FOUNDATIONS OF VOCABULARY: DOES STATISTICAL SEGMENTATION OF EVENTS CONTRIBUTE TO WORD LEARNING?

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Dani F. Levine
December 2017

Examining Committee Members:

Dr. Kathy Hirsh-Pasek, Advisory Chair, Developmental Psychology
Dr. Nora Newcombe, Examining Chair, Developmental Psychology
Dr. Thomas Shipley, Brain and Cognitive Sciences
Dr. Peter Marshall, Developmental Psychology
Dr. Tania Giovannetti, Brain and Cognitive Sciences
Dr. Roberta Michnick Golinkoff, External Member, University of Delaware

ABSTRACT

This dissertation evaluates the untested assumption that the individuation of events into units matters for word learning, particularly the learning of terms which map onto relational event units (Gentner & Boroditsky, 2001; Maguire et al., 2006). We predicted that 3-year-old children's statistical action segmentation abilities would relate to their verb comprehension and to their overall vocabulary knowledge (Research Question 1). We also hypothesized that statistical action segmentation would facilitate children's learning of novel verbs (Research Question 2).

Largely confirming our first prediction, children who were better able to statistically segment novel action sequences into reliable units had more sophisticated overall vocabularies and were quicker to select the correct referents of overall vocabulary items and verb vocabulary items; nevertheless, they did not have larger verb vocabularies. Unexpectedly, statistical action segmentation did not facilitate children's learning of verbs for statistically consistent action units. However, children showed greater learning of verbs labeling statistical action part-units than verbs labeling statistical action non-units, providing some evidence for our second prediction. In sum, this dissertation takes an important step towards understanding how event segmentation may contribute to vocabulary acquisition.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Kathy Hirsh-Pasek, for guiding me through this journey with unparalleled enthusiasm, honesty, and commitment. I would also like to thank Roberta Golinkoff, Nora Newcombe, Tim Shipley, Peter Marshall, and Tania Giovannetti for their insightful comments on this work.

I am grateful also for my exceptional community of colleagues, past and present, at the Temple Infant and Child Lab and the Child’s Play, Learning, and Development Lab at the University of Delaware. In particular, I would like to thank Tamara Spiewak Toub for being an amazing friend and supporting me emotionally as I worked through this dissertation, and for taking the time to provide detailed feedback on my presentation. I would also like to thank my lab twin, Junko Kanero, for acting as a sounding board for my ideas throughout these years, helping me troubleshoot E-Prime and eye tracker issues, and for being a selfless person who would drop everything when I needed support. I thank Tara Saunders, Jelani Medford, Ilissa Beasley, Laura Ayres, Marissa Lazar, and Mo Kasim for their instrumental assistance with data collection and analysis. Finally, I would like to thank Sarah Paterson, Brianna McMillan, Emily Hopkins, Molly Scott, and Lillian Masek for their thoughtful feedback and support.

I owe more than I can say to my parents, Steve and Jodi, for their unconditional love and support, which made it possible for me to work towards this achievement and to keep pushing forward through the challenges I take on. Thank you to my sister Heather Levine-Levy and brother-in-law Larry Levy, for being my closest friends and grounding me with silliness and happiness. Thank you to my new and wonderful family, the Donesons, who welcomed me with open arms and who supported me through this process. Most of all, thank you to my husband and best friend Steve Doneson, whose love and devotion not only helped me through this challenging time, but kept me happy and smiling every single day.

Finally, thank you to all of the children and parents who selflessly gave their time to make this research possible.

TABLE OF CONTENTS

CHAPTER

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Children must learn the shared meanings of a language to communicate and to succeed academically (Catts, Bridges, Little, & Tomblin, 2008; Young et al., 2002) and socially (McCabe & Meller, 2004; Snowling et al., 2006). Yet learning language is complex, forcing children to transform their continuous experiences into meaningful semantic units (Fisher & Gleitman, 2002; Göksun, Hirsh-Pasek, & Golinkoff, 2010). Mapping word to world requires breaking the continuous flow of sounds into discrete units such as words and requires parsing the flow of dynamic events into units like objects and actions.

Much of the research that has been conducted on word learning presumes that these speech and event units exist in the child's mind – only to be linked as the child learns words' referents (e.g., Arunachalam & Waxman, 2011; Brandone, Pence, Golinkoff, & Hirsh-Pasek, 2007; Gertner, Fisher, & Eisengart, 2006; Maguire et al., 2010; Scott & Fisher, 2012; Waxman, Lidz, Braun, & Lavin, 2009). However, work in areas such as speech perception and event perception demonstrates that the segmentation of speech and events into units presents additional challenges that may impact how children break into language (for a review, see Levine, Strother-Garcia, Hirsh-Pasek, & Golinkoff, in press). Further, work in speech perception shows that the process of breaking the flow of speech into language units plays a role in children's word learning (for reviews, see Romberg & Saffran, 2010; Saffran, 2014).

In this dissertation, we take the lead from speech perception research and ask whether children's parsing of events into units may contribute to their word learning. In particular, we take what Gleitman, Cassidy, Nappa, Papafragou, and Trueswell (2005) called "hard words" or verbs as one of our primary outcomes and investigate two research questions.

First, this work evaluates whether 3-year-old children's competence in parsing novel event streams into units is related to their knowledge of verbs and their overall vocabularies. Second, this research examines whether action segmentation may facilitate 3-year-olds' learning of novel verbs that map onto event units. This dissertation provides an evaluation of the untested assumption that event segmentation supports lexical acquisition.

Challenges of Word Learning

Patterns of lexical acquisition reveal challenges inherent in word learning, hinting at the importance of event segmentation. In particular, children's knowledge of nouns exceeds their knowledge of other types of words in comprehension and production (Gentner, 1982; Goldin-Meadow, Seligman, & Gelman, 1976; Haman et al., 2017). The noun bias is a cross-linguistic phenomenon, holding true not only for 'noun-friendly' languages such as English and French but also for 'verb-friendly' languages such as Mandarin and Korean, which use verbs in more prominent sentence locations and frequently drop noun phrases in conversation (Bornstein et al., 2004; Imai et al., 2008; Waxman et al., 2013).

The universality of this pattern suggests that the learnability of words may be derived not just from variability in sound patterns and syntax but from variability in words' referents. Mapping from word to world is not straightforward in any situation as famously noted by Quine in 1960; but for some types of words, linking sounds to referents is particularly problematic. Gentner and Boroditsky (2001) proposed the "Division of Dominance" continuum, hypothesizing that words vary in terms of their cognitive and linguistic dominance. At one extreme of this continuum are concrete nouns (e.g., *cup*, *ball*), whose referents are easily individuated independent of language (cognitive dominance); at the other extreme are determiners and conjunctions (e.g., *the, but*), whose meanings depend entirely on language (linguistic dominance); verbs and spatial prepositions fall in the middle of the continuum, with percepts that are less cohesive than objects such that they can be individuated in a variety of ways, relying on language to highlight the relevant units (Gentner & Boroditsky, 2001). Further, Maguire, Hirsh-Pasek, and Golinkoff (2006) suggested that words lie on a continuum of learnability based on four factors: the consistency of the *shape* of the word's referent, the ease with which the word's referent can be *individuated*, the *concreteness* of the word's referent to sensory systems, and the word's *imageability*, or the ease with which the word can evoke a mental image. Across both of these proposed continua, individuation of word referents is hypothesized to be a critical factor for word learning.

Even among nouns, some referents are more easily individuated than others. For example, children have more difficulty individuating small mobile objects than self-moving animate beings, and have even more difficulty individuating stationary amorphous objects (Gentner & Boroditsky, 2001). However, verbs provide a particularly complex individuation challenge for children.

Verbs express relations between participants and objects in events (e.g., the boy is *entering* the playground and *sprinting* to the basketball court as he *bounces* his basketball). Mapping verbs onto action units seems to demand that children recognize where one unit ends and another begins – all while continuous events fleetingly pass by and are replaced by other events. In this way, the processes of individuating objects versus individuating actions are distinct. For example, Maguire, Brumberg, Ennis, & Shipley (2011) demonstrated that while the boundaries of objects are perceived in a highly consistent way, the boundaries of formally equivalent action paths are perceived more variably. The spatiotemporal nature of actions makes them more difficult to individuate than objects, creating an additional challenge for verb learning.

How, then, are verb referents individuated? Consider the action verbs *push* and *throw*. There are particular concrete, reliable movement patterns in intentional, goal-directed actions such as *pushing* and *throwing* that can be observed repeatedly. Visual experience of these consistent movement regularities abounds, and, even without explicit awareness of the intent or goal structure in the movements, these recurrent experiences may lead viewers to individuate coherent action units (Baldwin, Baird, Saylor, & Clark, 2001; Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015; Newtson, Engquist, & Bois, 1977; Reynolds, Zacks, & Braver, 2007; Zacks, 2004; Zacks, Kumar, Abrams, & Mehta, 2009) that can serve as the referents of verbs. Indeed, evidence from cross-linguistic research suggests that languages are limited in how they can structure relational referents, constrained by ‘salient physical distinctions’ (Malt, Ameel, Imai, Gennari, Saji, & Majid, 2014; see also Majid, Jordan, & Dunn, 2015).

Physical action or manner verbs like *push* and *throw* are acquired earlier than verbs denoting intentions, such as *pour* and *spill* and mental state verbs such as *think* and *wish* (Fenson et al., 1994; Maguire et al., 2006; Poulin-Dubois & Forbes, 2002). Individuating events that support the learning of words with a strong foundation in perceptual experience may act as an important stepping stone for the learning of less concrete relational terms. That is, learning words for individuated actions within events might serve as a bootstrap for word to world mapping involving actions and events that must be inferred.

Statistical Action Segmentation

One process that would enable children to parse novel perceptual events into consistent units for language is statistical action segmentation, a form of statistical learning. Statistical learning is the mechanism by which predictably-structured patterns are extracted from continuous streams of information in the environment. This mechanism allows for rapid learning of reliable associations “by mere exposure” (Aslin, 2017). While research has recently begun to explore this process in the area of action segmentation, an extensive body of literature spanning over 20 years examines statistical learning in two related areas: speech segmentation and cross-situational word learning. From birth, infants track transitional regularities between syllables in the auditory speech stream, producing rudimentary but coherent “word” units (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Teinonen, Fellman, Naatanen, Alku, &

Huotilainen, 2009), and these abilities have been similarly demonstrated in school-age children (François, Chobert, Besson, & Schön, 2013; Mayo & Eigsti, 2012; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) and adults (Cunillera et al., 2009; Saffran et al., 1997). In addition, as early as 12 months and through school ages and adulthood, children track associative regularities or co-occurrence statistics between speech sounds and visual referents across learning situations, creating inchoate representations of word to world pairings (Scott & Fisher, 2012; Smith, Smith, & Blythe, 2011; Smith & Yu, 2008, 2013; Suanda, Mugwanya, & Namy, 2014; Vlach & Johnson, 2013; Vouloumanos & Werker, 2009; Yu & Smith, 2007). Statistical learning in these domains is constrained by attention (Fernandes, Kolinsky, & Ventura, 2010; Smith & Yu, 2013; Toro, Sinnett, & Soto-Faraco, 2005), memory (Palmer & Mattys, 2016; Vlach & DeBrock, 2017; Vlach & Johnson, 2013), social cues (Kuhl, 2007; MacDonald, Yurovsky, & Frank, 2017), and features of the stimulus that make it easier or more challenging to parse, such as the prosody of child-directed speech (Johnson & Jusczyk, 2001; Shukla, Nespor, & Mehler, 2007; Thiessen, Hill, & Saffran, 2005; see also Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989) and the sentence-final position of important words preceded by determiners that is also a hallmark of child-directed speech (Monaghan & Mattock, 2012; Yurovsky, Yu, & Smith, 2012).

Much less is known about how regularities in spatiotemporal streams of action may be tracked. Most research examining visual statistical learning in young children focuses on their ability to track simple ordering patterns of looming shapes (e.g., Fiser & Aslin, 2002); this ability is available from birth (Bulf, Johnson, & Valenza, 2011; Kirkham, Slemmer, & Johnson, 2002) and undergoes significant developments over time

(Kirkham, Slemmer, Richardson, & Johnson, 2007; Marcovitch & Lewkowicz, 2009; Slone & Johnson, 2015). However, several studies to date go beyond these overly simplistic tests of visual statistical learning to evaluate segmentation of novel action sequences.

The first of these studies familiarized 7- to 9-month-old infants with a 4-minute video in which 12 hand motions (i.e., *fine* action units) were grouped into triads (i.e., *coarse* action units) (Roseberry et al., 2011). The hand motions in each triad always appeared in the same order, but the ordering of the triads was variable. The authors asked whether infants who experienced this visual sequence would be able to distinguish the statistically intact triad units from triad part-units, which crossed a triad unit boundary and therefore had lower transitional probabilities (similar to infants' discrimination between statistical speech units and part-units; Aslin, et al., 1998). Indeed, infants succeeded at this task, looking longer toward intact units versus part-units, suggesting they had extracted coarse action units on the basis of transitional regularities between finer movements (Roseberry et al., 2011). In a follow-up study, Stahl and colleagues (2014) used a familiarization video of statistically-defined whole-body actions of an animated starfish in which the boundaries of fine action units were identical (i.e., the starfish always returned to a neutral body position), to ensure that infants were sensitive to transitional regularities rather than the transitional movements between actions. Infants still distinguished units from part-units, although in this study infants demonstrated the opposite pattern of greater looking toward part-units (Stahl et al., 2014).

In addition to those two studies examining infants' segmentation of novel body actions, a set of three studies explored children's sensitivity to a statistical action sequence involving the manipulation of six unique objects. The sequence contained two deterministic action pairs (in which the manipulation of one object always preceded the manipulation of a second object) – one pair leading to an effect (i.e., a light turning on) and the other having no effect – with other unpaired actions interspersed randomly (Monroy, Gerson, Domínguez-Martínez, Kaduk, Hunnius, & Reid, in press; Monroy, Gerson, & Hunnius, 2017a; Monroy, Meyer, Gerson, & Hunnius, 2017b). In two of those studies, 8- to 11-month-old infants and 18-month-old toddlers only learned to visually anticipate the second action of the deterministic pair leading to an effect (Monroy et al., in press; Monroy et al., 2017b), consistent with research showing that causal or enabling relations facilitate children's event memory (Bauer, 1992; Bauer & Mandler, 1989; Wenner & Bauer, 1999); in the third study, 19-month-old toddlers learned to predict both deterministic action pairs (Monroy et al., 2017a). Interestingly, children only succeeded at learning these regularities when an actor's hand performed the action sequence, but not when an identical self-propelled sequence was performed in a "ghost" condition, in which the hand was replaced by a spotlight (Monroy et al., 2017a). These findings add to the evidence that children segment novel event sequences using statistical learning, and additionally suggest that this process may be limited to agentive action.

