The Associative Learning Model versus the Hypothesis Testing Model: Using Error Patterns to Understand Preschoolers' Word Learning

Hannah Nye

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2017

Committee:

Amy Pace

Patricia Dowden

Sara Kover

John Thorne

Program Authorized to Offer Degree:

Speech & Hearing Sciences

University of Washington

Abstract

The Associative Learning Model versus the Hypothesis Testing Model:

Using Error Patterns to Understand Preschoolers' Word Learning

Hannah Nye

Chair of the Supervisory Committee:

Dr. Amy Pace, Ph.D., Assistant Professor

Speech & Hearing Sciences

A key question in childhood language development is how children are able to learn new words in spite of ambiguous real-world situations. Two potential explanations have dominated the literature to date. The classic model suggests that children catalog multiple possible referents while the competing model theorizes that children instead fixate on one possible referent. This research compares and contrasts these models by analyzing preschoolers' selection accuracy after they've made an error during a word-learning task. Findings suggest that neither model fully captures the complex problem of word learning during disambiguation tasks where linguistic cues can influence children's learning. Instead, the results found here may, in fact, be best explained more complex models of learning that incorporate elements from both models.

Learning new words is a truly remarkable feat that young children exhibit repeatedly on a daily basis. By the third year of life, children utilize a plethora of cues when assigning referents to novel words. For instance, they capitalize on the context (Carey, 1978), the speaker's behavioral and attentional patterns (Baldwin, 1991; Nappa, Wessel, McEldoon, Gleitman, & Trueswell, 2009; Tomasello & Farrar, 1986; L. Smith, 2000), various socio-pragmatic cues (Baldwin, 1993; Tomasello, 2000), and the linguistic and syntactic constraints surrounding a novel word (Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005). Although some word-learning strategies are more applicable for certain word types (e.g., children rely on syntactic cues more heavily for verbs) (Gleitman, 1990), they are all readily available to assist the typically developing child in correctly assigning new words to unclear referents.

Access to these many cues while learning new words may significantly reduce referential ambiguity, but it seems unlikely that it resolves all uncertainty (Quine, 1960). Consider, for example, a birthday party vignette. Adults and children across a span of ages are conversing in the same area at the same time. There may be music playing in the background. Visual stimuli takes the chaotic form of decorations, presents, snacks, desserts, and games. Suppose, in a situation like this, a mother says to her 3-year-old son, "Look at the pretty ribbon!" Despite the aforementioned word-learning heuristics at this child's disposal, some ambiguity is likely to remain. Without an explicit gestural cue, the child could reasonably assume that his mother is referring to the nearby birthday cake, or perhaps the wrapping paper on the gift, or even the gift in its entirety. It's even plausible for him to assume that she's speaking generally about the scene as a whole, or, conversely, some semantically unrelated detail such as an insect she noticed.

So how is it that children overcome these obstacles to learning new words? This thesis will consider two prominent word-learning models to address some of the unanswered questions regarding how children resolve referential ambiguity (also called indeterminacy). The remainder of the introduction will provide a brief review of the literature regarding indeterminacy, fast-mapping, and word retention and extension. Following this, it will present the theoretical and empirical background for each of the word-learning models, the related linguistic phenomena that support or contradict these models, and the need for continued research.

Referential Indeterminacy

Some researchers have posited that children simply ignore novel information in ambiguous settings and wait for more explicit instruction on a word's meaning; that is, word learning does not happen in such chaotic settings (Brent & Siskind, 2001). Others have argued that, even if a child does correctly identify the intended referent, this may not lead to word retention. From this perspective, mere disambiguation is not true word learning (Bion, Borovsky, & Fernald, 2013). However, considering that the average parent directs 300-400 utterances per hour to their children (Hart & Risley, 1995) and that a typical 6-year-old's expressive vocabulary is approximately 2,600 words and receptive vocabulary is 20,000-24,000 words (Owens, 1996; Stahl, 1999), it seems unlikely that children have the luxury of forgoing word learning in opaque settings. Indeterminacy is very likely an everyday occurrence; it seems that children must resolve it – to at least some extent – in order to acquire vocabulary as rapidly as they do.

