge (i.e., knowledge of the acceptable and unacceptable sound patterns within a language). That is, they found that infants computed transitional probabilities only between syllabic units that were consistent with the phonotactic patterns of the language. Thus, Saffran and Thiessen's work illustrates one case of domain-specific knowledge (i.e., phonotactic awareness) constraining domain-general statistical computations (i.e., tracking transitional probabilities). Other researchers have found that some statistical computations themselves may be domain-specific. For example, Meyer and Baldwin (2011) found that human learners could not compute conditional probabilities between segments in a task of action segmentation. In contrast, researchers in both speech segmentation (Aslin et al., 1998) and word learning (Klein & Yu, 2009) have found that learners readily compute conditional probabilities between units. Thus, there are multiple ways in which domain-generality and domain-specificity jointly contribute to children's language acquisition. The challenge for future work will be to go beyond the traditional notion that a learning process is either domain-general or domain-specific and to understand the myriad of ways in which domain-general and domain-specific mechanisms interact to produce behavior.

6.2 Limitations to the Cross-Situational Approach

Although cross-situational word learning research contributes to accounts of word learning and language acquisition in numerous ways, there also exist a number of limitations to this research. In this final section, I discuss some of these limitations, which include limitations to the current methodological approaches to cross-situational learning, limitations in the explanatory scope of the proposed process, and limitations in an understanding of the broader impacts of the research. Throughout this section, I also propose ways in which future research might begin to address these limitations.

6.2.1 Can cross-situational learning scale up to real-world learning environments?

The picture painted here of cross-situational word learning is that it is a statistical learning process. The empirical evidence presented in Chapters 3 and 4, and elsewhere (Smith & Yu, 2008; Suanda & Namy, 2012; Yu & Smith, 2007), that demonstrate human learners' ability to acquire word-to-referent mappings via sensitivity to co-occurrence statistics is central to this picture. Other lines of research also support this account. A number of observational studies for example, have documented that statistical properties of children's inputs, such as the frequency of words (Goodman, Dale, & Li, 2008; Huttenlocher, Haight, Bryk, Seltzer,

& Lyons, 1991), the contextual diversity with which words occur (Hills et al., 2010; Hoff & Naigles, 2002), and the density of particular words (Weizman & Snow, 2001) tends to correlate with better word learning. Additionally, computational investigations of word learning have revealed that models whose algorithms make use of distributional patterns in the input mimic the patterns of learning observed in children (Frank et al., 2009; Siskind, 1996; Yu, 2008).

When viewed from a distance, the experimental, observational, and computational findings provide compelling evidence for a statistical learning basis for vocabulary acquisition. However, a closer inspection of these approaches reveals few links among them other than their common theoretical underpinning. For example, computational modelers often use artificially generated inputs that are only loosely motivated by the observational data (for notable exceptions, see Frank et al., 2009; Yu, 2008; Yu & Ballard, 2007). Similarly, researchers who examine cross-situational word learning in artificial laboratory tasks employ task parameters that are not well motivated by the observational data either. This raises the question of whether the capacities driving learning in the laboratory tasks are the same as those underlying real-world learning. I argue that a tighter coupling between the observational, computational, and experimental methods is needed and is an important next step in testing the statistical learning account of children's word learning.

As a first step towards a more coupled approach, better estimates are needed not only of the language children hear but also of the corresponding visual input children see. Using infant head-mounted video cameras and eye trackers, researchers have begun to develop ways to capture a first-person visual of the infants' experience (Aslin, 2009; Franchak, Kretch, Soska, & Adolph, 2012; Yoshida & Smith, 2008). Combining this recently developed technology with the recordings of infants language environment provides one way in which we can begin to estimate at least part of the rich real-world co-occurrence statistics that characterize children's learning input. Once these statistics have been estimated, we can then use them to guide the learning parameters in experimental paradigms (see Gillette et al., 1999). Of interest is whether infants, children, and adult learners could acquire word-to-referent mappings from the co-occurrence statistics characteristic of those to which language learners are actually exposed. Finally, computational models can then be used as a window into the mechanism underlying learning from these real-world co-occurrence statistics.

6.2.2 Is cross-situational learning applicable to the acquisition of non-object labels?

A second limitation in cross-situational word learning research is that we know little about the explanatory scope of the proposed learning process. That is, as is the case with the majority of work on children's word learning, much of the existing cross-situational learning research (e.g., K. Smith et al., 2010; Suanda & Namy, 2012; Vouloumanos, 2008; Yu & Smith, 2007) focuses on the acquisition of object labels (but see Childers, 2011; Childers & Paik, 2009; Scott & Fisher, 2012 for a few notable exceptions). However even from the onset of lexical development, children also learn action words (e.g., "throw"), personal-social words (e.g. "bye-bye"), modifiers ("cold"), and functors (e.g., "and"). Although there is good reason to start the study of word learning by examining object names (e.g., relative to other lexical categories, object names, or nouns, are the single largest category in children's developing lexicons, see L. Bloom, 2000), a complete account of how children learn words must also be able to explain the acquisition of other lexical categories as well.