These studies demonstrate that statistical action segmentation is performed by infants and toddlers, and research with adults reveals a comparable capacity for segmenting novel events based on transitional regularities (Baldwin, Andersson, Saffran, & Meyer, 2008; Buchsbaum et al., 2015). However, to date no research has addressed the potential link between children's action segmentation and their lexical acquisition.

Bridging the gap between event segmentation and language learning will help to elucidate both of these complex cognitive processes. This dissertation addresses the following two research questions. The first explores the association between individual differences in action segmentation ability and vocabulary knowledge. In particular, we explored whether 3-year-old children who performed better on a statistical action segmentation task had greater verb vocabularies and overall vocabularies than children who performed worse on the segmentation task. The second question evaluates a potential causal relation, assessing experimentally whether action segmentation facilitates verb learning. Specifically, 3-year-olds first viewed a statistical action sequence and, following this segmentation opportunity, were tested on an action unit labeling task to determine whether subsequent learning of novel verbs for units, part-units, and non-units were differentially affected. The sections below present the specific aims for each research question and how each question addresses those objectives.

Research Question 1: Is Action Segmentation Linked with Vocabulary Knowledge?

To make the case that action segmentation contributes to lexical development, it is critical to establish that there is an association between children's action segmentation abilities and their vocabulary knowledge. There is reason to believe that segmentation of events will be related to language learning, with hints come from research linking segmentation of the sound stream to word acquisition (Evans, Saffran, & Robe-Torres, 2009; Junge, Kooijman, Hagoort, & Cutler, 2012; Kooijman, Junge, Johnson, Hagoort, & Cutler, 2013; Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006; Newman, Rowe, & Ratner, 2016; Singh, Reznick, & Xuehua, 2012; Swingley, 2005). That is – to the extent that children must find individual sound patterns that link to individual objects, actions, and events – then determining not only what counts acoustically as a word but also what counts visually as a referent will be key to the word learning process.

Prior research has examined relations between different aspects of non-verbal action processing and language outcomes, with compelling results. First, a study conducted by Kaduk and colleagues (2016) explored the possible relations between 9-month-olds' action processing and their language abilities at 9 and 18 months. The researchers showed infants videos of actions with expected outcomes (e.g., a man bringing a soft pretzel to his mouth) and actions with unexpected outcomes (e.g., the man bringing the pretzel to his ear) while measuring event-related potentials (ERPs) in response to these two event types. Prior research had shown that unexpected action outcomes elicited an N400-like ERP component suggestive of semantic action processing (Reid, Hoehl, Grigutsch, Groendahl, Parise, & Striano, 2009); the researchers thus hypothesized that individual differences in the ability to detect this violation in action structure would relate to individual differences in vocabulary size (Kaduk et al., 2016). Indeed, the 9-month-olds who showed an N400 response to unexpected action outcomes were those with higher-than-average vocabulary scores, assessed using the Swedish adaptation of the MacArthur Communicative Development Inventory, at both 9 and 18 months (Kaduk et al., 2016). Understanding the structure of familiar events may facilitate the acquisition of words that label objects and actions within those familiar events, leading to increased vocabulary size.

Another aspect of action processing that may relate to vocabulary is children's ability to form action categories of path (i.e., trajectory of motion) and manner (i.e., how an action is performed) (Konishi, Stahl, Golinkoff, & Hirsh-Pasek, 2016). Given that verbs label paths (e.g., *enter*, *ascend*) and manners (e.g., *jump*, *run*), Konishi and colleagues (2016) hypothesized that individual differences in the ability to extract invariant paths and manners when other event features vary would predict children's knowledge of verbs. Children's nonlinguistic action categorization skills were tested at 13-15 months, and verb knowledge was assessed at 27-33 months using a Verb Comprehension Task, which evaluated children's understanding of 36 action verbs (Konishi et al., 2016). Additionally, analyses controlled for overall receptive vocabulary at time 1, as measured by parent report using the short form of the MacArthur–Bates Communication Development Inventory. Findings supported the researchers' prediction: children who were better at forming path and manner categories at 13-15 months had larger verb vocabularies at 27-33 months, even after controlling for overall vocabulary (Konishi et al., 2016).

The robust relations found between individual differences in action processing and vocabulary in these two studies using diverse methods (Kaduk et al., 2016; Konishi et al., 2016) are intriguing, but leave many questions unanswered. First, while Kaduk and colleagues (2016) found that children's detection of action structure was related to individual differences in overall vocabulary, Konishi and colleagues (2016) found that children's action categorization abilities were specifically related to their knowledge of verbs, but not to their overall vocabulary when verbs were excluded. It may be that action categorization abilities are more directly linked to verb learning, while detecting action structure reflects action processing skill that relates to vocabulary more broadly.

However, there are many differences between these studies, and many alternative explanations for these disparate findings.

A second limitation of these studies is that they both utilize stimuli demonstrating *pre-individuated*, *familiar* actions. While this constraint enabled the researchers to test their hypothesized relations between specific aspects of action processing and vocabulary, it also obviated the need for children to segment events. In theory, the segmentation of novel, continuous events into action units may be a necessary prerequisite for forming path/manner categories and for understanding action structure, and should therefore also be related to verb and overall vocabulary knowledge.

By age three, children have acquired a large and rapidly growing number of verbs and other relational terms (Fenson et al., 1994); thus, this age is an optimal time to explore relations between action segmentation abilities and lexical knowledge. In this dissertation, we hypothesized that children who were better able to extract statistical action units from a novel stream of events would have greater verb and overall vocabularies than children who were poorer segmenters of novel events.

Research Question 2: Does Action Segmentation Facilitate Verb Learning?

Once we establish an association between event segmentation and vocabulary, it is imperative that we demonstrate that the mechanism of statistical action segmentation causally supports word learning. Again, we can turn to the speech perception literature as a guide for examining this question.

Research suggests that forming rudimentary speech units via speech segmentation supports children's word learning: giving 17-month-olds an opportunity to segment a novel language stream (i.e., artificial or foreign language) containing statistically consistent syllabic units (i.e., words) facilitated their ability to map statistically consistent syllabic labels, but not statistically inconsistent labels, onto objects (Graf Estes, Evans, Alibali, & Saffran, 2007; Hay, Pelucchi, Graf Estes, & Saffran, 2011). Even adults, who succeeded at learning all labels in a similar task, showed more rapid word learning for statistically consistent labels relative to inconsistent labels (Mirman, Magnuson, Graf Estes, & Dixon, 2008). Further, Graf Estes (2012) demonstrated this facilitation effect in 17-month-olds even when the voice at label learning was different than the voice from the speech segmentation sequence, suggesting that the segmented word units could be generalized beyond the perceptual, acoustic details of the initial stimulus. For event segmentation to facilitate verb learning, segmented event units would similarly need to be generalized beyond the initial percept to other exemplars of the unit (e.g., different actor, different location).

We hypothesize that event segmentation will facilitate verb learning in 3-year-olds. However, the exact ways in which this facilitation will manifest are difficult to predict given conflicting findings from prior research. While the infants in one action segmentation study demonstrated an attentional preference for action units over part-units (Roseberry et al., 2011), the infants in another, similarly structured study showed the exact opposite preference (Stahl et al., 2014). The aforementioned studies on word learning facilitation from speech segmentation also produced conflicting results: two studies found word learning facilitation for statistical words (syllables that always appeared together) but not part-words (syllables that crossed a ‘word’ boundary) (Graf

Estes et al., 2007; Hay et al., 2011), while a third, similarly structured study found comparable facilitation for statistical words and part-words (Graf Estes, 2012).

Those studies only compared units and part-units, although other speech segmentation studies with infants and some event segmentation studies with adults have utilized a third test item type, the non-unit. A non-unit is a novel ordering of syllables or actions from the statistical action sequence, with no internal consistency (i.e., transitional probabilities of 0.0). Speech segmentation studies reveal infants' ability to differentiate between statistical words and both part-words and non-words (Pelucchi, Hay, & Saffran, 2009; Saffran et al., 1996), although as with the infant event segmentation studies, the *direction* of infants' attentional preferences vary from study to study. In event segmentation studies with adults, participants consistently marked statistical action units as more familiar than action part-units following their segmentation of the action sequence (Baldwin et al., 2008; Buchsbaum et al., 2015). However, while one study's adults did not distinguish part-units from non-units (Baldwin et al., 2008), the other study's adults marked part-units as more familiar than non-units (Buchsbaum et al., 2015). Given the consistent potency of statistical sequence learning across studies despite the inconsistencies in participants' patterns of discrimination at test, this dissertation aims to assess whether statistical action segmentation facilitates verb learning, by comparing verb learning for statistical action units, part-units, and non-units following exposure to a novel statistical action sequence.

Hypotheses

If event segmentation supports children’s word learning then individual differences in verb vocabulary and overall vocabulary knowledge should be explained in part by individual differences in statistical action segmentation ability, even when covariates including age and nonverbal intelligence are controlled. Further, an experimental test of this link should reveal a facilitatory effect of statistical action segmentation on children’s learning of novel verbs.

CHAPTER 2

METHODOLOGY

This dissertation evaluates two questions regarding the relation between children’s event segmentation and lexical acquisition. Research Question 1 investigates whether statistical action segmentation abilities relate to vocabulary knowledge at a developmental window when children’s vocabularies, particularly for verbs and other relational terms, are growing rapidly. Specifically, we tested the relation between 3-year-olds’ performance on a statistical action segmentation task and their verb and overall vocabulary knowledge, controlling for age and general intelligence. Research Question 2 examines whether the formation of action representations through statistical action segmentation facilitates the learning of novel verbs which label newly segmented units.

These two research questions were examined using one group of 3-year-olds, as described below. All experimental protocols reviewed in the following sections were approved by the Temple University Institutional Review Board.

Participants

A total of 64 typically developing 3-year-olds (*Mean age* = 41.82 months; *SD* = 3.79; 41 females) participated in the present study. Children were from monolingual English-speaking, upper middle-income households in two suburbs of northeastern cities in the United States. The majority of the study sample was White Caucasian (87.5% were White, 7.8% were Black/African American, 1.6% were Asian, and 3.2% were multiracial). Based on the tasks they completed, a subgroup of participants was selected for each research question. The sample size was determined for each research question based on power analyses and study design, and data collection continued until the predetermined sample sizes for both questions were reached. If children's attention to any test trial was below 2 seconds (i.e., the minimum time required to distinguish two statistical action units) or if their attention to verb training was below 3.33 seconds (i.e., the minimum time required to learn the verb-action unit pairing), they were not included in the analysis for the corresponding research question (see Data Analysis section for more information). An additional 18 participants were excluded from all analyses because of experimenter error (4), technical failure (3), or fussiness (11).

Children were included in the analysis for Research Question 1 if they completed all four tasks required to address the question (Segmentation Task, Verb Comprehension Task, Quick Interactive Language Screener, and Zoo Locations task; see Stimuli for the detail of each task). The analysis of Research Question 2, on the other hand, included all children who completed the Action Unit Labeling Task (see Stimuli for the detail of the task). Of the larger sample, 54 were included in Research Question 1 analyses, and 40 were included in Research Question 2 analyses.

Stimuli

Segmentation Task

Statistical Action Sequences

Two novel statistical action sequences performed by a single male actor were created for the Statistical Sequence Phase of the Segmentation Task. For each sequence, the actor was recorded on video performing 12 dynamic human body actions (Appendix

A). All actions began and ended at the same body position and were equivalent in duration (1-s cycle). This allowed for recombination of the actions into any order with Adobe Premiere Pro CS6 software, and ensured that no cues to sequence structure were provided by transitional movements (Stahl et al., 2014). The two sequences' actions were distinct both in the movements they involved and in the actor's hand formation throughout the actions – for one sequence the actor's hands were closed, and for the other sequence his hands were open while performing the actions (Appendix A).

For each sequence, the 12 actions were grouped into four distinct three-action units, with within-unit transitional probabilities (TPs) of 1.0. Each sequence had a total of 42 units. Two units occurred twice as often as the other two units (14 times vs. 7 times), creating high- and low-frequency action units (Figure 1a). The arrangement also created *part-units* spanning the last action of one high-frequency unit and the first two actions of another high-frequency unit (Figure 1a,b). Part-units had a within-unit TP of 0.5 between the first and second action and 1.0 between the second and third action (Figure 1b). With this sequence structure, the frequency of each part-unit as a whole is equivalent to the frequency of each low-frequency unit (Aslin et al., 1998). The full duration of each 42-unit sequence (composed of 126 actions) was 126-seconds.

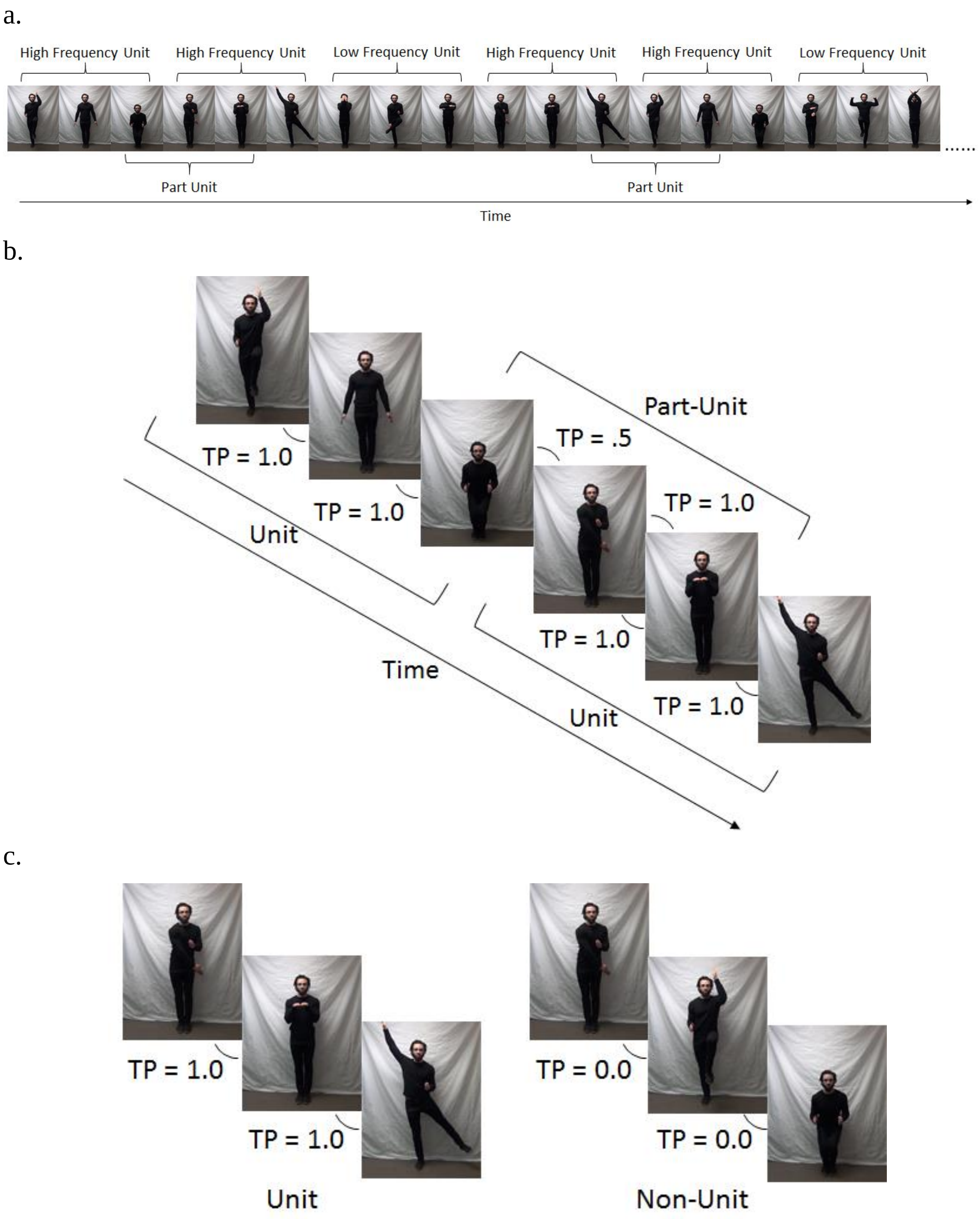


Figure 1. Actions from the Segmentation Task. Examples of (a) sequence of actions from one of the Statistical Sequence Phases, (b) units and part-units from the sequence, and (c) a pairing of a unit with a non-unit. Actions are depicted here as static pictures, though they were presented as dynamic events in the study.