Nevertheless, some researchers have maintained that referential ambiguity may actually be quite rare in children's daily word-learning experiences (Yurovsky, L. Smith, & Yu, 2013). Considering the multitude of strategies children have at their disposal to resolve indeterminacy, it is possible that there will, eventually, always be sufficient cues to identify the correct referent for each novel word. For example, during the collection of a corpus of parent-child dyad interactions recorded from the child's eye-level perspective, these researchers found that some naming events are highly informative (e.g., the labeled toy takes up much of the child's view) while others are more ambiguous (e.g., other items may be in view, a distractor item may take up much of the view). However, aggregating across instances resulted in high accuracy of correct referent identification by adult labelers (Yurovsky et al., 2013). These findings suggest children may rely more heavily on highly informative learning events and discount, or ignore, the ambiguous ones.

Fast-mapping

Whether referential ambiguity is common or rare in real-world settings and whether children use indeterminate information to learn actual words, children irrefutably possess the ability to resolve it. In fact, they are notably adept at completing disambiguation tasks using a process called "fast-mapping." Dozens of experimental studies have shown that when children are presented with a set of objects, only one of which is novel, and also presented with a novel label, they typically assign it to the unfamiliar object (Carey & Bartlett, 1978; Clark, 1990; Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Markman, 1991; Markman & Wachtel, 1988; Merriman & Bowman, 1989). In laboratory settings, children recognize that the novel object must be assigned to the novel label based on mutual exclusivity; that is, all the other targets present already have known labels.

Fast-mapping is a skill that can be measured behaviorally as early as 16- to 18-months-old and is increasingly refined throughout toddlerhood. (Byers-Heinlein & Werker, 2009; Halberda, 2003; Markman, Wasow, & Hansen, 2003; Mervis & Bertrand, 1994). This skill also continues to develop and support children's word learning through the preschool years (Dollahan, 1985). More importantly, children's ability to fast map may predict their language outcomes in middle childhood. Rajan et al. (in press) found that novel word-object association ability at 21-months uniquely accounted for 22% of the variance in childhood receptive vocabulary five to eight years later, and that this variance was independent of any difference in general intellectual functioning. Thus, it is reasonable to conclude that fast-mapping skills are of great import, and that understanding how children accomplish this feat would provide valuable insight into how children learn words.

Retention and Extension

As important as fast-mapping is, it is only one step in the word-learning process. After children decide which referent a novel word refers to, they must then use this same word-referent pairing in other contexts as well (Barrett, 1999). Accurate referent selection does not automatically lead to word learning and retention; for example, Horst and Samuelson (2008) found that children who accurately fast-mapped novel labels did not reduplicate their selections on retention trials conducted just five minutes later. To learn the word "dog" accurately and completely, a child must recognize that the label refers not only to the family dog, but also to other members of the canine species. This extension to new settings, new referents, and new contexts, is required in varying degrees for all words. With this in mind, word-learning theories must consider both the initial exposure and subsequent trials to accurately represent how a child may learn words. Thus, learning a novel word is, at the very least, a two-step process: first, disambiguation, second, extension.

The Models

In response to the research on referential ambiguity and fast-mapping, two preeminent theories have emerged to account for children's word-learning ability in ambiguous settings. The traditional theory is known as the associative learning model; it is a gradual view of word learning. In this model, children require repeated exposures to novel words before they settle on a single meaning; during the learning process they accept numerous meanings (with evolving weights) until some learning threshold is reached, at which point all competing understandings are discarded (Akhtar & Montague, 1999; Roembke & McMurray, 2016; Scott & Fisher, 2012; K. Smith, A. Smith, & Blythe, 2011; L. Smith, 2000; L. Smith & Yu, 2008; Yu & L. Smith, 2011). Alternatively, the more recent hypothesis testing model suggests that children maintain only *one* hypothetical word-referent pairing at a time, which is either verified or discarded as subsequent exposures align or conflict. This model is sometimes referred to as "propose-but-verify" (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013; Woodard, Gleitman, & Trueswell, 2016). Both are discussed in greater depth below.