There are at least three reasons that a cross-situational word learning account based on object name learning might not easily translate to the acquisition of other lexical categories. First, the candidate referents for lexical categories such as verbs are often less obvious than those for nouns (see Gentner & Boroditsky, 2001; Snedeker & Gleitman, 2004). For example, consider a child who sees a ball kicked by a boy and hears the sentence, "the boy kicked the ball". Whereas the candidate object referents are relatively obvious (the boy and the ball), the candidate action referent is less so. What is the action to which the word "kick" refers? Does it refer to *causing the ball to move*, *causing the ball to move with one's foot*, *causing the ball to spin*, *causing the ball to fly*, *intending to harm the ball*, etc (see Gentner & Boroditsky, 2001)? In an influential set of studies, Gleitman and colleagues showed adults muted video clips of mothers talking to their children with beeps inserted when certain words were uttered, and then asked the adults to guess the word uttered. What Gleitman and her colleagues have found is that relative to nouns, adult learners do a very poor job at picking out verbs (e.g., "throw", Gillette et al., 1999; Snedeker & Gleitman, 2004) and mental-content words (e.g., "believe", Pappafragou et al., 2007), as well as abstract nouns (e.g., "thing" Kako, 2005). Thus as Gleitman, Gentner and their colleagues (Gentner, 1982; Gentner & Boroditsky, 2001; Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005) have argued, there may be more to learning lexical categories such as verbs besides simply tracking the cross-situational observations in which these words occur.

A second reason that tracking co-occurrence statistics of non-nouns may be difficult is that their referents have limited temporal availability (Gentner & Boroditsky, 2001; Merriman & Tomasello, 2005). That is, in the example above, unlike the boy and the ball that remain visible throughout the event, the kick is a fleeting action. Thus, analyses of the candidate referents of "kick" must rely less on direct perception and thus likely result in less precise representations than object representations.

A final potential limit to extending cross-situational learning accounts to the acquisition of non-noun categories touches on what Merriman and Tomasello describe as the *self-other distinction* (Merriman & Tomasello, 1995). That is, because actions can be performed either by oneself or by another individual, the actions may yield very different information depending on the actor. For example, in the case of a self-performed action, the learner may have insight into the intention and goal of the action. In contrast in the case of another's actions, the learner will have much less of these internal insights but might instead have a visual of the actors full body performing the action, something that is missing when viewing one's own actions. Thus relative to comparing instances of an object across situations, which are similar regardless of actor, comparing instances of an action across situations may be more problematic.

Thus there are a number of issues to consider in extending the cross-situational word learning account to the acquisition of non-noun lexical categories. Although a number of studies stemming from the comparison and categorization literatures point to the general benefit of cross-situational information in the learning of action words (Childers, 2011; Childers & Paik, 2009) and object property terms (e.g., Waxman & Klibanoff, 2000), additional work is still needed to address the implications of the above issues for the nature and relative role of cross-situational word learning in the acquisition of different lexical categories.

6.2.3 Is cross-situational learning ability related to real-world vocabulary growth?

A final limitation to the current research, and cross-situational word learning research more generally, is that there is little direct evidence that implicates cross-situational and statistical learning abilities in word learning beyond the laboratory setting. This is particularly surprising given the large amounts of variance to-be-accounted for in children and adults' vocabulary sizes (Anglin, 1993; Fenson et al., 1994; Stanovich & Cunningham, 1992). Existing individual differences research has suggested a multitude of factors that account for some portion of this variance, including socio-economic and early experiential factors (Arriaga, Fenson, Cronan, & Pethick, 1998; Huttenlocher et al., 1991; Hoff, 2003; Hoff & Naigles, 2002; Pan, Rowe, Singer, & Snow, 2005; Rowe, Levine, Fisher, & Goldin-Meadow, 2009), social learning competence (Gliga et al., 2012; Morales, et al., 2000; Morales, Mundy, & Rojas, 1998; Mundy, Fox, & Card, 2003; Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, & Tager-Flusberg, 2007), cognitive and memory capacities (Bowey, 2001;

Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992; Gupta, 2003), literacy and reading achievement (Cunningham & Stanovich, 1991; Perfetti, Wlotko, & Hart, 2005; Stanovich & Cunningham, 1992) and meta-linguistic skills (McBride-Chang et al., 2005, 2008; Metsala, 1999; Smith & Tager-Flusberg, 1982). To the extent that vocabulary acquisition is a statistical learning process, as is proposed here, then we should expect that statistical learning capacities should also correlate with measures of vocabulary size and other metrics of word learning.