Action Units

For the purpose of each Segmentation Test Phase (see Design section for more detail), videos of action segments from the sequence were created. These included the two high-frequency units, two low-frequency units, and two part-units. In addition, videos of six non-units were created for the six test trials. Non-units were novel arrangements of three actions from the sequence, with internal TPs of 0.0. Each non-unit was paired with one of the six units/part-units, and therefore the structure of non-units was determined by the unit/part-unit with which it was paired. The non-unit paired with each unit/part-unit began with the same initial action as the unit/part-unit, and all three actions within each non-unit were equivalent in frequency to the three actions in the unit/part-unit with which it was paired (see Figure 1c).

The actions were slowed down for test trials by 33% relative to their duration in the full sequence because piloting revealed that children required more time to process two action units (i.e., one on each side of the screen) as they were presented simultaneously, while still maintaining the intensity and cohesion of the action unit. Therefore, test trials assessed action units of a 4-second duration, as compared to the 3-second action units embedded in the sequence.

Action Unit Labeling Task

Statistical Action Sequence

A novel statistical action sequence was created for the Statistical Sequence Phase of the Action Unit Labeling (AUL) Task to address the second question. The structure of this sequence was identical to the structure of the sequences described above for the

Segmentation Task in Research Question 1. However, a female actor performed the actions in this sequence (Appendix A).

Action Units

Stimuli for the AUL Phase of the AUL Task included videos of the two low-frequency units and two part-units derived directly from the sequence, as well as two non-units, created by recombining the two high-frequency units into two novel, distinct triads. High-frequency units themselves were not included because these units were equivalent to low-frequency units in terms of transitional probabilities between actions, and we were interested only in effects of statistical learning of these probabilities – rather than effects of frequency of exposure. In addition, a male actor was video-recorded performing the 12 actions of the sequence, and these action videos were recombined into the same two low-frequency units, two part-units, and two non-units. Actions were slowed down by 33% relative to their duration in the full sequence.

Four novel verbs were used to label action units. Two verbs were trained and tested, and two untrained verbs were used just for testing purposes (Table 1). All auditory stimuli were recorded by a female native English speaker.

Table 1

List of verb-action pairings participants were trained and tested on

Action Unit (order counterbalanced)	**Trained Verb**	**Untrained Verb**
Low-Frequency Unit	Keefing	Javing
Part-Unit/Non-Unit	Pilking	Moding

Additional Assessments

Three additional assessments were included in the study. The first assessed nonverbal intelligence, specifically a measure of visual-spatial working memory, and the second and third assessed language outcomes.

Visual-Spatial Working Memory Task

Children’s nonverbal visual-spatial working memory was assessed with the Zoo Locations task of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV; Wechsler, 2012). This task has an average internal consistency reliability of .90 and has a *g* (general intelligence) loading of .55 for children ages 2:6 through 3:11 (Raiford & Coalson, 2014). In this task, children view animal cards arranged on a gridded zoo for a specified time (three or five seconds) and are asked to reproduce the arrangement (Figure 2). The test begins with a sample item and proceeds until participants answer incorrectly on two consecutive trials. Zoo layouts progressively increase from two to nine zoo locations, and the number of animal cards placed during a single trial ranges from one to seven. Children are given corrective feedback and are offered a second chance through the second trial; they then receive only one attempt on subsequent trials. Children receive one point for each correct response, with scores ranging from 0 to 20.

Figure 2. Example layout from the WPPSI-IV Zoo Locations task. Participants reproduce the previously viewed arrangement of animal cards on the gridded zoo.

Language Assessments

Verb Comprehension Task

Children's knowledge of verbs was assessed with a touchscreen-adapted version of the Verb Comprehension Task (VCT; Konishi et al., 2016). Concurrent validity of this task with parental report of verb knowledge on the MacArthur–Bates Communication Development Inventory is .42 for 27- to 33-month-olds (Konishi et al., 2016). This task presented children with a split-screen, side-by-side video display of two different actions performed by the same human actor, with pre-recorded audio of a female native English speaker prompting children to choose the appropriate action for a given verb. For example, in one of the trials, children heard the prompt, "Where is she *marching*?" while a video clip of a woman marching and a video clip of the woman spinning were presented side by side (Figure 3).

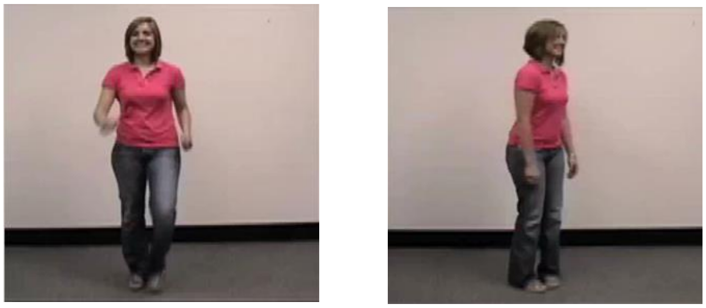


Figure 3. Example from the Verb Comprehension Task. Participants select the video that corresponds to a given word (e.g., “Where is she *marching*?”).

Paired actions were performed by the same actor, but seven female actors appeared throughout the task in order to keep children engaged. The verbal prompt was presented once for each trial. Trials lasted 10 seconds; if children did not respond within this window, an incorrect response was marked and the task moved on to the next trial. The task took approximately 4 to 7 minutes to administer, and included two types of trials: practice and test trials.

Practice trials displayed two familiar objects side by side on the touchscreen and prompted children with audio asking them to select the appropriate image for a given noun. For example, one trial showed children images of a banana and cookie and asked, “Can you show me the banana?” Three practice trials at the beginning of the VCT were designed to familiarize children with the side-by-side display, audio prompts, and the task of selecting the appropriate referent by touching the image on the screen. Nine additional practice trials were placed throughout the VCT – one after every four test trials – to ensure children’s continued attention and effort throughout the task. Children’s mean accuracy for the twelve nouns was 98%, indicating children understood and maintained attention throughout the task.

Test trials assessed children’s knowledge of 36 verbs, by presenting each of 18 verb pairs twice for a total of 36 trials. The first 18 trials asked children to identify one of the verbs in each pair (e.g., “Where is she *spinning*?”), and the second 18 trials, presented in the same order, asked children to identify the other verb in the pair (e.g., “Where is she *marching*?”). Twenty-nine trials tested transitive verb pairs, whose referents were depicted using the same object (e.g., *kicking* the balloon vs. *throwing* the balloon), and seven trials tested intransitive verb pairs (e.g., *spinning* vs. *marching*) (Appendix B).

Response accuracy and reaction times were recorded using E-prime 2.0 (Psychological Software Tools, Inc.). Accuracy was coded as the proportion of the 18 verb pairs correctly identified. Reaction time was coded as the average reaction time for correct test trial pairs.

Quick Interactive Language Screener

The Quick Interactive Language Screener (QUILS), a touchscreen 15-minute assessment designed for 3- through 5-year-olds, was used to evaluate children’s language comprehension skills in three distinct component areas: Vocabulary, Syntax, and Process, with high internal consistency reliabilities (0.79, 0.79, and 0.87, respectively) and concurrent validity with the Peabody Picture Vocabulary Test 4th Edition (.67, .54, .58, respectively) and the Preschool Language Scale 5th Edition, Auditory Comprehension subtest (.59, .54, .62, respectively) (Golinkoff et al., 2017). The QUILS begins with three practice items to ensure children understand how to select their answer choices on the screen before the scored items begin. The Vocabulary component provides a comprehensive measure of children’s acquired vocabulary products, by assessing children’s knowledge of closed-class words (prepositions and conjunctions) and openclass words (nouns and verbs; Figure 4). The Syntax component evaluates the grammar children know, including their understanding of WH-questions, past tense, prepositional phrases, and embedded clauses. The Process component assesses children's ability to acquire novel nouns, adjectives, and verbs using linguistic processes such as fast mapping and syntactic bootstrapping, and the ability to convert novel active verbs to the passive form.

Figure 4. Example of a verb test item from the QUILS Vocabulary component. Participants select the image that corresponds to a given word (e.g., "Who is *unlocking* something?").

Each of the three QUILS components assesses 16 test items, for a total of 48 test items. A male native English speaker provides the auditory prompts for all items, and children are allotted 20 seconds following each prompt to select a response. If no option is selected, the prompt is repeated, and an additional 15 seconds is allotted before the test moves on to the next item. The QUILS proceeds automatically from one item to the next, and the ordering of items is identical for all participants.

The QUILS produces raw accuracy scores for each QUILS component, ranging from 0 to 16. In addition, reaction time data were averaged for accurate items in each QUILS component to provide an additional measure of children's language knowledge in each area.

Procedure and Design

All participants met individually with the experimenter in a lab setting. First, the Segmentation Task was presented in two blocks – one for each Statistical Sequence Phase and its respective Segmentation Test Phase – interleaved with the Action Unit Labeling Task (Table 2), because pilot testing revealed that this order was more engaging for children. The study moved automatically through these tasks and, to keep children engaged, presented short, fun animated videos between these tasks. Children were seated in front of a large monitor (30-inches for 35 children tested at one site and 24-inches for 29 children tested at the second site). A hidden video camera attached to the top center of the monitor captured eye gaze for off-line coding for all participants, and a Tobii X60 table-mounted eye tracker recorded eye movements for the participants tested at one of the two sites (N=35). Prior to beginning the experiment, the eye-tracker was calibrated to accurately track children's eyes. For the calibration process, children watched a red dot move around the screen to nine calibration points. The eye-tracker recorded children's binocular eye fixations at a sampling rate of 60 Hz.

Children's eye gaze for the Segmentation and AUL Tasks were either coded online by the eye tracker or offline, frame-by-frame (29.97 frames/s) by researchers blind to condition. Offline coding of eye gaze was necessary for children tested without an eye tracker (N = 29), and for children whose eye gaze was not properly tracked by the eye tracker (N = 4). Intercoder reliability was computed between pairs of coders on a random selection of 20% of the videos to establish reliability for looks coded as left, right, center, or away. Intercoder reliability was *r* = .93.

Table 2
Study procedure: Order of tasks

Task	Apparatus	Research Question
Segmentation Task Block 1	Video Camera and/or Eye-tracker	Research Question 1
Action Unit Labeling Task		Research Question 2
Segmentation Task Block 2		Research Question 1
Zoo Locations Task	Not Applicable	
Verb Comprehension Task	Touchscreen Laptop	
Quick Interactive Language Screener		

Segmentation Task

Statistical Sequence Phase

First, a video of one of two statistical action sequences played continuously for its full 126-second duration, presented in Block 1. The other sequence was presented in Block 2. The order of presentation of the two sequences was counterbalanced across participants.

Segmentation Test Phase

Immediately following each Statistical Sequence Phase was a Segmentation Test Phase of 6 test trials, designed using a preferential looking paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987), which has been used in prior research assessing event categorization (Pruden, Roseberry, Göksun, Hirsh-Pasek, & Golinkoff, 2013). Each test trial presented a statistical action unit or part-unit on one side of a split-screen and a non-unit on the other side of the screen (Figure 5). At the beginning of each test trial, children were prompted with the question, “Which video do you remember seeing?” For each test trial, the action unit/part-unit and action non-unit played side by side simultaneously twice, with a blank black screen presented briefly (duration of 0.6 seconds) between stimulus presentations, to avoid creating a transition from the last action of the unit with the first action of the unit. The two presentations of each action unit, combined with the blank screen between iterations, produced test trials of 8.6 seconds in duration.

Each test trial was preceded by a 3-second centering trial to redirect children’s attention to the middle of the video monitor (Figure 5). Centering trials displayed a video of a laughing baby accompanied by audio instructions to look at the screen (e.g., “Look up here!”).

The six test trials assessing action segmentation for each sequence presented children with the two high-frequency units, two low-frequency units, and the two part-units – all paired with distinct non-units. The sides of the screen in which the unit/part-unit and non-unit were presented were randomized for each trial to minimize side bias, and the order of trial presentation was randomized for each participant to minimize order effects.

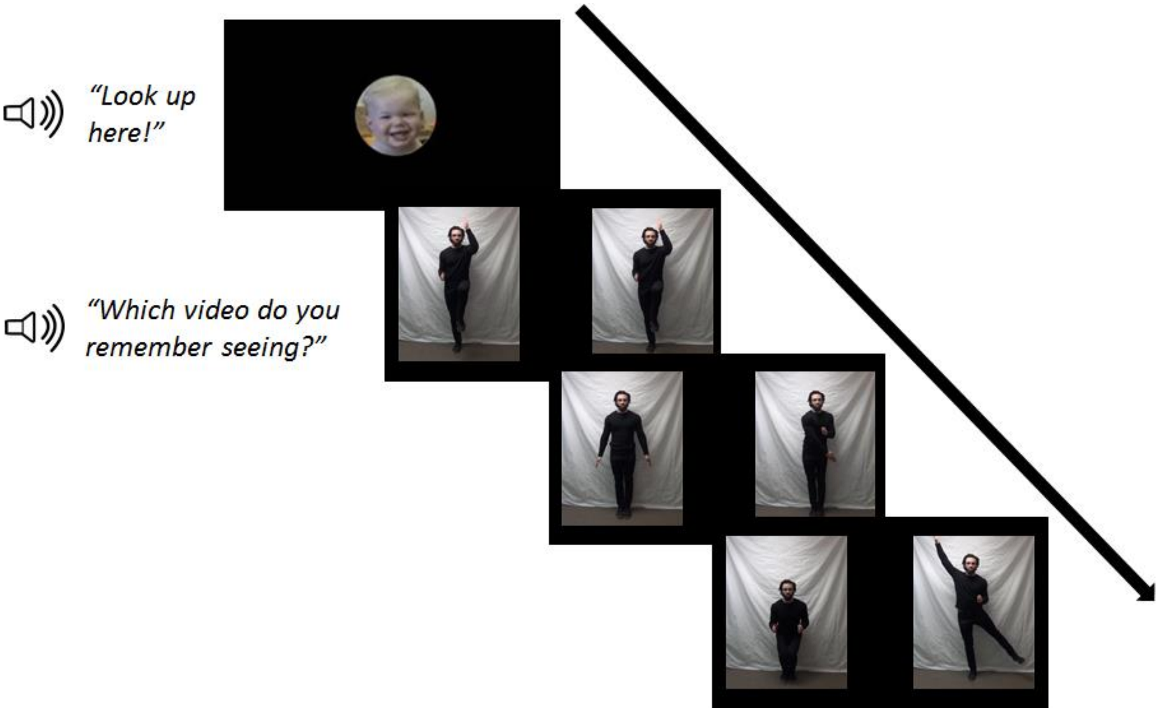


Figure 5. Schematic of how the Segmentation Test Phase proceeded. This image shows a trial testing children's fixation to a high-frequency action unit vs. an action non-unit using a split-screen preferential looking paradigm. The centering trial preceding each test trial was presented for 3 seconds accompanied by audio instructing the child to look at the screen. Each pair of action units had a duration of 4 seconds and was presented twice in each test trial, with the two presentations separated by a 0.6 second duration black screen (not shown here).