The Associative Learning Model (ALM)

The associative learning model (ALM) endeavors to explain how children collect information across multiple exposures to words. In this model, when children are exposed to a novel label and are unsure of the correct referent, they store several possible meanings for the unknown word. As subsequent exposures occur, children implicitly compute the statistical likelihood of various word-referent pairings until some learning threshold is reached and they settle on a definite referent for the word. This model has also been referred to as cross-situational statistics, cross-situational word learning, or statistical word-referent learning. In this thesis, associative learning is not meant to encompass all instances of possible associative learning in childhood development, but rather specifically refer to the theory of children holding multiple word-referent associations at fluctuating weights until they settle on a decision.

The ALM was originally derived from John Locke's (1690) writing "An Essay Concerning Human Understanding," which established empiricism as a viable explanation for human learning. This theory was more directly applied to language acquisition in the 1990's, when computational models of the ALM first became prevalent (Siskind, 1996). The earliest explanation of word learning focused on how children acquire nouns. L. Smith (2000) found that the nouns children learn early are predominantly count nouns of solid objects holding similar shapes (e.g. animals, foods, body parts). In light of this, she theorized that children build shape biases: they catalog each exposure to a particular noun and eventually settle on a particular set of shape-based inclusion and exclusion criteria for that label. This is an example of cross-situational associative learning at work. According to the ALM, children begin with relatively loose word-referent pairings that become more refined with each exposure (L. Smith & Yu, 2008).

Adult Studies and Computer Models

Over time, the plausibility of the ALM has been established through computer models and experiments with adults. Numerous computer simulations have supported the theory that novel words can be learned and organized into meaningful semantic networks using cross-situational associative learning, whereby information is carried across trials (Fazly, Alishahi, & Stevenson, 2010; Frank, Goodman, & Tenenbaum, 2009; Siskind, 1996; Vogt & A. Smith, 2005; Yu & Ballard, 2007; Yurovsky, Fricker, Yu, & L. Smith, 2014). Since human learning can be confounded by variables that do not affect computers, the ALM has also been demonstrated in word-learning tasks with adult participants (K. Smith et al., 2011; Yu & L. Smith, 2007). Yurovsky and colleagues (2014) demonstrated that prior exposure was an informative source for word learners. In this study, adults were exposed to two blocks of ambiguous word-learning trials; all adults received the same words and stimuli for block one, and then, in the second block, some participants received an entirely new set of words to learn, and some received some new words and some words that they had incorrectly learned from block one. They found that participants with partial knowledge from the first block of trials, even when they did not correctly learn those words, learned significantly more mappings in the second block than participants who received an entirely new set of words to learn. Even though knowledge of word-object pairings was below learning threshold in the first block, it was sufficient to inform pairings in the second block. This implies that knowledge of potential targets – accumulated over multiple exposures – increased word-learning power.

Another particularly well-designed, recent study considered whether learners maintain multiple hypotheses in parallel. In this experiment, participants attempted to learn eight novel word-referent pairs in the presence of both frequent and infrequent competitor targets; their task was to determine which novel object always occurred with the various words and correctly pair words and referents. Results showed that participants lingered their gaze longer on more frequently occurring competitor targets, even if they ultimately chose the correct target (Roembke & McMurray, 2016). This provides excellent evidence that learners are paying attention to competing referents and accumulating meaningful information across trials. According to this research, word learners entertain multiple hypotheses and utilize knowledge of past possible targets to inform their process.

Child Studies

Since the primary focus of developing the ALM is to understand how children (not adults or computers) learn novel words, research has been conducted on developing children as well. In one of the classic studies, 2-, 3-, and 4-year-olds were shown a novel object and taught a novel adjective. In subsequent trials, children extended the novel adjective (based on shape or texture, depending on their testing condition) to other exemplars, thus indicating they maintained information about word meaning across trials (Akhtar & Montague, 1999). However, since children were explicitly told which object the adjective applied to, they did not have to initially map the meaning, only extend it.

L. Smith and Yu (2008) examined whether or not children use associative learning to resolve referential ambiguity for mapping nouns. During training trials, two novel words and two novel shapes were presented to 12- to 14-month-old infants, with no spatial or temporal relationship to explain which label described which object. On subsequent test trials, only one word was spoken while two shapes were displayed: a shape that had been previously present with the same label (thus the target), and a shape that had not. Using manually-coded eye gaze as an outcome measure, results suggest that infants preferentially looked at the target, implying that they utilized information from ambiguous word-learning settings and retained it across trials. The authors more recently conducted a similar study using more advanced eye-tracking technology, and found results consistent with their previous, manually-coded eye gaze study; both studies supported the ALM (L. Smith & Yu, 2011). Additionally, similar patterns have been documented during verb learning tasks with 2.5-year-olds (Scott & Fisher, 2012). Thus, word learning patterns that are consistent with the ALM have been demonstrated across word classes in childhood lexical development.