Studies of statistical learning in other domains provide some evidence that performance in statistical learning tasks correlates with language competence. First, a number of researchers have examined statistical language learning in individuals with language-related disabilities, such as specific language impairment (Evans, et al., 2009; Tomblin et al., 2007), agrammatic aphasia (Christiansen et al., 2010), dyslexia (Howard, Howard, Japikse, & Eden, 2006), and other language learning disabilities (Grunow, Spaulding, Gomez, & Plante, 2006; Plante, Gomez, & Gerken, 2002), and have found that, compared to non-impaired and typically developing controls, these individuals have impaired statistical learning capacities in addition to impaired language functioning. More recently, scholars have examined variability in the statistical learning of typically developing adult populations and have found that performance on these tasks is correlated with a number of measures of language comprehension and processing (Conway, Bauernschmidt, Huand, & Pisoni, 2010; Kidd, 2012; Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010), as well reading skills (Arciuli & Simpson, 2012), even after individual differences in other capacities, such as working memory, have been accounted for.

Thus, it is reasonable to hypothesize that performance on cross-situational word learning would correlate with vocabulary skills. Of course even if it did, such data would not illuminate the precise causal relation between statistical word learning and vocabulary acquisition. Nonetheless, such an approach would be a valuable first step in mapping out the role of statistical learning in natural vocabulary growth. Perhaps subsequent longitudinal research that examines whether pre-linguistic statistical learning capacities predict early vocabulary size or training studies that examine the impact of promoting cross-situational learning on vocabulary growth could provide better insight into the causal relation underlying any link between statistical learning and real-world word learning.

6.3 Conclusions

How do children learn so many words so quickly despite the inherently ambiguous nature of reference? One way this paradox can be resolved is to posit that children have at their disposal powerful mechanisms that allow them to infer reference at the moment new words are encountered. Forty years of empirical investigations into this issue leaves no doubt that at least older word learners do have access to such processes. However, children's word learning capacity may include additional solutions to solving the problem of referential ambiguity.

A growing body of evidence over the past 5 years suggests that cross-situational word learning may be once such candidate process. As demonstrated in the studies herein, children can acquire novel word-to-referent mappings across a handful of ambiguous naming events, even when the only cue to mapping is the cross-situational co-occurrence statistics. Investigation into children's cross-situational learning capacities is still in its infancy and lags far behind investigation of children's fast mapping capacities. Thus, many questions remain regarding the scope of this learning process and its role at different points in development. Further, little is known about the underlying mechanisms that give rise to such learning. As the current simulation studies underscore, multiple existing theoretical accounts can be sculpted to be consistent with the empirical data. These findings suggest that a combination of theory and model development may be needed, in addition to empirical investigations, to advance our understanding of the nature of cross-situational word learning. I argue that this work has great promise and potential. Notions of cross-situational word learning are reminiscent of recent documentations of statistical learning across many aspects of language learning. This suggests that the current cross-situational learning work may contribute not only to a more complete account of vocabulary growth, but also to a more unified account of language acquisition.

References

Akhtar, N. (2004). Contexts of early word learning. In D.G. Hall & S.R. Waxman (Eds.), *Weaving a lexicon* (pp. 485–507). Cambridge, MA: MIT Press.

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635-645.

Akhtar, N., & Montague, L. J. (1999). Lexical acquisition: The role of cross-situational learning. *First Language, 19*, 347-358.

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word Learner: A Debate on lexical acquisition* (pp. 115-135). New York, NY: Oxford University Press.

Anglin, J. M. (1993). Vocabulary development: A Morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10), v-238.

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science, 35*, 1105-1138.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*, 286-304.

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low- and middle-income families. *Applied Psycholinguistics, 19*, 209-223.

Aslin, R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science, 86*, 561-565.

Aslin, R. N., & Newport, E. L. (2008). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions* (pp. 15-29). Mahwah, NJ: Lawrence Erlbaum Associates.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: from acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*, 170-176

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875-890.

Baldwin, D. A. (1993a). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395-418.

Baldwin, D. A. (1993b). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832-843.

Baldwin, D. A., Andersson, A., Saffran, J. R., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106*, 1382-1407.

Bates, E., & Elman, J. (1996). Learning rediscovered. *Science, 274,* 1849-1850.