Action Unit Labeling Task

Statistical Sequence Phase

Following the Segmentation Task Block 1, the AUL Task began with a video of a novel statistical action sequence, featuring a female actor, playing continuously for its full 126-second duration.

Action Unit Labeling Phase

Immediately following the Statistical Sequence Phase was the AUL Phase. In this phase, two verbs were trained and tested using a stringent test of word learning (Brandone et al., 2007; Golinkoff, Jacquet, Hirsh-Pasek, & Nandakumar, 1996; Hollich et al., 2000; Roseberry, Hirsh-Pasek, & Golinkoff, 2014; Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009) which makes use of the Intermodal Preferential Looking Paradigm (IPLP; Golinkoff et al., 1987; Hirsh-Pasek & Golinkoff, 1996), an established method of examining language comprehension. The stringent test of word learning assesses children's ability to extend trained words to novel exemplars, their ability to use the principle of mutual exclusivity (Markman, 1991) to avoid mapping different labels onto already-labeled actions, and their ability to retain trained mappings.

Children learned one verb paired with a statistically consistent action unit and one verb paired with either a part-unit or a non-unit. Children were randomly assigned into the following groups: ten children learned labels for a statistically consistent action unit first and a non-unit second (*Mean age* = 41.58 months; *SD* = 4.42; 7 females); ten children learned labels for a statistically consistent action unit first and a part-unit second (*Mean age* = 42.27 months; *SD* = 3.88; 7 females); ten children learned labels for a non-unit first and a statistically consistent action unit second (*Mean age* = 40.08 months; *SD* = 2.58; 5 females); and ten children learned labels for a part-unit first and a statistically consistent action unit second (*Mean age* = 43.41 months; *SD* = 3.93; 8 females). The order of verb learning was counterbalanced to capture potential effects of verb order, because prior studies have found that young children (30- to 35-month-olds) trained on two novel verbs are better able to learn the second verb than the first (Roseberry et al., 2009).

For each verb, the AUL Phase began with a Salience Trial, followed by Verb Training, followed by Extension, New Verb, and Recovery Test Trials (Figure 6). For the Salience Trial and three Test Trials, two action units (or two part-units or non-units) played side by side simultaneously twice, and for the Verb Training, one of those action units (or part-units or non-units) played on the full screen twice. A blank black screen was presented briefly (duration of 0.6 seconds) between the two stimulus presentations within each trial, to avoid creating a transition from the last action of the unit with the first action of the unit. The two presentations of each action unit, combined with the blank screen between iterations, produced trials of 8.6 seconds in duration.

Each Trial and Verb Training was preceded by a 3-second centering trial to return children's attention to the middle of the screen, in order to avoid side bias based on a prior trial. The centering trial displayed a video of a laughing baby accompanied by audio specific to the upcoming Trial or Verb Training. For centering trials preceding a Salience Trial, the audio instructed children to look at the screen (i.e., "What's up here?"). For centering trials preceding Verb Training, the audio introduced the novel word to be trained (e.g., "*Pilking*!"), because it has been shown that hearing a verb prior to viewing the labeled action may facilitate verb learning (Tomasello & Kruger, 1992). Finally, for centering trials preceding Test Trials, the audio presented the verb that would be tested in the upcoming Test Trial, in order to prepare children to look for the matching action unit.

Salience Trial. Children viewed a preview of the exact test clips, featuring a male actor, that would be presented during the Extension, New Verb, and Recovery Test Trials. Accompanying this preview were audio instructions directing children's attention to the screen (i.e., "What's going on up here?").

Measuring looking time to this split-screen presentation prior to Verb Training allowed for detection of a priori preferences for either of the paired clips. The two actions in each pair of clips were experienced equally in the full sequence – with equivalent frequency and equivalent within-unit TPs (e.g., two part-units) – therefore, no a priori preference was expected. An absence of preference in the Salience Trial indicates that preferences during the Test Trials were due to an effect of Verb Training.

Verb Training. After the Salience Trial, children were trained on the novel verb-action unit pairings. For each verb, the action unit that appeared during Verb Training was randomized (e.g., part-unit 1 or part-unit 2). Participants watched a video clip of the action unit, performed by the same female actor who had performed the full sequence, paired with audio presenting a novel verb using full syntax (“Look at her *pilking*! She’s *pilking*! She is *pilking*!”). The verb was presented three times, and the audio presenting the verb occurred throughout the action unit triad, to increase the likelihood that the verb would be attached to the full unit and not to a single action within that unit. This Verb Training is similar to children’s experiences of learning real verbs, many of which also have component pieces, such as the consecutive action steps reflected in the verbs ‘tying’, ‘washing,’ and ‘baking’.

Extension Test Trial. Following Verb Training was the Extension Test Trial, the first test trial for each trained verb. Audio asked participants to find the action unit labeled by the trained verb (“Where is *pilking*? Can you find *pilking*? Look at *pilking*!”). This trial was designed to assess participants’ ability to generalize the trained verb to a novel exemplar of the action unit, performed by a novel actor. Studies have shown that if children learned the verb-action pairing, they should prefer looking at the action from training when asked to find the target word at test (e.g., Roseberry et al., 2009).

New Verb Test Trial. The second test trial, the New Verb Test Trial, asked participants to find a novel action that was not labeled during Verb Training (“Where is *moding*? Can you find *moding*? Look at *moding*!"). This trial was designed to test whether participants had truly learned that the verb (e.g., *pilking*) maps onto the particular action unit (e.g., part-unit 2) by testing whether they would avoid mapping other verbs to that action unit (Hollich et al., 2000). Based on the principle of *mutual exclusivity* (Markman, 1991), there should be a preference for linking an action with only a single verb. If children learned the target verb (e.g., *pilking*) that was trained, then upon hearing a novel verb (e.g., *moding*) they should not look at the action unit that was paired with the trained verb. They may look instead toward the nonmatching action unit (e.g., part-unit 1) or may show no preference for either action unit (Roseberry et al., 2009), indirectly demonstrating their unwillingness to map a new label to a previously labeled action unit.

Recovery Test Trial. The final test trial, the Recovery Test Trial, was identical to the Extension Test Trial, asking children to renew their attention to the action unit labeled during Verb Training. If children successfully mapped the trained verb onto the action unit with which it was paired and retained that mapping, then they should recover their attention toward that action unit when the trained verb is presented once more.

Verb learning would be reflected by a quadratic pattern of increased looking to the labeled action unit during the Extension Test Trial, decreased looking to this unit during the New Verb Test Trial, and renewed looking during the Recovery Test Trial. In addition, children's looking toward the labeled action unit should be greater than chance for the Extension and Recovery Test Trials if children learned the verb.

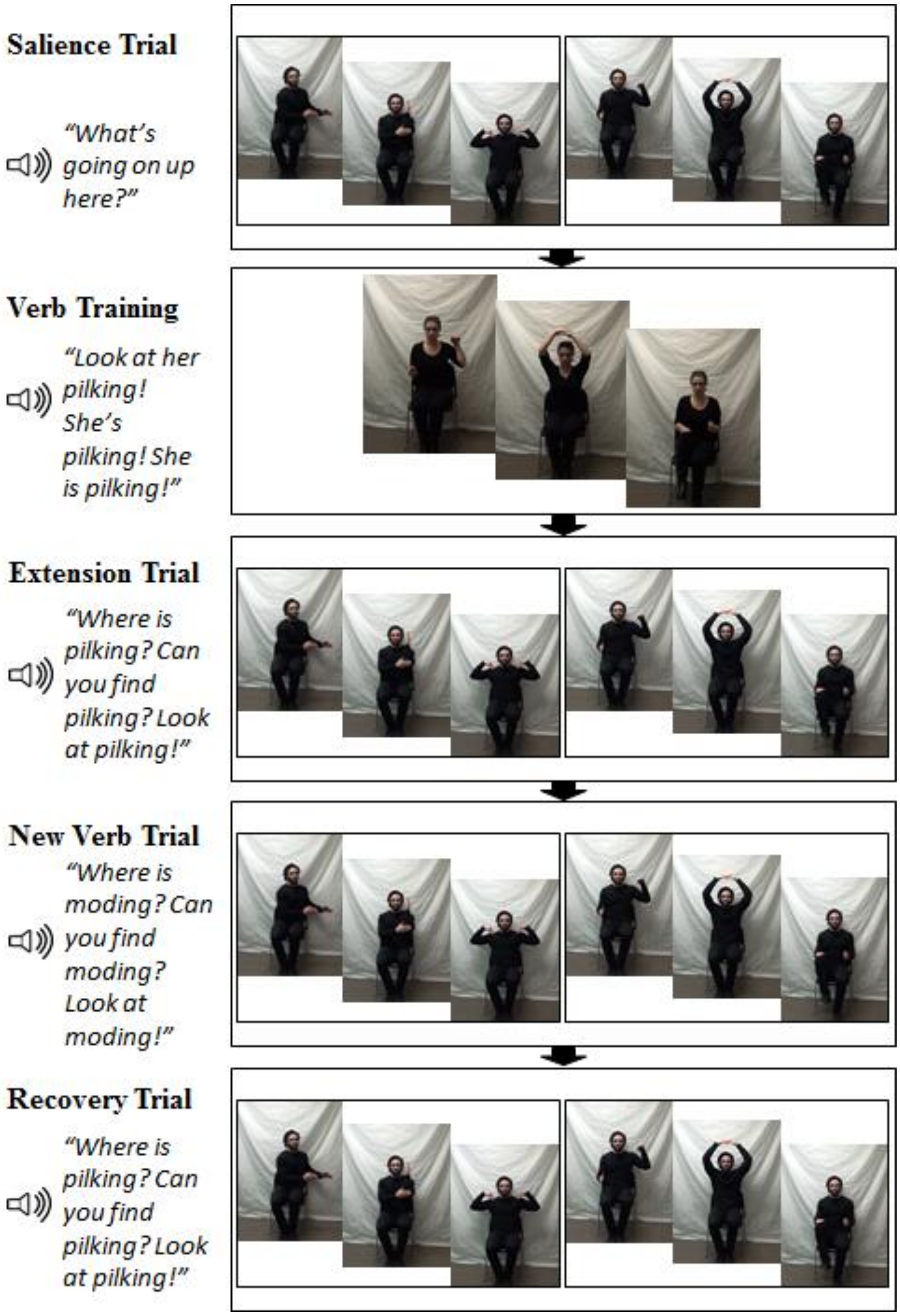


Figure 6. Schematic of how the Action Unit Labeling (AUL) Phase proceeded for each verb. The AUL Phase, presented immediately following the Statistical Sequence Phase, trained and tested children on two novel verbs. For each verb, a Salience Trial was followed by Verb Training, followed by three Test Trials. Randomization determined which of the two action units presented during the Salience Trial would be paired with the verb during Verb Training. A centering trial preceded each Trial and Verb Training (not shown here).

Additional Assessments

Following the Segmentation and AUL Tasks, children performed the Zoo Locations task, the VCT, and QUILS (Table 2). For the Zoo Locations task, participants were seated at a table directly across from the experimenter, and children's performance was coded online by the experimenter. For the VCT and QUILS, participants were seated in front of a touchscreen laptop (13-inch screen), and the software programs coded participants' responses automatically for accuracy and reaction time. Reaction time measures were included because growing evidence suggests that vocabulary knowledge is not all-or-nothing but exists along a continuum (Fernald & Marchman, 2012; Hendrickson, Poulin-Dubois, Zesiger, & Friend, 2017; Seston, Golinkoff, Ma, & Hirsh-Pasek, 2009).

The entire session took 40-50 minutes (Segmentation and AUL Tasks, Zoo Locations task, VCT, and QUILS). The fixed order of this procedure was used to prevent the language assessments (VCT and QUILS) from influencing children's performance on the Segmentation and AUL Tasks and the Zoo Locations task. For example, the QUILS Vocabulary component assesses children's knowledge of prepositions such as *above* and *below*; familiarization with these words, particularly for children who understood them, could have improved children's spatial working memory strategies on the Zoo Locations task.

Data Analysis

Research Question 1: Is Action Segmentation Linked with Vocabulary Knowledge?

Looking time for trials in the Segmentation Task was coded as the looking time to the target unit (i.e., high-frequency action unit, low-frequency action unit, or part-unit) divided by the total looking time to the target unit and non-target unit (i.e., non-unit). To ensure children had enough time to distinguish the two test items, children needed to view, at a minimum, 2 seconds of each test trial, which was enough to see at least one full action (1.33 seconds) and half of a second action (0.67 seconds) in a triad. The proportion of looking to the target unit was averaged over the six trials for each segmentation sequence – across the three unit types (high-frequency unit, low-frequency unit, part-unit) – to create a measure of action segmentation ability. We predicted 3-year-olds' action segmentation ability would explain a significant amount of variance in VCT verb knowledge and QUILS vocabulary knowledge.

Research Question 2: Does Action Segmentation Facilitate Verb Learning?

Looking time for Salience and Test Trials in the AUL Task was coded as the total looking time to the target unit (i.e., the action unit labeled during Verb Training) divided by the total looking time to the target unit and non-target unit. To ensure children had enough time to distinguish the two test items, children needed to view, at a minimum, 2 seconds of each trial, which was enough to see at least one complete action (1.33 seconds) and half of a second action (0.67 seconds). Children also needed to view a minimum of 3.33 seconds of each Verb Training, to ensure children were exposed to the verb paired with at least two complete actions (2.66 seconds) and half of the third action in the triad (0.67 seconds). We hypothesized that 3-year-olds would show greater learning of verbs labeling statistically consistent action units (i.e., low-frequency units) than verbs labeling inconsistent action units (i.e., part-units or non-units), and greater learning of verbs labeling part-units than non-units. Verb learning would be reflected by a quadratic pattern, of increased looking to the labeled action unit during the Extension Test Trial, decreased looking to this action unit during the New Verb Test Trial, and renewed looking during the Recovery Test Trial, when the trained verb was again presented. Additionally, verb learning would be reflected by greater than chance looking during the Extension and Recovery Test Trials, and chance or less than chance looking during the New Verb Test Trial.