Phenomena Supporting the ALM

There is other research investigating the learning mechanisms that may undergird the ALM; one prominent body of work focuses on the statistical learning perspective. Statistical learning refers to the ability to extract patterns from input in the environment and has been demonstrated in visual (Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014) as well as auditory development (Saffran, Aslin, & Newport, 1996). The quintessential example of statistical learning in the auditory domain is children's well-documented ability to segment continuous speech into distinct words using transitional probabilities (i.e., the likelihood that one sound follows another, based on phonotactic probability) to identify word boundaries. As early as 8-months-old, children are sensitive to probabilities in artificial languages and use this information to segment "words." They identify which sound combinations are statistically likely to be a word boundary in their native language and which are not (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). There is evidence that statistical learning actually facilitates a child's ability to learn the meanings of words (Estes, Evans, Alibali, & Saffran, 2007). This demonstrates that children are capable of utilizing statistical probability on large amounts of data. Such statistical prowess is necessary if children are using the ALM to learn words; they must be constantly analyzing multiple possible referents for novel words and deciding which is most statistically probable. Since children use this kind of statistical reasoning on phoneme clusters, it's plausible they do the same with novel word meanings.

A related body of literature exploring children's early expressive errors due to partial word knowledge can also be applied to support the ALM. When children are learning to use new words, they tend to overextend them (e.g. they call all round objects "ball") or underextend them (e.g. "dog" refers to their family dog, but not to other dogs) (Barrett, 1986; Barrett, 1996; Meints, Plunket, & Harris, 1999). These phases in children's lexical development provide evidence that word definitions are fluid and refined over time, not simply "accurate" or "inaccurate." This fluidity suggests that repeated exposures to the same word across contexts may serve to strengthen a child's representation of that lexical item, driving word learning in the manner hypothesized by the ALM.

Research Contradicting the ALM

More recent research has revealed some limitations with the research methodology in the ALM literature. When this model has been studied in both children and adults, novel words and referents have typically been introduced in ambiguous settings, and researchers documented whether the participant was able to learn the novel word after repeated ambiguous exposures. However, some researchers have pointed out that reporting aggregate measures of final outcomes indicates nothing about the strategy adopted across each trial (Medina et al., 2011; Trueswell et al., 2013). In other words, whether or not the participants were eventually able to learn the word in ambiguous settings does not indicate whether they utilized a refinement strategy such as the ALM or a guessing-then-dismissing strategy. Combining data across participants exaggerates this effect. Trueswell et al. (2013) notes, "If learning is characteristically abrupt, due to sudden insight on a single learning instance, and if these insights happen on different trials, across individuals, average performance will look gradual over that period of time" (p. 151). In order to truly understand which learning strategy is being adopted, analyses must consider trial-by-trial trends within individuals.

In addition to the limitations of the ALM research methodology, there are data that directly contradict the feasibility of this model. Some studies have shown, for example, that as referential ambiguity increases, adults' ability to use cross-situational strategies decreases (Medina et al., 2011; K. Smith et al., 2011). This poses a problem for the ALM, which claims to assist word learning in the *most* ambiguous settings. Similarly, the ALM proposes that learners refine and narrow their word definitions over time, but, in some instances, an associative strategy across trials caused adult learners to move toward an *inaccurate* definition that was more generic, in order to account for the variety of circumstances in which the word appeared. For example, when the target was supposed to be "ball," participants tended to move towards the more generic word "toy" (Gillette et al., 1999). Additionally, data has revealed that order of input matters: that is, words are easier to learn if they are introduced in high context settings before they are introduced in low context settings (Medina et al., 2011). If children are storing and utilizing all exposures to unlearned words as the ALM suggests, then theoretically order of those exposures should not impact their final accuracy. Taken together, these studies present significant complications for the ALM.