Belanger, J., & Hall, D. G. (2006). Learning proper names and count nouns: Evidence from 16- and 20-month-olds. *Journal of Cognition and Development, 7*, 45-72.

Bjork, R. A. (2011). On the symbiosis of learning, remembering, and forgetting. In A. S. Benjamin (Ed.), *Succesfull remembering and successful forgetting: A*

Festchrift in honor of Robert A. Bjork (pp. 1-22). London, UK: Psychology Press.

Bloom, L. (2000). The intentionality model of word learning: How to learn a word, any word. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word Learner: A Debate on lexical acquisition* (pp. 19-50). New York, NY: Oxford University Press.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: The MIT Press.

Bloom, P. (2004). Myths of Word Learning. In D.G. Hall & S.R. Waxman (Eds.), *Weaving a lexicon* (pp. 205–224). Cambridge, MA: MIT Press.

Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science, 34*, 620-642.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An Instance-based learning approach. *Discourse Processes, 45*, 122-159.

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development, 4*, 179-202.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development, 8*, 278-302.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding Processes in Human Memory* (pp. 85-124). New York, NY: Winston & Sons.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics, 22*, 441-469.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language, 29*, 591-610.

Brown, R. (1957). Linguistic determination and the part of speech. *Journal of Abnormal and Social Psychology, 55*, 1-5.

Campbell, A. L., & Namy, L. L. (2003). The role of social-referential context in verbal and nonverbal symbol learning. *Child Development, 74*, 549-563.

Carey, S. (1978). The child as word-learner. In M. Halle, J. Bresnan, & G.A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264-293). Cambridge, MA: MIT Press.

Carey, S. (1994). Does learning a language require the child to reconceptualize the world? *Lingua, 92*, 143-167.

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

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

Casasola, M. (2005). When less is more: How infants learn to form an abstract categorical representation of support. *Child Development, 76*, 279-290.

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive simple causality. *Developmental Psychology, 29*, 421-433.

Caza, G. A., & Knott, A. (2012). Pragmatic bootstrapping: A Neural network model of vocabulary acquisition. *Language Learning and Development, 8*, 113-135.

Childers, J. B. (2011). Attention to multiple events helps two-and-a-half-year-olds extend new verbs. *First Language, 31*, 3-22.

Childers, J. B., & Paik, J. H., (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language, 36*, 201-224.

Christiansen, C. M., Louise Kelly, M., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition, 116*, 382-393.

Cohen, J. (1992a). A power primer. *Psychological Bulletin, 112*, 155-159.

Cohen, J. (1992b). Statistical power analysis. *Current Directions in Psychological Science, 1*, 98-101.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112*, 347-382.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 1119-1130.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 24-39.

Conway, C. M., Bauernschmidt, A., Huand, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition, 114*, 356-371.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*, 264-274.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua, 92*, 81-104.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162-170

Diesendruck, G. (2007). Mechanisms of word learning. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 257-276). Malden, MA: Blackwell Publishing.

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: an alternative account of Samuelson and Smith. *Developmental Science, 7*, 33-41.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General, 120*, 190-202.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*, 174-187.

Evans, J. L., Saffran, J. R., & Tore-Robes, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321-335.

Evey, J. A., & Merriman, W. E. (1998). The prevalence and the weakness of an early name mapping preference. *Journal of Child Language, 25*, 121-147.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science, 34*, 1-47.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), v-173.

Ferman, S., & Karni, A. (2010). No childhood advantage in the acquisition of skill using an artificial language rule. *PLoS ONE, 5*, 1-10.

Fernald, A., & Hurtado, N., (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science, 9*, F33-F40.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*, 499-504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: General*, 28, 458-467.

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science, 35*, 367-380.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development, 82*, 1738-1750.

Frank, M. C., Goodman, N. H., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*, 578-585.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (in press). Social discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development.*

Freyd, P., & Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior, 21*, 282-295.

Frishkoff, G. A., Collins-Thompson, K., Perfetti, C. A., & Callan, J. (2008). Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavior Research Methods, 40*, 907-925.

Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading, 15*, 71-91.

Gampe, A., Liebal, K., & Tomasello, M. (2012). Eighteen-month-olds learn novel words through overhearing. *First Language, 32*, 385-397.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200-213

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A Longitudinal study. *Developmental Psychology, 28*, 887-898.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development (Vol. II): Language, thought, and culture* (pp. 31-53). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman, & S.C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge, UK: Cambridge University Press.

Gentner, D. & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487-513.

Gillette, J., Gleitman, H., Gleitman, L. R., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135-176.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3-55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. Language, Learning and Development, 1, 23-64.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*, 95-112.