CHAPTER 3

RESULTS

The results are organized around the two research questions. In the first section, we address the question of whether 3-year-old's statistical action segmentation abilities and their vocabulary knowledge are linked. From the subset of participants who completed the Segmentation Task, Zoo Locations task, VCT and QUILS (N=54), we report on the reliability of our measure of statistical action segmentation ability, present group-level analyses of children's action segmentation performance, examine correlations between all of the key variables, and, finally, we present hierarchical regression analyses predicting vocabulary knowledge from action segmentation ability, controlling for covariates.

In the second section, we explore whether action segmentation facilitates verb learning in 3-year-olds. From the subset of participants who completed the AUL Task (N=40), we report preliminary analyses showing no differences in overall attention between conditions and demonstrating an absence of visual preference for actions prior to verb training. Finally, we present results of a mixed design ANOVA assessing whether the statistical coherency of actions influenced the success of verb learning, with follow up analyses comparing visual preferences during successful verb learning trials to chance levels.

Research Question 1: Is Action Segmentation Linked with Vocabulary Knowledge?

We predicted that 3-year-olds would demonstrate a reliable ability to statistically segment novel actions, as has been demonstrated in infants and adults (e.g., Baldwin et al., 2008; Stahl et al., 2014), and that 3-year-olds who were better able to extract statistical action units would have greater verb and overall vocabulary knowledge.

Table 3 presents descriptive statistics for total visual attention to the Statistical Sequence Phase and action segmentation accuracy for blocks 1 and 2 of the Segmentation Task, as well as the key language measures and the nonverbal measure of working memory. Action segmentation accuracy was calculated as the proportion of children's looking time to the target unit (i.e., high-frequency unit, low-frequency unit, and part-unit) relative to the non-unit, averaged across the six test trials for each block.

Of note, the distribution for children's accuracy on the VCT showed substantial skewness (skew = -1.98, SE skew = .32) and kurtosis (kurtosis = 4.98, SE kurtosis = .64), with 22.22% of children answering all verb vocabulary items correctly. This distribution artificially reduces performance differences between participants, and could reduce the likelihood of uncovering significant correlations with other measures. Thus, lack of correlations involving this measure should be interpreted with caution. Other than verb accuracy on the VCT, there was no substantial skewness or kurtosis in the other variables.

Table 3

Descriptive statistics for key variables for Research Question 1

Variable	*M*	*SD*	Range
Statistical Sequence Phase Attention (out of 126-s)			
Block 1	84.20	19.84	(21.98, 121.87)
Block 2	62.49	21.64	(21.80, 126.00)
Action Segmentation Accuracy (out of 1.0)			
Block 1	.50	.09	(.25, .67)
Block 2	.49	.06	(.37, .63)
VCT Verb Accuracy (out of 18)	15.48	2.83	(4, 18)
VCT Verb RT	3.83	.81	(2.67, 6.27)
QUILS Vocab Accuracy (out of 16)	9.17	2.76	(4, 14)
QUILS Vocab RT	5.02	2.01	(2.70, 13.02)
QUILS Syntax Accuracy (out of 16)	8.56	3.11	(2, 15)
QUILS Syntax RT	5.92	2.50	(2.49, 12.27)
QUILS Process Accuracy (out of 16)	7.89	3.56	(2, 15)
QUILS Process RT	3.65	1.73	(1.10, 9.93)
Zoo Locations WM Task (out of 20)	8.30	2.08	(3, 12)

Notes. N = 54 for all variables except Block 2 of Statistical Sequence Phase Attention and Action Segmentation Accuracy, N = 22. VCT = Verb Comprehension Task; QUILS = Quick Interactive Language Screener; RT = Reaction time; WM = Working memory.

We evaluated whether action segmentation ability was a reliable construct by assessing the association between children's mean accuracy scores for the two blocks of the Segmentation Task. The correlation between accuracy scores was significant ($r(22) = .69$, $p < .001$), suggesting the Segmentation Task reliably captured children's statistical action segmentation. Given this reliability between the two sequences and the fact that less than half of the children completed segmentation of the second sequence, subsequent analyses were conducted using just the data from the first block of the Segmentation Task for each child.

Next, we explored at the group level children's performance on the Segmentation Task. Mean overall accuracy on the Segmentation Task ($M = .50$, $SD = .09$) did not differ from chance ($t(53) = -0.36$, $p = .72$, $d = 0.05$). Further, there was no difference in accuracy on trials assessing children's segmentation of high-frequency units, low-frequency units, and part-units (*F*(2,106) = 0.40, *p* = .67, η_p^2 = .01; Figure 7). Despite this null result indicating group-level action segmentation was at chance, the subsequent analyses address the key research question of how individual differences in action segmentation ability may relate to individual differences in vocabulary knowledge. Given the absence of any difference between children's parsing of high-frequency units, low-frequency units, and part-units, these three measures were averaged and combined into one measure of action segmentation ability.

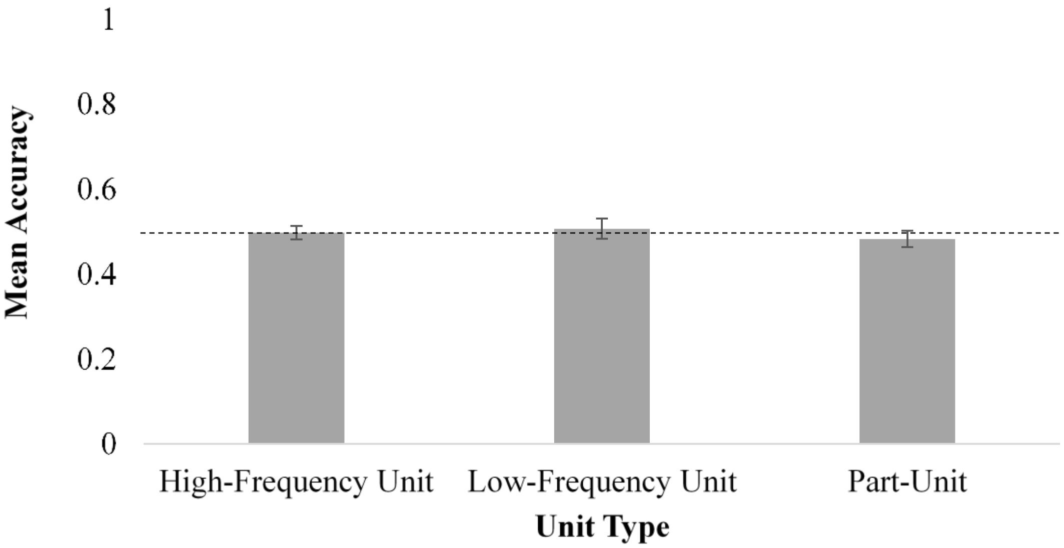


Figure 7. Mean accuracy of children choosing the correct answer across the three Unit Types (High-Frequency Unit, Low-Frequency Unit, and Part-Unit). The dotted line indicates the chance level accuracy (50%). Error bars represent the ±1 standard error.

To assess whether individual differences in action segmentation ability predict verb and overall vocabulary knowledge, we first explored zero-order correlations between action segmentation ability, accuracy and reaction times on the VCT and QUILS Vocabulary, as well as the potential covariates age, performance on the Zoo Locations working memory task, attention to the Statistical Sequence Phase, as well as QUILS Syntax and Process (Table 4).

Table 4

Correlations between action segmentation ability, VCT and QUILS Vocabulary measures, and potential covariates

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. ASA	–	.14	-.48***	.46***	-.44***	.09	-.37**	.19	-.34*	.01	.35**	.05
2. VCT Verb AC		–	-.45***	.47***	-.23	.47***	-.29*	.35**	-.41**	.33*	.40**	.19
3. VCT Verb RT			–	-.58***	.38**	-.38**	.69***	-.55***	.55***	-.27†	-.34*	-.09
4. QUILS Voc AC				–	-.49***	.53***	-.48***	.61***	-.38**	.32*	.38**	.08
5. QUILS Voc RT					–	-.08	.41**	-.28*	.44***	-.26†	-.36**	-.11
6. QUILS Synt AC						–	-.25†	.48***	-.25†	.21	.13	.10
7. QUILS Synt RT							–	-.38**	.69***	-.17	-.31*	-.18
8. QUILS Proc AC								–	-.23†	.25†	.22	.17
9. QUILS Proc RT									–	-.23†	-.48***	-.25†
10. Age										–	.22	.11
11. WM											–	.15
12. Attention												–

Notes. ASA = Action segmentation ability; VCT = Verb Comprehension Task; QUILS = Quick Interactive Language Screener; Voc = Vocabulary; Synt = Syntax; Proc = Process; AC = Accuracy; RT = Reaction time; WM = Working memory.

† *p* < .10. * *p* < .05. ***p* < .01. ****p* < .001.

Action segmentation ability showed a positive correlation with QUILS Vocabulary accuracy ($r(54) = .46$, $p < .001$), and a comparable negative correlation with mean reaction time to correctly answered items on that measure ($r(54) = -.44$, $p < .001$). There was no relation between action segmentation ability and accuracy on the VCT ($r(54) = .14$, $p = .31$), but there was a moderate negative correlation with reaction time for correctly identified verbs on the VCT ($r(54) = -.48$, $p < .001$). Action segmentation ability was also negatively correlated with reaction time on QUILS Syntax ($r(54) = -.37$, $p < .01$) and QUILS Process ($r(54) = -.34$, $p = .01$), though these correlations were weaker, and correlations with accuracy scores on those measures were nonsignificant (QUILS Syntax accuracy: $r(54) = .09$, $p = .50$; QUILS Process accuracy: $r(54) = .19$, $p = .17$). Correlations between visual attention to the Statistical Sequence Phase and all other variables, including action segmentation ability, were nonsignificant (all *p*'s > .05). Working memory was significantly correlated with action segmentation ability and most language measures, and age was correlated with some language measures, though not with action segmentation ability (see Table 4).

Partial correlations were calculated for all significant correlations between action segmentation ability and vocabulary measures, controlling for covariates age and working memory. Partial correlations remained significant for VCT verb reaction time ($r(54) = -.44$, $p < .01$) as well as QUILS Vocabulary accuracy ($r(54) = .41$, $p < .01$) and reaction time ($r(54) = -.39$, $p < .01$).

Hierarchical regression analyses were then conducted to examine the prediction of vocabulary knowledge from action segmentation ability (Table 5). Children's age in months was entered first, followed by visual-spatial working memory, followed by action segmentation ability scores in the third and final step.

The results of the hierarchical regressions indicate that adding action segmentation ability in step 3 significantly improved all three models. Controlling for age and visual-spatial working memory, action segmentation ability (i.e., the ability to distinguish statistical action units and part-units from non-units) explained 13% unique variance in QUILS Vocabulary accuracy scores ($\Delta F(1,50) = 10.24$, $p < .01$), 13% unique variance in QUILS Vocabulary reaction time ($\Delta F(1,50) = 8.91$, $p < .01$), and 16% unique variance in VCT verb reaction time ($\Delta F(1,50) = 11.73$, $p < .01$). Based on these models, when action segmentation ability increases by one standard deviation, QUILS Vocabulary accuracy scores are predicted to increase by .39 standard deviations, QUILS Vocabulary reaction times are predicted to decrease by .38 standard deviations, and VCT verb reaction times are predicted to decrease by .43 standard deviations (Table 5).

Given the significant intercorrelations between language measures of vocabulary, syntax and process (Table 4), and the hypothesized unique relation between action segmentation ability and vocabulary, we ran the same hierarchical regressions adding in a step to control for QUILS Syntax and Process accuracy and reaction time measures (Table 6). Even after controlling for age, working memory, and QUILS Syntax and Process accuracy and reaction time measures, action segmentation ability explained 7% unique variance in QUILS Vocabulary accuracy scores ($\Delta F(1,46) = 8.20$, $p < .01$), 7% unique variance in QUILS Vocabulary reaction time ($\Delta F(1,46) = 5.17$, $p = .03$), and 5% unique variance in VCT verb reaction time ($\Delta F(1,46) = 6.16$, $p = .02$). Based on these models, when action segmentation ability increases by one standard deviation, QUILS Vocabulary accuracy scores are predicted to increase by .29 standard deviations, QUILS Vocabulary reaction times are predicted to decrease by .30 standard deviations, and VCT verb reaction times are predicted to decrease by .25 standard deviations.

Table 5

Hierarchical regression analyses predicting children's vocabulary knowledge from action segmentation ability controlling for age and working memory

	QUILS Vocabulary Accuracy				QUILS Vocabulary Reaction Time				VCT Verb Reaction Time			
	B	*SE B*	*β*	*F*	*B*	*SE B*	*β*	*F*	*B*	*SE B*	*β*	*F*
Step 1												
Age	.24	.10	.32*	6.00*	-.14	.07	-.26†	3.78†	-.06	.03	-.27†	3.94†
Step 2												
Age	.19	.10	.25†		-.10	.07	-.19		-.04	.03	-.20	
WM	.43	.17	.33*	6.61**	-.30	.13	-.31*	4.94*	-.12	.05	-.30*	4.64*
Step 3												
Age	.21	.09	.28*		-.12	.07	-.22†		-.05	.03	-.23†	
WM	.24	.17	.18		-.17	.13	-.17		-.05	.05	-.14	
ASA	12.05	3.77	.39**	8.61***	-8.50	2.85	-.38**	6.77***	-3.88	1.13	-.43**	7.66***

Notes. For QUILS Vocabulary accuracy, R^2 = .10 for Step 1; ΔR^2 = .10 for Step 2; ΔR^2 = .13 for Step 3. For QUILS Vocabulary reaction time, R^2 = .07 for Step 1; ΔR^2 = .09 for Step 2; ΔR^2 = .13 for Step 3. For VCT verb reaction time, R^2 = .07 for Step 1; ΔR^2 = .08 for Step 2; ΔR^2 = .16 for Step 3. QUILS = Quick Interactive Language Screener; VCT = Verb Comprehension Task; WM = Working memory; ASA = Action segmentation ability. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6

Hierarchical regression analyses predicting children's vocabulary knowledge from action segmentation ability controlling for age, working memory, and language syntax and process