The Hypothesis Testing Model (HTM)

Gleitman's (1990) early work highlighted the ALM's shortcomings not long after its introduction, but it was over two decades before she and her colleagues presented an alternative: the hypothesis testing model (HTM), sometimes referred to as propose-but-verify. This model suggests that when children are faced with indeterminacy, they form *one* hypothesis about the accurate word-referent pairing, and then in subsequent exposures they either confirm this hypothesis or discard it.

The HTM originates from Gillette and colleagues' (1999) work using a now widespread research methodology known as the human simulation paradigm. In this experiment, adult participants watched video clips in which a mother interacted with her toddler. The audio was removed from these videos, and, in its place, was a beep tone which signaled when a target word was uttered; this was repeated across several scenes. The participant's job was to watch all the scenes and guess which word the tone stood for. This methodology has been implemented in numerous additional studies (Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina & Trueswell, 2013; Piccin & Waxman, 2007; Yurovsky et al., 2013), most notably Medina and colleagues (2011). This latter study came to the conclusion that the learners were holding a single hypothesis in mind and applying that very same hypothesis across subsequent trials unless it was disconfirmed. Trueswell, Gleitman and colleagues (2013), tested this model directly and then named it: propose-but-verify. Thus, a competing model to the ALM emerged.

Adult Studies and Computer Models

Since the HTM is relatively new, there are only a handful of studies considering this single-hypothesis approach. As aforementioned, Medina et al. (2011) found that if adult participants initially received a highly informative vignette, they were likely to guess correctly and maintain the hypothesis over time, especially after watching additional vignettes in which their guess was verified. This result seems intuitive. However, the authors also found that if participants guessed incorrectly after an initial event, they were less likely to be correct on subsequent guesses – thus implying that learners were not maintaining alternative targets from past vignettes. When the learners guessed incorrectly, they were very unlikely to guess correctly on the next trial (only 11% accuracy). Notably, their performance was nearly identical to that of other participants who viewed those same videos in isolation without cross-situational comparison (9% accuracy). This finding contradicts the ALM, suggesting that participants had little or no capacity to access alternative hypotheses that they had previously encountered. In fact, a false hypothesis, once formed, detracted from a learner's ability to be flexible in the face of disconfirming evidence.

Trueswell et al. (2013) used response patterns and eye-tracking with adult learners to assess whether participants were maintaining information explicitly or implicitly. This research found that when participants had been incorrect on a given trial, they were at chance accuracy for the next trial, thus implying that they did not explicitly reference alternative information from previous trials. They also found that eye-movements indicated no implicit memory from previous trials; participants were more likely to perseverate on the target only after a correct trial, never after an incorrect one. The authors repeated these results in highly simplified settings, with only two targets available across many trials, and they still found that participants were rarely, if ever, showing evidence of maintaining alternative hypotheses. Finally, they compared a computer simulation to their results from human participants and found a surprisingly similar pattern of learning, lending further credibility to the HTM.

Child Studies

To date, only one study has applied the HTM to child participants. Woodard et al. (2016) designed an "animal safari" experiment in which 2- and 3-year-olds were exposed to novel animals and nonsense words. They were shown two animals at a time during the first exposure, and then, in the second exposure, they were shown either the animal they had selected to go with the novel word ("same" condition), or else the animal that had co-occurred in the first exposure that they had not selected ("switch" condition). They found that children could remember a previously selected target and use it to inform later selections. However, the children did not show enough recollection of other potential targets to inform accuracy above chance. This data supports a single-hypothesis model of word learning in child learners. Unlike with the ALM, however, the HTM has yet to be studied on other word classes (i.e., verbs, adjectives) for child learners.

Phenomena Supporting the HTM

The HTM is further supported by evidence that children and adults may learn differently. Adult studies have been referenced as evidence for both the ALM and the HTM, but it is possible that such studies are actually uninformative when it comes to answering the question of how children learn words. Consider, for example, the game “20 Questions.” It has been documented that children under the age of seven tend to adopt a less efficient strategy than adults: they guess individual items (e.g., “is it a penguin?”) rather than asking broader strategic questions (e.g., “is it an animal?”) (Courage, 1989). Unlike adults, children are more likely to guess object by object than to rule out an entire category at once. This suggests that children may be prone to adopt a more local strategy, while adults may be more global learners. Additional evidence for this phenomenon is found in children’s response patterns to the aforementioned human simulation paradigm. Piccin and Waxman (2007) found that 7-year-olds required more guesses than adults in order to arrive at a target word, and even when they did arrive, they were significantly more likely to abandon their correct guess in subsequent trials. Taken together, these results imply that children are more likely to seek their best guess for each exposure, rather than integrating information across trials. This could suggest that children’s learning patterns are meaningfully different than those of adults, and should accordingly be studied separately.