Gliga, T., Elsabbagh, M., Hurdy, K., Charman, T., Johnson, M. H., & The BASIS Team (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development, 83*, 926-938.

Goldberg, A. (1999). The emergence account of argument structure semantics. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 197-212). Mahwah, NJ: Lawrence Erlbaum Associates.

Golinkoff, R. M., & Hirsh-Pasek, K. (2000). Word learning: Icon, index, or symbol? In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 3-18). New York, NY: Oxford University Press.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science, 15*, 30-33

Golinkoff, R. M., Hirsh-Pasek, K., Bloom, L., Smith, L. B., Woodward, A. L., Akhtar, N., Tomasello, M., & Hollich, G. (2000). *Becoming a word learner: A debate on Lexical Acquisition*. New York, NY: Oxford University Press.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21*, 125-155.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178-186.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*, 515-531.

Gordon, J., Schumm, J. S., Coffland, C., & Doucette, M. (1992). Effects of inconsiderate versus considerate text on elementary students' vocabulary learning. *Reading Psychology, 13*, 157-169.

Grunow, H., Spaulding, T. J., Gomez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders, 39*, 158-170.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology, 56*, 1213-1236.

Hall, D. G., Lee, S. C., & Belanger, J. (2001). Young children's use of syntactic cues to learn proper names and count nouns. *Developmental Psychology, 37*, 298-307.

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action-based verb extension. *Child Development, 82*, 674-686.

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

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58,* 78-79.

Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology, 108,* 229-241.

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.

Hintzman, D. L., & Stern, L. D. (1978). Contextual variability and memory for frequency. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 539-549.

Hochmann, J. R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition, 115*, 444-457.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*, 1368-1378.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. Child Development, 73, 418-433.

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: contextual repetition promotes learning from storybooks. *Frontiers in Psychology, 2*: 17.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*, 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*, 706-713.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151-195.

Howard, J. H. Jr., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning but not on implicit spatial context learning. *Neuropsychologia, 44*, 1131-1144.

Hudson Kam, C. L., & Newport, E. L. (2009) Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*, 30-66.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236-248.

Ichinco, D., Frank, M. C., & Saxe, R. (2009). Cross-situational word learning respects mutual exclusivity. In Taatgen, N., & van Rijn, H. (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2214-2219). Austin, TX: Cognitive Science Society.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: age-related difference in implicit sequence learning across the human lifespan. *Developmental Science, 15*, 496-505.

Jaswal, V. K., & Hansen, M. B. (2006). Learning words: children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science, 9*, 158-165.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21*, 767-787.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009). Frequency and contextual diversity effects in cross-situational word learning. In Taatgen, N., & van Rijn, H. (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2220-2225). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2010). Cross-situational statistical learning: Implicit or intentional? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2362-2367). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012a). An associative model of adaptive inference for learning word-referent mappings. *Psychonomic Bulletin & Review, 19*, 317-324.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012b). Actively learning nouns across ambiguous situations. In N. Miyake, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Kako, E. (2005). Information sources for noun learning. *Cognitive Science, 29*, 223-260.

Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development, 45*, 469-473.

Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18-month-old and 24-month-old infants' use of function words to determine reference. *Child Development, 77*, 325-338.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307-321.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science, 14*, 925-934.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*, 171-184.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35-B42.

Klein, K. A., & Yu, C. (2009). Joint or conditional probability in statistical word learning: Why decide? In Taatgen, N., & van Rijn, H. (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 3105-3110) Austin, TX: Cognitive Science Society.

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*, 295-323.

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

Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching, 42*, 341-364.

Laumann Long, L., & Shaw, R. J. (2000). Adult age differences in vocabulary acquisition. *Educational Gerontology, 26*, 651-664.

Lee. J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*, 69-92.

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect.* New York, NY: Mouton de Gruyter.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science, 31*, 581-612.

Lieven, E. V. M. (1994). Cross-linguistic and cross-cultural aspects of language addressed to children. In C. Gallaway, & B.J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56-73). Cambridge, UK: Cambridge University Press.

Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review, 79*, 1-13.

MacWhinney, B. (2000). *The CHILDES project*. Mahwah, NJ: Lawrence Erlbaum.

Magic Touch. [Add-On Touch Screen]. (2009). Garland, TX: Keytec, Inc.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science, 11*, 628-634.

Marazita, J. M., & Merriman, W. E. (2004). Young children's judgment of whether they know names for objects: The metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language, 51*, 458-472.

Markman, E. M. (1987). How children constrain the possible meanings of words. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 255-287). Cambridge, UK: Cambridge University Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14*, 57-77.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M.R. Gunnar, & M. Maratsos (Eds.) *Modularity and constraints in language and cognition: The Minnesota symposia on child psychology* (pp. 59-101). Hillsdale, NJ: Lawrence Erlbaum Associates.