	QUILS Vocabulary Accuracy				QUILS Vocabulary Reaction Time				VCT Verb Reaction Time			
	B	*SE B*	*β*	*F*	*B*	*SE B*	*β*	*F*	*B*	*SE B*	*β*	*F*
Step 1												
Age	.24	.10	.32*	6.00*	-.14	.07	-.26†	3.78†	-.06	.03	-.27†	3.94†
Step 2												
Age	.19	.10	.25†		-.10	.07	-.19		-.04	.03	-.20	
WM	.43	.17	.33*	6.61**	-.30	.13	-.31*	4.94*	-.12	.05	-.30*	4.64*
Step 3												
Age	.08	.08	.11		-.08	.07	-.14		-.01	.02	-.06	
WM	.27	.15	.20†		-.15	.14	-.15		-.02	.04	-.06	
Synt AC	.25	.10	.28*		.09	.09	.14		-.02	.03	-.08	
Proc AC	.25	.09	.33**		-.09	.09	-.17		-.07	.03	-.29*	
Synt RT	-.27	.16	-.24†		.13	.14	.16		.15	.04	.46**	
Proc RT	.09	.24	.06	9.70***	.25	.22	.22	3.12*	.05	.07	.10	11.52***
Step 4												
Age	.11	.07	.14		-.10	.07	-.18		-.02	.02	-.09	
WM	.17	.14	.13		-.07	.14	-.08		-.01	.04	.01	
Synt AC	.25	.09	.29**		.09	.09	.13		-.02	.03	-.09	
Proc AC	.24	.09	.31**		-.08	.08	-.14		-.06	.03	-.27*	
Synt RT	-.19	.15	-.18		.07	.14	.09		.13	.04	.40**	
Proc RT	.13	.22	.08		.23	.21	.20		.04	.06	.08	
ASA	8.96	3.13	.29**	10.76***	-6.75	2.97	-.30*	3.65**	-2.22	.90	-.25*	11.84***

Notes. For QUILS Vocabulary accuracy, R^2 = .10 for Step 1; ΔR^2 = .10 for Step 2; ΔR^2 = .35 for Step 3; ΔR^2 = .07 for Step 4. For QUILS Vocabulary reaction time, R^2 = .07 for Step 1; ΔR^2 = .09 for Step 2; ΔR^2 = .12 for Step 3; ΔR^2 = .07 for Step 4. For VCT verb reaction time, R^2 = .07 for Step 1; ΔR^2 = .08 for Step 2; ΔR^2 = .44 for Step 3; ΔR^2 = .05 for Step 4. QUILS = Quick Interactive Language Screener; VCT = Verb Comprehension Task; WM = Working memory; Synt = QUILS Syntax; Proc = QUILS Process; AC = Accuracy; RT = Reaction time; ASA = Action segmentation ability. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Research Question 2: Does Action Segmentation Facilitate Verb Learning?

The second research question asked whether 3-year-olds' verb learning for novel action units would be facilitated by prior opportunities to statistically segment an action sequence in which those units were embedded. Word learning would be reflected in a quadratic pattern of increased looking during the Extension Test Trial, decreased looking during the New Verb Test Trial, and renewed looking during the Recovery Test Trial. We predicted that children would exhibit this quadratic pattern of word learning for statistically consistent action units, and expected children would show weaker word learning for part-units, and weaker or no word learning for non-units. Given that some prior research has found children learn verbs better later on in an experiment (Roseberry et al., 2009), we also predicted a possible effect of verb order.

Preliminary analyses were conducted to establish that there were no differences between conditions on visual attention to the Statistical Sequence Phase or the Verb Training. For attention to the Statistical Sequence Phase, a one-way ANOVA revealed no differences between conditions ($F(3,36) = .57$, $p = .64$, $\eta_p^2 = .04$; Table 7). For Verb Training, a 2 (Inconsistent Action Unit: part-unit, non-unit) x 2 (Verb Order: consistent action unit labeled first, inconsistent action unit labeled first) x 2 (Action Unit Type: consistent action unit, inconsistent action unit) mixed design ANOVA revealed no main effect on attention to Verb Learning of Inconsistent Action Unit ($F(1,36) = 0.91$, $p = .35$, $\eta_p^2 = .02$), Verb Order ($F(1,36) = 0.99$, $p = .33$, $\eta_p^2 = .03$), or Action Unit Type ($F(1,36) = 0.01$, $p = .99$, $\eta_p^2 < .01$), and no significant interactions (all p's > .10) (Table 7). These analyses suggest that group-level differences on Test Trials cannot be attributed to differences in attention to either the Statistical Sequence Phase or Verb Training preceding the test trials.

Table 7
Descriptive statistics for attention variables for Research Question 2

	Condition (First Verb Trained, Second Verb Trained)							
	LF Unit, Non-Unit		LF Unit, Part-Unit		Non-Unit, LF Unit		Part-Unit, LF Unit	
Attention Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Statistical Sequence Phase (out of 126-s)	69.28	21.56	75.24	23.34	80.55	21.06	75.73	7.56
Verb Training (out of 8.6-s)								
Verb 1 (“*keefing*”)	7.91	1.25	7.72	1.46	8.09	0.84	8.07	0.67
Verb 2 (“*pilking*”)	7.62	1.58	7.16	1.59	7.85	1.17	7.48	1.56

Note. LF Unit = Low-Frequency Unit.

An additional preliminary analysis was conducted to establish that children did not show a preference for either 3-action unit presented during Salience Trials. Specifically, a mixed design analysis of variance (ANOVA) was conducted to compare, across conditions, children’s a priori preference during the Salience Trial for the unit that would be labeled in the subsequent Verb Training. There were two between-subjects factors, Inconsistent Action Unit (part-unit, non-unit) and Verb Order (consistent action unit labeled first, inconsistent action unit labeled first), and one within-subjects factor, Action Unit Type (consistent action unit, inconsistent action unit). A 2 (Inconsistent Action Unit) x 2 (Verb Order) x 2 (Action Unit Type) ANOVA revealed no main effect on Salience preference of Inconsistent Action Unit ($F(1,36) = 0.05$, $p = .83$, $\eta_p^2 < .01$), Verb Order ($F(1,36) = 1.83$, $p = .18$, $\eta_p^2 = .05$), or Action Unit Type ($F(1,36) = 1.57$, $p = .22$, $\eta_p^2 = .04$), and no significant interactions (all *p*’s > .10). This analysis indicated that group-level differences on test trials could not be attributed to a priori preferences for the labeled or non-labeled action unit.

To evaluate whether verb learning was facilitated by statistical segmentation, a mixed design ANOVA with two between-subjects factors and two within-subjects factors was conducted. The between-subjects factors were Inconsistent Action Unit (part-unit, non-unit) and Verb Order (consistent action unit labeled first, inconsistent action unit labeled first), and the within-subjects factors were Action Unit Type (consistent action unit, inconsistent action unit) and Test Trials (Extension, New Verb, Recovery). If, as expected, children learned labels for the statistically consistent action units but not the statistically inconsistent action units, then a within-subjects quadratic contrast should result for the Action Unit Type by Trials interaction. If children showed differential learning for labels for part-units versus non-units, then an effect of Inconsistent Action Unit may emerge. Finally, if, similar to prior research, children learn verbs better later in the experiment, then an effect of Verb Order may emerge.

The within-subjects quadratic contrast for Action Unit Type by Trials was not significant ($F(1,36) = 0.02$, $p = .89$, $\eta_p^2 < .01$), and there was no main effect of Action Unit Type ($F(1,36) = 0.14$, $p = .71$, $\eta_p^2 < .01$), Trials ($F(2,36) = 0.60$, $p = .55$, $\eta_p^2 = .02$), Inconsistent Action Unit ($F(1,36) = 0.18$, $p = .67$, $\eta_p^2 = .01$), or Verb Order ($F(1,36) = 1.45$, $p = .24$, $\eta_p^2 = .04$). However, a significant within-subjects quadratic contrast emerged for the interaction of Verb Order by Action Unit Type by Trials ($F(1,36) = 6.56$, $p = .01$, $\eta_p^2 = .15$; Figure 8). No other interactions were significant.

We conducted Tukey-corrected *post-hoc* simple effects contrasts to determine under what conditions the quadratic effect of Trial occurred. The quadratic effect of Trial was significant for inconsistent action units when the inconsistent action unit was the second verb labeled ($t(38) = 2.48$, $p < .05$, $r = .37$). Children's proportion of looking to the labeled action unit decreased from the Extension Trial ($M = .57$, $SD = .26$) to the New Verb Trial ($M = .47$, $SD = .20$), and was renewed during the Recovery Trial ($M = .58$, $SD = .24$). However, when the inconsistent action unit was the first verb labeled, children's proportion of looking to the labeled action was unchanged from the Extension Trial ($M = .50$, $SD = .16$) to the New Verb Trial ($M = .53$, $SD = .17$) to the Recovery Trial ($M = .46$, $SD = .17$) ($t(38) = -1.15$, $p > .05$, $r = .18$). The quadratic effect of trial was not significant for statistically consistent action units, whether the unit was labeled first ($t(38) = -0.35$, $p > .05$, $r = .06$) or second ($t(40) = 1.18$, $p > .05$, $r = .19$).

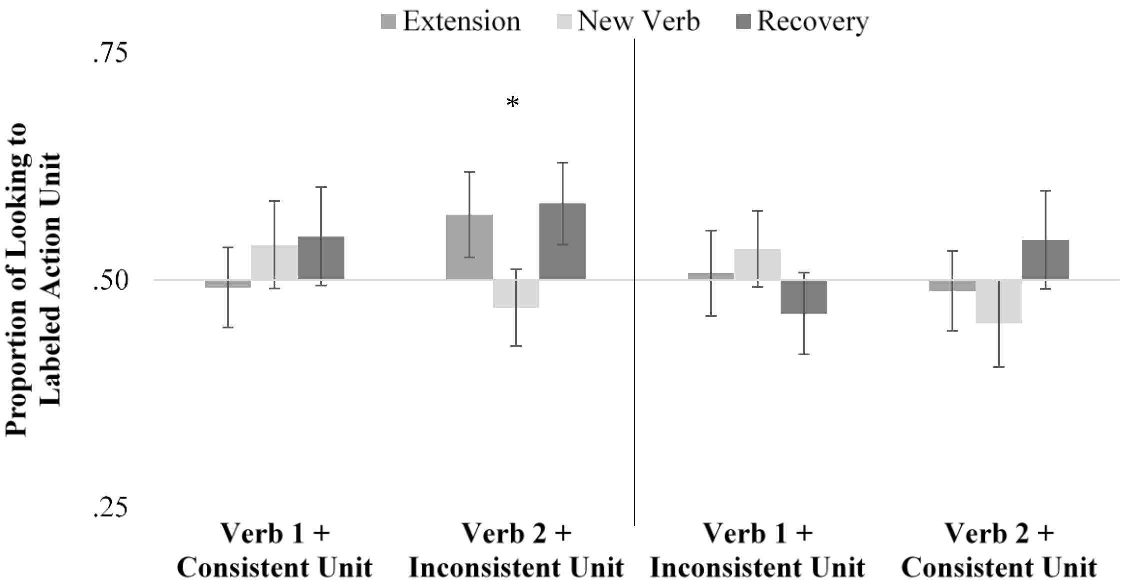


Figure 8. Proportion of looking time to the labeled action unit across three test trials for two verbs under two verb order conditions. Error bars represent the ±1 standard error. * denotes a significant quadratic pattern, *p* < .05.

Additional one-sample *t*-tests were conducted to determine whether children's accuracy was different from chance (50%) on the three test trials in the conditions in which verb learning occurred: verb 2 + part-units and verb 2 + non-units. One-sample *t*-tests revealed that for non-units, children were at chance in their looking to the correct action unit for all three trials (all *p*s > .10), but for part-units, children showed greater-than-chance looking to the labeled action unit on Extension and Recovery Test Trials that was trending toward significance (Extension: *M* = .64, *t*(9) = 2.01, *p* = .08, *d* = .64; Recovery: *M* = .64, *t*(9) = 2.20, *p* = .06, *d* = .69), while their looking to the New Verb Test Trial was at chance (*M* = .53, *t*(9) = 0.70, *p* = .50, *d* = .22) (Figure 9).

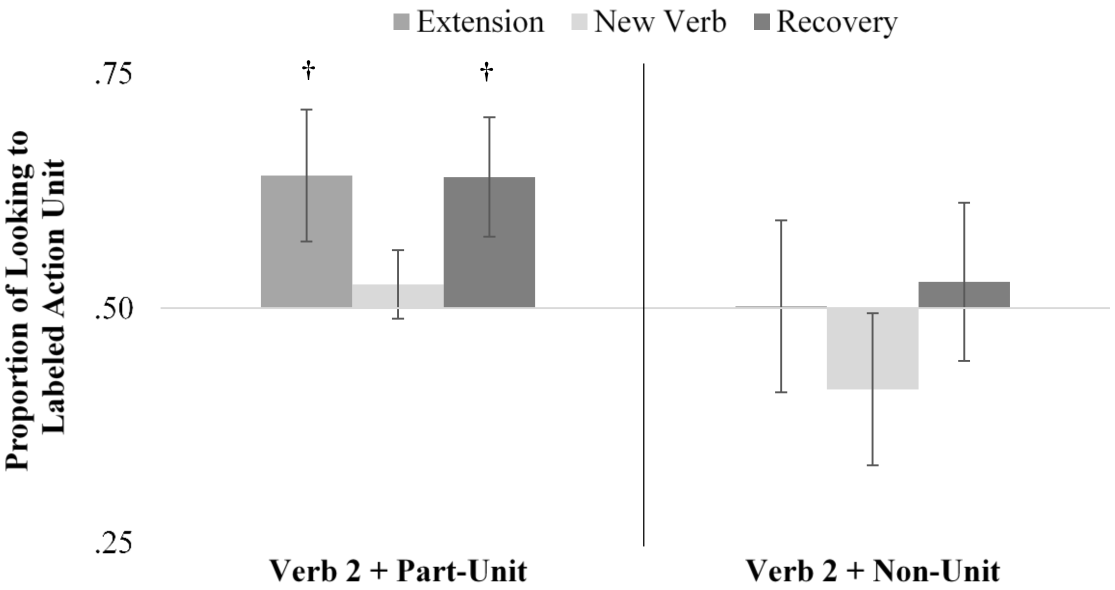


Figure 9. Proportion of looking time to the labeled action unit for verb 2, Part-Unit and Non-Unit. Error bars represent the ±1 standard error. † denotes a trending difference from chance, *p* < .10.

CHAPTER 5

DISCUSSION

Mapping word to world – pairing speech units with event units – requires children to segment speech and event streams into reliable units. While speech segmentation has been linked to word learning (for reviews, see Romberg & Saffran, 2010; Saffran, 2014), the problem of event individuation for lexical acquisition has remained unexplored in research. If anything, studies of word learning in children largely take for granted the challenge of individuating relational referents, presenting children with novel labels alongside videos of pre-packaged action units (e.g., Maguire et al., 2010; Waxman et al., 2009). Yet, the referents of many of the words forming children's growing vocabularies do not come in neat pre-individuated packages (Gentner & Boroditsky, 2001; Maguire et al., 2006). Relational terms, including verbs, are challenging for young word learners to acquire relative to concrete nouns (e.g., Bornstein et al., 2004). This dissertation takes the lead from speech perception research and evaluates the untested assumption that a spontaneous cognitive process of event segmentation plays a role in vocabulary acquisition.