Markman, E. M., & Hutchinson, J. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16*, 1-27.

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

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

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

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

McBride-Chang, C., Wagner, R. K., Muse, A., Chow, B., W-Y., & Shu, H. (2005). The role of morphological awareness in children's vocabulary acquisition. *Applied Psycholinguistics, 26*, 415-435.

McBride-Chang, C., Tardif, T., Cho, J-R., Shu, H., Fletcher, Stokes, S. F., et al., (2008). What's in a word? Morphological awareness and vocabulary knowledge in three languages. *Applied Psycholinguistics, 29*, 437-462.

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

McMurray, B., Horst, J. S., Toscano, J. C., & Samuelson, L. K. (2009). Integrating connectionist learning and dynamical systems processing: case studies in speech and lexical development. In J. P. Spencer, M. S. C. Thomas & J. L. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 218-249). New York, NY: Oxford University Press.

Merriman, W. E. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 331-358). Mahwah, NJ: Lawrence Erlbaum Associates.

Merriman, W. E., & Tomasello, M. (1995). Verbs are words too. In M. Tomasello, W. E., Merriman (Eds.), *Beyond Names for Things: Young Children's Acquisition of Verbs* (pp. 1-18). Hillsdale, NJ: Lawrence Erlbaum Associates.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3-19.

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology, 69*, 199-221.

Meyer, M., & Baldwin, D. (2011). Statistical learning of action: The role of conditional probability. *Learning & Behavior, 39*, 383-398.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An Individual differences study. *Language Learning, 62*, 302-331.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1*:31.

Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition, 123,* 133-143.

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior & Development, 21*, 373-377.

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology, 21*, 283-298.

Mundy, P., Fox, N., & Card, J. (2003). EEG coherence, joint attention and language development in the second year. *Developmental Science, 6*, 48-54.

Murphy, G. L. (2002). *The Big Book of Concepts.* Cambridge, MA: The MIT Press.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24,* 237-270.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E., Curtis (Eds.), *The Nature of Vocabulary Acquisition* (pp. 19-36). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233-253.

Namy, L. L. (2009). Early word learning and other seemingly symbolic behaviors. In A. Woodward & A. Needham (Eds.), *Learning and the Infant Mind* (pp. 249-262). New York, NY: Oxford University Press.

Namy, L. L. (2012). Getting specific: Early general mechanisms give rise to domain-specific expertise in word learning. *Language Learning and Development, 8*, 47-60.

Namy, L. L., Gentner, D., & Clepper, L. E. (2007). How close is too close? Alignment and perceptual similarity in children's categorization. *Cognition, Brain, & Behavior, 11*, 647-659.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131*, 5-15.

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development, 69*, 295-308.

Namy, L. L. & Waxman, S. R. (2000). Naming and exclaiming: Infants' sensitivity to naming contexts. *Journal of Cognition and Development, 1*, 405-428.

Namy, L. L., Knight-Schwartz, J. M., & Smith, L. B., (2011). *Learning the cues that signal naming events.* Manuscript submitted for publication.

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science, 6*, 136-142.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Nicol Medina, T., Snedeker, J., Trueswell, J. C., & Gleitman., L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences, 108*, 9014-9019.

Ochs, E., & Shieffelin, B. B., (1984). Language acquisition and socialization: Three developmental stories and their implications. In R. A. Shweder & R. A. LeVine (Eds.) *Culture theory: Essays on mind, self, and emotion* (pp. 276-320). Cambridge, UK: Cambridge University Press.

Ochs, E., & Shieffelin, B. B. (2008). Language socialization: An historical overview. In P. A. Duff & N. H. Hornberger (Eds.), *Encyclopedia of language education: Language socialization* (pp. 3-15). New York, NY: Springer.

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*, 763-782.

Pappafragou, A., Cassidy, K., & Gleitman, L. R. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition, 105*, 125-165.

Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg., H. (2007). Children with autism illuminate the role of social information in word learning. *Child Development, 78,* 1265-1287.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1281-1292.

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

Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development, 3*, 295-323.

Pinker, S. (1989). *Learnability and cognition: The acquisition of grammatical structure.* Cambridge, MA: MIT Press.

Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders, 35*, 453-462.

Plunkett, K., Sinha, C., Moller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science, 4*, 293-312.

Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning and Verbal Behavior, 22*, 133-152.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development, 77*, 266-280.

Pye, C. (1986). Quiche Mayan speech to children. *Journal of Child Language, 13*, 85-100.

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

Quine, W. V. O. (1968). Ontological relativity. *The Journal of Philosophy, 65,* 185-212.

Quine, W. V. O. (1987). Indeterminacy of translation again. *The Journal of Philosophy, 84*, 5-10.

REALbasic (Version 5.1) [Computer Software]. (2008). Austin, TX: REAL Software Inc.

Regier, T. (2003). Emergent constraints on word-learning: A computational review. *Trends in Cognitive Sciences, 7*, 263, 268.

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

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research, 52*, 16-30.

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartnets helps them learn new vocabulary words. *Journal of Educational Psychology, 86*, 54-64.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science, 1*, 906-914.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*, 339-349.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy, 15*, 608-635.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies of Second Language Acqusition, 21*, 589-619.

Rowe, M. L., Levine, S. C., Fisher, J. A., & Goldin-Meadow, S. (2009). Does linguistic input play the same role in language learning in children with and without early brain injury? *Developmental Psychology, 45*, 90-102.

Sabbagh, M. A., & Baldwin, D. A. (2005). Understanding the role of communicative intentions in word learning. In N. Eilan, C. Hoerl, T. McCromack, & J., Roessler (Eds.), *Joint Attention: Communication and Other Minds* (pp. 165-184). New York, NY: Oxford University Press.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493-515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 177-196.

Saffran, J. R. (2009). What can statistical learning tell us about infant learning? In A. L. Woodward, & A. Needham (Eds.), *Learning and the infant mind* (pp. 29-46). New York, NY: Oxford University Press.

Saffran, J. R. (2010). What is statistical learning, and what statistical learning is not. In S. P. Johnson (Ed.), *Neoconstructivism: The New Science of Cognitive Development* (pp. 180-194). New York, NY: Oxford University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27-52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101-105.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*, 484-494.

Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology, 38*, 1016-1037.

Samuelson, L. K., & Bloom, P. (2008). The shape of controversy: What counts as an explanation of development? *Developmental Science, 11*, 183-184.

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94-104.

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

Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997). Parital word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research, 29*, 531-553.

Scott, R. M., & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition, 122*, 163-180.

Shore, W. J., & Durso, F. T. (1990). Partial knowledge in vocabulary acquisition: General constraints and specific detail. *Journal of Educational Psychology, 82*, 315-318.

Shu, H., Anderson, R. C., & Zhang, H. (1995). Incidental learning of word meanings while reading: A Chinese and American cross-cultural study. *Reading Research Quarterly, 30*, 76-95.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition, 106*, 833-870.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39-91.

Smith, C. L., & Tager-Flusberg, H. (1982). Metalinguistic awareness and language development. *Journal of Experimental Child Psychology, 34*, 449-468.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2009). Reconsidering human cross-situational learning: A Revision to Yu & Smith's (2007) experimental paradigm. In Taatgen, N., & van Rijn, H. (Eds.), *Proceedings of the 31st Annual*

Conference of the Cognitive Science Society (pp. 2711-2716) Austin, TX: Cognitive Science Society.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2010). Cross-situational learning: An Experimental study of word-learning mechanisms. *Cognitive Science, 35*, 480-498.

Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51-80). New York, NY: Oxford University Press.

Smith, L. B. (2001). How domain-general processes may create domain-specific biases. In M. Bowerman, & S.C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 101-131). Cambridge, UK: Cambridge University Press.

Smith, L. B. (2003). Learning to recognize objects. *Psychological Science, 14*, 244-250.

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science, 34*, 1287-1314.

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

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

Smith, L. B., & Yu, C. (in press). Attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development.*

Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*, 342-353.

Smith-Hefner, N. J. (1988). The linguistic socialization of Javanese children in two communities. *Anthropological Linguistics, 30*, 166-198.

Snedeker, J. (2009). Word learning. In L.R. Squire (Ed.) *Encyclopedia of neuroscience* (pp. 503-508). Amsterdam, Netherlands: Elsevier.

Snedeker, J., & Gleitman, L. R. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon* (pp. 257-293). Cambridge, MA: MIT Press.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language, 4*, 1-22.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition, 20*, 51-68.

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

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research, 69*, 261-286.

Swanborn, M. S. L., & de Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning, 52*, 95-117.

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

Thiessen, E. D. (2009). Statistical learning. In E.L. Bavin (Ed.) *The Cambridge handbook of child language* (pp. 35-50). Cambridge, UK: Cambridge University Press.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of sress and statistical cues to word boundaries by 7- to 9-month-old. *Developmental Psychology, 39*, 706-716.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*, 1-42.

Tomasello, M. (2003). *Constructing a language: A Usage-based theory of language acquisition.* Cambride, MA: Harvard University Press.