In particular, statistical action segmentation – parsing streams of action based on movement regularities – is a potent candidate mechanism for explaining how children find meaningful units within the continuous and novel flow of events. In this dissertation, we ask (1) whether individual differences in statistical action segmentation abilities relate to children's verb and overall vocabulary knowledge and (2) whether action segmentation facilitates verb learning. We hypothesized, based on prior literature, that children who were better able to statistically segment novel streams of human action would have greater verb and overall vocabulary knowledge (Research Question 1). Further, we predicted that children's learning of novel verbs would be facilitated by statistical action segmentation, with children better able to map verbs onto statistically consistent action units than less statistically consistent part-units, and non-statistically consistent non-units (Research Question 2). As discussed below, the study supported some of our hypotheses, while also presenting some unexpected results.

Research Question 1: Is Action Segmentation Linked with Vocabulary Knowledge?

Our first question addresses the proposed link between action segmentation and lexical development. More specifically, we asked whether individual differences in 3-year-olds' statistical action segmentation ability would explain unique variance in their verb and overall vocabulary knowledge, assessed in terms of accuracy and reaction time. We predicted that, if children's statistical segmentation of novel actions contributes to their vocabulary acquisition, then action segmentation ability should make an independent contribution to vocabulary knowledge, controlling for age and nonverbal intelligence. We additionally tested whether this link holds after controlling for other aspects of children's linguistic knowledge, namely their knowledge of grammatical structure (syntax) and their ability to use particular linguistic strategies to acquire novel words and structures (process). Both syntax knowledge and language process are important for word learning (for review, see Levine et al., in press), but, in theory the contribution of these linguistic skills to word learning should be independent from the contribution of the non-linguistic cognitive process of action segmentation. This question concerning a link between action segmentation and vocabulary knowledge has not previously been examined in the literature.

Hierarchical regression analyses indicated that individual differences in 3-year-olds' statistical action segmentation abilities predicted overall vocabulary accuracy and reaction time, as well as their verb vocabulary reaction time when controlling for age and nonverbal intelligence. Children who looked longer, on average, to statistically coherent action units relative to non-units had more sophisticated overall vocabularies on the QUILS and were faster to select correct vocabulary referents on the QUILS and VCT. This finding supports our hypothesis of a link between action segmentation ability and vocabulary knowledge, and supports the theory that individuation of word referents matters for word learning (Gentner & Boroditsky, 2001; Maguire et al., 2006). Moreover, individual differences in action segmentation ability predicted all three aforementioned measures of vocabulary even when children's knowledge of syntax and their ability to use language-learning strategies (assessed using QUILS) were controlled, providing further support for the idea that action segmentation is a distinct cognitive process linked to vocabulary development.

Unexpectedly, however, no relation was found between action segmentation ability and children's accuracy on the VCT. Broadly, there are two possible explanations, one methodological and one theoretical, for why we found action segmentation ability was linked with QUILS Vocabulary accuracy but not VCT verb vocabulary accuracy. First, while the distribution for 3-year-olds' QUILS Vocabulary accuracy was roughly normally distributed, their VCT accuracy scores showed a ceiling effect. Indeed, while the QUILS assessment was designed for 3- through 5-year-olds and has been tested with hundreds of children establishing reliability and validity (Golinkoff et al., 2017; Levine et al., 2015), the VCT was originally designed for slightly younger children (27- to 33-month-olds) and had not been tested or validated with 3-year-olds. It is possible that the skewness and kurtosis in this distribution reduced the probability that correlations with other variables such as action segmentation ability would be produced; however, VCT verb accuracy was significantly correlated with other language measures and with age and working memory, making this explanation less plausible.

A second possible explanation for these disparate findings is that the link with overall vocabulary is robust because it takes into account words that vary substantially in how they are individuated. That is, the QUILS Vocabulary component assesses words that vary greatly in terms of their ease of individuation, from concrete nouns such as ‘sailor,’ to relational nouns such as ‘doorknob,’ to action verbs such as ‘unlocking,’ to spatial prepositions such as ‘between,’ to conjunctions such as ‘because.’ In contrast, the VCT tested *only* action verbs (which may not vary substantially in how they are individuated), and the variability in children’s performance on the VCT may be explained more by individual differences in factors specific to verb learning (e.g., Konishi et al., 2016). However, we cannot fully disentangle these possibilities with the present research.

Importantly, our finding that action segmentation is linked with overall vocabulary (but not necessarily verb vocabulary) contrasts with a prior study linking non-linguistic action categorization uniquely to verb vocabulary (Konishi et al., 2016), while aligning with research linking understanding of action structure to overall vocabulary (Kaduk et al., 2016). Konishi and colleagues (2016) found that the non-linguistic ability of infants to form invariant categories of action’s path and manner – such as forming a category of *jogging* after seeing this manner occur *around*, *through*, and *behind* a tent – was uniquely predictive of children's later verb vocabulary, but not overall vocabulary. However, research by Kaduk and colleagues (2016) showed that the non-linguistic ability of infants to detect unexpected outcomes of familiar events – such as a man bringing a pretzel to his ear – was linked to concurrent and later overall vocabulary size. To succeed in this latter non-linguistic task, children need to have learned via experience and perhaps partially through statistical action segmentation that food is typically brought to the mouth rather than other parts of the head. While Konishi and colleagues' (2016) manner/path categorization task may be tapping into a process unique to verb learning, this dissertation and Kaduk and colleagues' (2016) study may be tapping into a common cognitive process of finding reliable structure in action, serving as the foundation for mapping words onto objects, actions, and events.

It must be noted that children's accuracy on the Segmentation Task was at chance for high- and low-frequency units, as well as part-units. This pattern should not be taken as an indication that children failed the task, because their performance was moderately consistent across the two segmentation sequences. Rather, their chance performance at the group level may be an indication of the difficulty of this task. While we attempted to ease task demands by keeping the first action of each side-by-side triad the same and presenting action units more slowly in the Segmentation Test Phase than the Statistical Sequence Phase, the complexity of viewing two side-by-side action triads may have necessitated significant looks to both sides before children could make an informed selection. Future analyses should examine children's segmentation accuracy separately for the first and second portions of each test trial to determine whether children's segmentation performance for the latter half of the test trials is above chance. Importantly, although group performance was at chance, there were substantial individual differences in our measure of segmentation accuracy, which provided the basis for examining relations between action segmentation ability and language.

Our results provide compelling evidence for our hypothesized link between action segmentation ability and lexical knowledge, both in terms of what words children know and how quickly they can respond to these known words. However, links between action segmentation and overall vocabulary were more evident than links with verb knowledge alone. Future research will be needed to evaluate the exact ways in which statistical action segmentation may contribute to word learning, and Research Question 2 is a first step for this research.

Research Question 2: Does Action Segmentation Facilitate Verb Learning?

Our second question examined experimentally whether statistical action segmentation facilitated children's learning of novel verbs for action units. We hypothesized that children would demonstrate greater learning of verbs for statistically consistent action units compared to statistically inconsistent action units, and further predicted effects of the type of inconsistent action unit, with greater learning of verbs for part-units than non-units. Following the Statistical Sequence Phase, in which children viewed a statistical action sequence, we utilized a stringent test of verb learning, assessing children's ability to extend a trained verb to a novel exemplar of the action unit (Extension Trial), to look away from that action unit when a new verb was presented (New Verb Trial), and to renew looking toward that action unit when the trained verb was presented again (Recovery Trial). We also compared children's visual attention toward the correct referent on those test trials to chance levels, as a further test of word learning.

Contrary to predictions, there was no evidence that children learned words for statistically consistent action units. Rather, the mixed design ANOVA revealed a significant interaction, with children successfully learning words for statistically inconsistent action units (i.e., part-units and non-units), and doing so only when this verb was learned second. Verb learning was evidenced by the expected quadratic pattern of decreased looking to the matching action from the Extension Trial to the New Verb Trial, and renewed looking to the matching action at the Recovery Trial. However, when comparing accuracy to chance levels, findings for part-units and non-units diverged. Extension, Recovery, and New Verb Trials were all at chance for non-units, but for part-units, the predicted patterns of greater-than-chance looking during Extension and Recovery Trials were trending toward significance, and looking during the New Verb Trial was at chance. Thus, while we predicted verb learning would be greatest for statistically consistent action units and lowest for non-units, with part-units somewhere in between, results suggest that children's verb learning was strongest for part-units, less robust for non-units, and not apparent for statistically consistent action units.

Children's greater learning of verbs for action part-units than non-units supports our hypothesis that statistical segmentation facilitates verb learning. This finding is consistent with research demonstrating that statistical speech segmentation facilitates the mapping of part-word labels onto objects, compared to no prior segmentation opportunities (Graf Estes, 2012). However, in that research, speech segmentation also facilitated children's mapping of statistically consistent word labels onto objects (Graf

Estes, 2012), whereas here no facilitation effect was found for statistically consistent action units. Further, other speech segmentation studies have found facilitation effects of segmentation for statistically consistent syllabic units *but not* part-units or non-units (Graf Estes et al., 2007; Hay et al., 2011). Yet, two prior studies assessing statistical action segmentation in infants found conflicting attentional biases: one favoring statistically consistent action units (Roseberry et al., 2011), and the other favoring part-units (Stahl et al., 2014). Moreover, no prior studies have examined word learning facilitation effects of statistical *action* segmentation on word learning, so there is no direct comparison.

One possibility for these unexpected findings is that children mapped verb labels onto individual actions rather than statistically consistent and inconsistent triadic action units. Despite the statistical nature of the segmentation sequence and the statistical learning demonstrated in research question one, it may be that the ambiguity of the referent, given that the actor always returned to a common body position between actions, led children to map verb labels onto individual actions rather than the triadic action unit. If that is the case, the findings may reflect word learning facilitation based on frequency rather than transitional probability. That is, the individual actions comprising part- and non-unit triads were those actions that comprised the high-frequency unit triads, appearing twice as often in the segmentation sequence as the individual actions comprising the statistically consistent (i.e., low-frequency) action unit triads, and this increased frequency may have played a role in facilitating word learning. Indeed, frequency of object experience contributes to the acquisition of nouns (Clerkin, Hart, Rehg, Yu, & Smith, 2016), and it would not be surprising that the frequency of action experience would play a role in verb learning. However, the goal of this study was for children to map verbs onto action unit triads, the frequency of which was equal for low-frequency units and part-units, and was non-existent for non-units, which never appeared (in that order) in the sequence. The finding that children learned words for non-units and not low-frequency units suggests that at least some children may have mapped labels onto individual actions rather than triadic action units, and unfortunately, without follow-up research, it is not possible to determine exactly what children perceived as the verb's referent during the AUL Phase.

But this explanation alone would not fully explain the results, because of the superior verb learning demonstrated for action part-units relative to non-units. The individual actions comprising part- and non-units were identical, and the only difference between these two units was the transitional probabilities between actions (i.e., 0.5 and 1.0 for part-units, 0.0 and 0.0 for non-units). Thus, this result suggests an effect of transitional probability on children's verb learning in addition to the effect of action frequency.

The effect of frequency seems to be stronger than the effect of transitional probability, because children failed to learn verbs for low-frequency units at all. However, unlike the low-frequency action unit triad, the part-unit triad contained within it an action dyad that was both high transitional probability (TP of 1.0) and high frequency. Verb learning may be accomplished most effectively when a verb referent has high internal reliability and is also experienced with high frequency. This possibility is intriguing, but would need to be tested directly in future research.

Unlike studies of speech perception, which unambiguously show that statistical speech segmentation facilitates word learning (Graf Estes, 2012; Graf Estes et al., 2007; Hay et al., 2011), the finding from this dissertation shows that facilitation of word learning from statistical action segmentation is a comparatively weak and equivocal effect. This distinction may reflect the realities of word learning outside the laboratory. There is flexibility in how relational referents such as the referents of verbs are packaged for language (Gentner & Boroditsky, 2001; Goksun et al., 2010; Tomasello & Brandt, 2009), and children, accustomed to this variability, may not find the absence of statistical coherency in action units as a sufficient deterrent for mapping novel labels. For example, when children are learning the verb ‘slide,’ they may initially map this label onto the action sequence of a child climbing and then gliding down a sliding board; but this word mapping must be flexible, as children will learn that the word applies more generally to movement along a smooth surface. In contrast, statistical coherency in speech units may act as a stronger constraint on word learning, because there is a finite “inventory of sounds (the phonemes)” in any given language (Aslin, 2017) and the string of syllables that form a word is a fixed unit that is a constant across all instances and uses of the word. Given this logic, it would not be surprising if statistical speech segmentation plays a more direct and less ambiguous role in facilitating word learning (Graf Estes, 2012; Graf Estes et al., 2007; Hay et al., 2011) than does statistical action segmentation.

Future Directions

This dissertation found robust links between children’s statistical action segmentation abilities and their vocabulary knowledge, yet, links with verb vocabulary and facilitation of verb learning were only weakly supported. We suggest that event segmentation may be a critical cognitive process for lexical acquisition; however, there are alternative explanations for the demonstrated association between action segmentation ability and vocabulary knowledge. First, it is possible the link is explained by effects of language on event segmentation. That is, children with larger vocabularies may be better able to parse events because of their greater ability to describe the events they are viewing. Indeed, Zacks, Tversky, and Iyer (2001) showed that adults' event segmentation is improved when they verbally described the events they were parsing, compared to segmenting the events silently. If language is also used, albeit to a lesser extent, when parsing events silently, then larger vocabularies may lead to improved event segmentation. Another possible explanation for the link is that a third factor such as previous event experience or event knowledge may affect both vocabulary knowledge and the capacity to segment events. In fact, adults with more expertise in a particular domain do demonstrate superior event segmentation within that domain (Bläsing, 2015; Levine, Hirsh-Pasek, Pace, & Golinkoff, 2017). Clearly, there may be complex pathways between event experience, event segmentation ability, and vocabulary knowledge, and future research would be necessary to truly bolster the claim that event segmentation supports lexical acquisition. Specifically, this claim requires support from (1) longitudinal research, examining the contribution of action segmentation to vocabulary knowledge controlling for earlier vocabulary knowledge, and (2) experimental research, following up on our second research question, to determine the ways in which action segmentation may facilitate word learning.

To further our understanding of how statistical action segmentation may contribute to lexical acquisition, we need to probe more deeply what cognitive process or processes are being represented by children's action segmentation performance. Do statistical action segmentation abilities reflect just this process as described, or do they reflect visual statistical segmentation abilities more generally, or, even more broadly, statistical segmentation abilities for all continuous input including the speech stream? Mounting evidence, albeit from adult studies, suggests that statistical learning is not a uniform ability; that is, statistical learning ability is reliable within individuals, within *but not across* modalities and domains (Conway & Christiansen, 2005; Siegelman, Bogaerts, Christiansen, & Frost, 2017; Siegelman & Frost, 2015). However, the challenges of word learning theoretically seem to require three statistical learning processes: speech segmentation to unitize novel speech streams, action segmentation to unitize the stream of events, and cross-situational learning to learn word-to-world mappings (Levine et al., in press). Future research should evaluate how these three processes may make independent contributions and work together to promote word learning.

If future research confirms the independent contribution of event segmentation to lexical development, the implications of this relation will warrant further attention. For example, just as speech segmentation, which contributes to language acquisition (Evans et al., 2009; Junge et al., 2012; Kooijman et al., 2013; Newman et al., 2006, 2016; Singh et al., 2012; Swingley, 2005) is scaffolded by adults' use of child-directed speech (as opposed to adult-directed speech) (Johnson & Jusczyk, 2001; Kemler Nelson et al., 1989; Monaghan & Mattock, 2012; Shukla et al., 2007; Thiessen et al., 2005; Yurovsky et al., 2012), there is some evidence that children's action segmentation could be scaffolded by adults’ use of child-directed action, which highlights action boundaries using eye gaze, action repetition, and exaggerated movements (Brand, Baldwin, & Ashburn, 2002; Brand, Hollenbeck, & Kominsky, 2013; Brand & Shallcross, 2008; Williamson & Brand, 2014). Adults may be able to assist children with the task of individuating events into units, and in so doing support their lexical development.

Conclusions

Word learning requires children to uncover reliable mappings between speech units and event units. Although much research has focused on how children find reliable units in speech and how they figure out to what a novel word refers (for review, see Levine et al., in press), children also must, in theory, find the reliable, candidate referents of novel words from the flux and flow of continuous events. This dissertation evaluated whether finding reliably-patterned units of events might assist children with the challenges of word learning. The task of finding these non-linguistic units is purportedly more difficult for relational terms (Gentner & Boroditsky, 2001; Maguire et al., 2006), which become an increasingly large portion of children’s vocabularies across development (Fenson et al., 1994; Gentner & Boroditsky, 2009). We suggest that children may succeed at learning a range of complex relational terms in part through their ability to spontaneously segment events using the powerful mechanism of statistical learning.

The findings from this study largely confirm the hypothesized link between event segmentation and lexical acquisition. We found that 3-year-olds who were better able to statistically segment novel action sequences had more sophisticated overall vocabularies, but not specifically larger verb vocabularies, and were faster to select correct overall and verb vocabulary referents, controlling for age, nonverbal intelligence, and other linguistic knowledge and skills. We also found some experimental evidence for these links: although children failed to map verbs onto statistically consistent units, they showed greater verb learning for statistical action part-units than non-units, suggesting a facilitatory effect of statistical action segmentation on verb learning. These results not only build on recent work indicating links between action perception and vocabulary size (Konishi et al., 2016; Kaduk et al., 2016), but they also specifically begin to bridge the gap between children's segmentation of events and their growing knowledge of words which map onto segmented event units. This dissertation adds a unique contribution in revealing that children's automatic non-linguistic segmentation of ongoing events may assist in simplifying the complexities of word learning.

REFERENCES CITED

Arunachalam, S., & Waxman, S. R. (2011). Grammatical form and semantic context in verb learning. *Language Learning and Development*, *7*(3), 169-184.

Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*, e1373.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321-324.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382-1407.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708-717.

Bauer, P. J. (1992). Holding it all together: How enabling relations facilitate young children's event recall. *Cognitive Development*, *7*, 1-28.

Bauer, P. J., & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1-to 2-year-olds' recall of events. *Developmental Psychology*, *25*(2), 197-206.

Bläsing, B. E. (2014). Segmentation of dance movement: effects of expertise, visual familiarity, motor experience and music. *Frontiers in Psychology*, *5*.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., ... & Vyt, A. (2004). Cross-Linguistic Analysis of Vocabulary in Young Children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*(4), 1115-1139.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for ‘motionese’: modifications in mothers’ infant-directed action. *Developmental Science*, *5*(1), 72-83.

Brand, R. J., Hollenbeck, E., & Kominsky, J. F. (2013). Mothers’ infant-directed gaze during object demonstration highlights action boundaries and goals. *IEEE Transactions on Autonomous Mental Development*, *5*(3), 192-201.

Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, *11*(6), 853-861.

Brandone, A. C., Pence, K. L., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Action speaks louder than words: Young children differentially weight perceptual, social, and linguistic cues to learn verbs. *Child Development*, *78*(4), 1322-1342.

Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology*, *76*, 30-77.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research*, 51(6), 1569-1579.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2016, September). *How everyday visual experience prepares the way for learning object names.* Paper presented at 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics, Cergy-Pontoise, France. doi: 10.1109/DEVLRN.2016.7846803

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*(1), 24-39.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., ... & Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*, *48*(3), 541-553.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321-335.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i-185.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*(1), 203-222.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics*, *72*(6), 1522-1532.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822-15826.

Fisher, C., & Gleitman, L. R. (2002). Language acquisition. In H. F. Pashler (Series Ed.) & R. Gallistel (Vol. Ed.), *Steven's handbook of experimental psychology, Vol. 1: Learning and motivation* (3rd ed., pp. 445–496). New York, NY: Wiley.

François, C., Chobert, J., Besson, M., & Schön, D. (2012). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*(9), 2038-2043.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading Technical Report; no. 257*.

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

Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from Navajo. In V. Gathercole (Ed.), Routes to language: Studies in honor of Melissa Bowerman (pp. 5–36). New York, NY: Psychology Press.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684-691.

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

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). Trading spaces carving up events for learning language. *Perspectives on Psychological Science*, *5*(1), 33-42.

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. *Cognition*, *4*(2), 189-202.

Golinkoff, R. M., de Villiers, J. G., Hirsh-Pasek, K., Iglesias, A., & Wilson, M. (2017). *QUILS: Quick Interactive Language Screener*. Baltimore: Brookes.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23-45.

Golinkoff, R. M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child Development*, *67*(6), 3101-3119.

Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. *Frontiers in psychology, 3*.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*(3), 254-260.

Haman, E., Łuniewska, M., Hansen, P., Simonsen, H. G., Chiat, S., Bjekić, J., ... & Gagarina, N. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from cross-linguistic lexical tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, 1-26.

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93-106.

Hendrickson, K., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2017). Assessing a continuum of lexical–semantic knowledge in the second year of life: A multimodal approach. *Journal of Experimental Child Psychology*, *158*, 95-111.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... & Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the society for research in child development*, i-135.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, *79*(4), 979-1000.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of memory and language*, *44*(4), 548-567.

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science, 15*(4), 463-473.

Kaduk, K., Bakker, M., Juvrud, J., Gredebäck, G., Westermann, G., Lunn, J., & Reid, V. M. (2016). Semantic processing of actions at 9 months is linked to language proficiency at 9 and 18 months. *Journal of Experimental Child Psychology, 151*, 96-108.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55-68.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559-1571.

Konishi, H., Stahl, A. E., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Individual differences in nonlinguistic event categorization predict later motion verb comprehension. *Journal of Experimental Child Psychology, 151*, 18-32.

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology, 4*.

Kuhl, P. K. (2007). Cracking the speech code: How infants learn language. *Acoustical Science and Technology*, *28*(2), 71-83.

Levine, D., Hirsh-Pasek, K., Pace, A., & Michnick Golinkoff, R. (2017). A goal bias in action: The boundaries adults perceive in events align with sites of actor intent. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 916-927.

Levine, D., Pace, A., Yust, P., Brezack, N., Morini, G., Tabori, A., … & Hirsh-Pasek, K. (2015, October). *Developing a computer-administered language assessment for diverse children ages 3 to 6*. Presented at Division for Early Childhood's Annual International Conference on Young Children with Special Needs and Their Families, Atlanta, GA.

Levine, D., Strother-Garcia, K., Hirsh-Pasek, K., & Golinkoff, R. M. (In press). Names for things…and actions and events: Following in the footsteps of Roger Brown. In E. M. Fernandez & H. S. Cairns (Eds.) *Handbook of Psycholinguistics*. New York, NY: Wiley-Blackwell.

MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, *94*, 67-84.

Maguire, M. J., Brumberg, J., Ennis, M., & Shipley, T. F. (2011). Similarities in object and event segmentation: A geometric approach to event path segmentation. *Spatial Cognition & Computation*, *11*(3), 254-279.

Maguire, M., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A unified theory of word learning: Putting verb acquisition in context. *Action meets word: How children learn verbs*, 364-391.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., Imai, M., Haryu, E., Vanegas, S., ... & Sanchez-Davis, B. (2010). A developmental shift from similar to language-specific strategies in verb acquisition: A comparison of English, Spanish, and Japanese. *Cognition*, *114*(3), 299-319.

Majid, A., Jordan, F., & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences*, *49*, 1-18.

Malt, B. C., Ameel, E. Imai, M., Gennari, S. P., Saji, N., & Majid, A. (2014). Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language, 74*, 107-123.

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science*, *12*(6), 1020-1025.

Markman, E. M. (1991). Categorization and naming in children: Problems of induction. Cambridge, MA: MIT Press.

Mayo, J., & Eigsti, I. M. (2012). Brief report: A comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. *Journal of Autism and Developmental Disorders, 42*(11), 2476-2485.

McCabe, P. C., & Meller, P. J. (2004). The relationship between language and social competence: How language impairment affects social growth. *Psychology in the Schools, 41*(3), 313-321.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1), 271-280.

Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word–referent mappings. *Cognition*, *123*(1), 133-143.

Monroy, C. D., Gerson, S. A., Domínguez-Martínez, E., Kaduk, K., Hunnius, S., & Reid, V. (in press). Sensitivity to structure in action sequences: An infant event-related potential study. *Neuropsychologia*.

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017a). Toddlers’ action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, *157*, 14-28.

Monroy, C., Meyer, M., Gerson, S., & Hunnius, S. (2017b). Statistical learning in social action contexts. *PloS one, 12*(5), e0177261.

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, *42*(4), 643-655.

Newman, R. S., Rowe, M. L., & Ratner, N. B. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*(5), 1158-1173.

Newtson, D., Engquist, G. A., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35*(12), 847-862.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology*, *69*(12), 2390-2401.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674-685.

Poulin-Dubois, D., & Forbes, J. N. (2002). Toddlers' attention to intentions-in-action in learning novel action words. *Developmental Psychology*, *38*(1), 104-114.

Pruden, S. M., Roseberry, S., Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Infant categorization of path relations during dynamic events. *Child Development*, *84*(1), 331-345.

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

Raiford, S. E., & Coalson, D. L. (2014). *Essentials of WPPSI-IV assessment*. Hoboken, NJ: John Wiley & Sons.

Reid, V. M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: evidence for semantic processing systems. *Developmental Psychology*, *45*(3), 620-629.

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613-643.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956-970.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video?. *Child Development*, *80*(5), 1360-1375.

Roseberry, S., Richie, R., Hirsh-Pasek, K., Golinkoff, R. M., & Shipley, T. F. (2011). Babies catch a break 7-to 9-month-olds track statistical probabilities in continuous dynamic events. *Psychological Science*, *22*(11), 1422-1424.

Saffran, J. (2014). Sounds and meanings working together: Word learning as a collaborative effort. *Language Learning*, *64*(s2), 106-120.

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

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*(2), 101-105.

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163-180.

Seston, R., Golinkoff, R. M., Ma, W., & Hirsh-Pasek, K. (2009). Vacuuming with my mouth?: Children's ability to comprehend novel extensions of familiar verbs. *Cognitive Development*, *24*(2), 113-124.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1-32.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B*, *372*, 1711.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105-120.

Singh, L., Reznick, J. S., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*(4), 482-495.

Slone, L. K., & Johnson, S. P. (2015). Infants' statistical learning: 2-and 5-month-olds' segmentation of continuous visual sequences. *Journal of Experimental Child Psychology*, *133*, 47-56.

Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480-498.

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

Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, *9*(1), 25-49.

Snowling, M. J., Bishop, D. V. M., Stothard, S. E., Chipchase, B., & Kaplan, C. (2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *Journal of Child Psychology and Psychiatry, 47*(8), 759-765.

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*(5), 1821-1826.

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395-411.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86-132.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71.

Tomasello, M., & Brandt, S. (2009). Flexibility in the semantics and syntax of children's early verb use. *Monographs of the Society for Research in Child Development*, *74*(2), 113-126.

Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, *19*, 311-333.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34.

Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, *93*, 217-230.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375-382.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611.

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H. J. (2013). Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives*, *7*(3), 155-159.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67-95.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence – fourth edition technical manual and interpretive manual.* San Antonio, TX: Psychological Corporation.

Wenner, J. A., & Bauer, P. J. (1999). Bringing order to the arbitrary: One-to two-year-olds' recall of event sequences. *Infant Behavior and Development*, *22*(4), 585-590.

Williamson, R. A., & Brand, R. J. (2014). Child-directed action promotes 2-year-olds' imitation. *Journal of Experimental Child Psychology*, *118*, 119-126.

Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry*, *43*(5), 635-645.

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

Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: scaffolding from child-directed speech. *Frontiers in Psychology*, *3*.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979-1008.

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*(2), 201-216.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29-58.

APPENDIX A

STATIC DEPICTIONS OF EACH ACTION IN THE THREE SEGMENTATION SEQUENCES

Stretch Out

Touch Ear

Side Step

Bend Arms

Punch Right

Reach Over

“L” Stretch

Crouch

Reach Across

Punch-Kick

Alligator Arms

Stretch Up

Raise Arm/Leg

Reach Under

Stretch Down

Align Hands

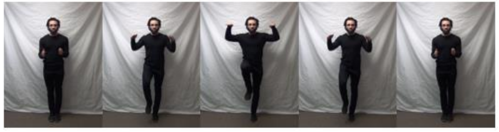
Bend Knees

Diagonal Stretch

Touch Arm

Cover Eyes

Bend Up Arms/Leg



Foot Touch

Cross Arms

Fold Arms

Arch

Knee Touch

Twist

Arm Dangle

Parallel Hands

Diagonal Arms

Elbow Touch

Shoulder Touch

Puppet Hand

Clap

Hands Out

Body Cross

APPENDIX B

WORD PAIRS TESTED IN THE VERB COMPREHENSION TASK

Intransitive verb pair		
Run	Jump	
Stretch	Clap	
March	Spin	
Dance	Cry	

Transitive verb pair		*Direct object*
Feed	Hug	Cookie Monster
Pour	Drink	Juice
Shake	Open	Gift
Read	Rip	Newspaper
Roll	Bounce	Basketball
Lift	Pull	Pooh bear
Drop	Bite	Ice cream
Kiss	Tickle	Teddy bear
Squeeze	Blow	Balloon
Kick	Throw	Balloon
Lick	Break	Lollipop
Wash	Rock	Baby
Cut	Tie	Ribbon
Eat	Push	Cake

Noun pair (practice trials)		
Cookie	Banana	
Goldfish	Donut	
Firetruck	Bird	
Orange	Plane	
Squirrel	Grapes	
Rocket	Giraffe	
Crab	Pancakes	
Elephant	Carrot	
Duck	Chocolate	
Tiger	Burger	
Corn	Hat	
Zebra	Strawberry	

Note: During test, intransitive and transitive verb pairs and noun pairs were interspersed.