Tomasello, M., & Akhtar, N. (2000). Five questions for any theory of word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 179-186). New York, NY: Oxford University Press.

Tomasello, M., & Barton, M. (1994). Learning words in nonostensive contexts. *Developmental Psychology, 30*, 639-650.

Tomasello, M., & Krueger, A. C. (1992). Acquiring verbs in ostensive and nonostensive contexts. *Journal of Child Language, 19*, 311-333.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language, Learning and Development, 3*, 269-293.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245-1260.

Turk-Browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*, 552-564.

Verkoeijen, P. P. J. L., Pikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 796-800.

Vlach, H. A., & Sandhofer, C. M. (2010). Desirable difficulties in cross-situational word learning In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Cognitive Science Society* (pp. 737-742). Austin, TX: Cognitive Science Society. (pp. 2470-2475).

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context dependent word learning. *Journal of Experimental Child Psychology, 108,* 394-401.

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's learned words. *Frontiers in Psychology, 3*: 46.

Vogt, P. (2012). Exploring the robustness of cross-situational learning under zipfian distributions. *Cognitive Science, 36*, 726-739.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition, 107,* 729-742.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology, 45*, 1611-1617.

Wagner, R. K., Muse, A. E., & Tannenbaum (2007). *Vocabulary Acquisition: Implications for Reading Comprehension.* New York, NY: The Guilford Press.

Wagovich, S. A., & Newhoff, M. (2004). The single exposure: Partial word knowledge growth through reading. *Journal of Speech–Language Pathology, 13*, 316-328.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36,* 571-581.

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*, 265-279.

Werker, J. F., Fennell, C. T., Corcoran, K., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*, 1-30.

Werner, H., & Kaplan, E., (1950). Development of word meaning through verbal context: An experimental study. *The Journal of Psychology, 29*, 251-257.

Whitehurst, G. J., Kedesdy, J., & White, T. (1982). A functional analysis of meaning. In S. A. Kuczaj II (Ed.), *Language development (Vol. I): Syntax and semantics* (pp. 397-427). Hillsdale, NJ: Lawrence Erlbaum Associates.

Whitmore, J. M., Shore, W. J., & Smith, P. H. (2004). Partial knowledge of word meanings: Thematic and taxanomic representations. *Journal of Psycholinguistic Research, 33*, 137-164.

Wilkinson, K. S., & Houston-Price, C. (in press). Once upon a time, there was a pulchritudinous princess…: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics*.

Woodward, A. L. (2000). Constraining the problem space in early word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 81-114). New York, NY: Oxford University Press.

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70*, 65-77.

Woodward, A. L., & Markman, E. M., (1998). Early word learning. In W. Damon, D. Kuhn, & R. S. Siegler (Eds.), *Handbook of child psychology (Vol. 2): Cognition, perception, and language* (371-420). New York, NY*:* John Wiley & Sons.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology, 30*, 553-566.

Wysoci, K. & Jenkins, J. R. (1987). Deriving word meanings through morphological generalization. *Reading Research Quarterly, 22*, 66-81.

Xu, F., & Tennenbaum, J. B. (2007a). Word learning as Bayesian inference. *Psychological Review, 114*, 245-272.

Xu, F., & Tenenbaum, J. B. (2007b). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*, 288-297.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12,* 1-7.

Yoshida, K. A., Rhemtulla, M., & Vouloumanos, A. (2012). Exclusion constraints facilitate statistical word learning. *Cognitive Science, 36,* 933-947.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using head camera to study visual experience. *Infancy, 13*, 229-248.

Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 545-559.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development, 4*, 32-62.

Yu, C. & Ballard, D. H., (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*, 2149-2165.

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

Yu, C. & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science, 14*, 165-180.

Yu. C, Smith, L. B., Klein, K., & Shiffrin, R. M. (2007). Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same? In D. S. McNamara & J. G. Trafton (Eds.) *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 737-742). Austin, TX: Cognitive Science Society.

Yu, C., Zong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: evidence from eye tracking. *Frontiers in Psychology, 3*: 148.

Yuan, S., & Fisher, C. (2009). Really? She blicked the baby? Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science, 20,* 619-626.

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 715-720). Austin, TX: Cognitive Science Society.

Yurovsky, D., Fricker, D., Yu, C., & Smith, L. B. (2010). The active role of partial knowledge in cross-situational word learning. In S. Ohlsson, & R. Catrambone (Eds.) *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1589-1594). Austin, TX: Cognitive Science Society.

Zechmeister, E B., Chronis, A. M., Cull, W. L., D'Anna, & Healy, N. A. (1995). Growth of a functionally important lexicon. *Journal of Reading Behavior, 27*, 201-212.