Furthermore, word learning requires both fast-mapping and extension of the novel word, and there is introductory evidence in this arena that supports the HTM. Bion and colleagues (2013) specifically studied how disambiguation and retention are related. They found that young children (18-, 24-, and 30-month-olds) who spent a greater proportion of time looking at the novel object during fast-mapping training trials were better at retaining the word-referent link on subsequent test trials, independent of age or vocabulary size. Although this study doesn't consider the implications of these results as applied to word-learning models, a theoretical application supports the HTM. All participants were exposed to the same number of trials, so the ALM would predict relatively similar outcomes across participants based on exposure to the same targets. However, the HTM would predict that children who disambiguated more successfully on training trials (theoretically forming a stronger single hypothesis) would more successfully verify that hypothesis on subsequent trials, and the results align with this prediction.

Research Contradicting the HTM

One recent study specifically contradicts the theory behind the HTM. A key hallmark of this theory is the idea that learners consider only one hypothesis at any given time. Thus, if they have chosen the wrong hypothesis and are exposed to a situation that forces them to discard said hypothesis, they won't remember alternative possible referents: they will be at chance accuracy. Roembke and McMurray (2016) found that this is not necessarily the case as trials increase. In early trials, participants were roughly at chance (~33% accuracy between three targets) after an incorrect conjecture, just as the HTM predicts. However, by later trials, participants were significantly above chance (~50% accuracy), even after an incorrect guess. The likelihood of a correct answer after selecting an incorrect hypothesis on the preceding trial improves over time, which could not be the case with a pure HTM. The authors suggest that this pattern may be explained by considering the effect of a previous trial's performance as a *product* of learning, not a mechanism of it. However, it should be noted that this study was done with adult participants, and it has already been discussed how this may be a limiting factor.

The Present Study

Rationale

Previous research disproportionately represents adult subjects, which is problematic because of evidence that children may employ different learning mechanisms than adults. As a result, there is dire need for additional studies that include child subjects. Additionally, the few studies that do include children have small sample sizes. In the ALM literature, the largest child sample size is 72 subjects (Akhtar & Montage, 1999); for the HTM, a single study has been conducted with children, and it included only 32 children (Woodard et al., 2016). It is essential to consider the skills of fast-mapping and word learning in populations in infancy, toddlerhood, and preschool, when these skills are most functional. Larger sample sizes with child participants are needed to gain a better appreciation of this complex learning mechanism during development.

The existing research, with the exception of human simulation paradigm studies, typically involves a word-learning design in which the participant has no cues as to which novel word is associated with which novel referent during the exposure trial; rather, the word-referent association must be learned over time (Roembke & McMurray, 2016; Scott & Fisher, 2012; K. Smith et al., 2011; L. Smith & Yu, 2008; Trueswell et al., 2013; Woodard et al., 2016; Yu & L. Smith, 2007; Yu & L. Smith, 2011; Yurovsky et al., 2014). While this approach results in fewer variables (and there is thus more control over the data), it does not approximate children's real-world learning experience in which children may draw on other cues to attempt to resolve referential ambiguity. This present study introduces linguistic cues to more closely simulate a child's real-world experience of learning novel words. This may increase the ecological validity of this paradigm and make the results more generalizable to real word-learning settings.

Finally, much attention in the ALM and HTM literature has been given to patterns of successful learning over time (Akhtar & Montague, 1999, Medina et al., 2011; Scott & Fisher, 2012; K. Smith et al., 2011; L. Smith & Yu, 2008; Yu & L. Smith, 2007; Yu & L. Smith, 2011; Yurovsky et al., 2014), but there has been traditionally less focus on error patterns. The errors children make can provide useful insight into the nature of their word-learning processes, but this is a largely untapped source of information. For example, Medina et al. (2011) found that accuracy dropped between the first and second human simulation vignettes in their study, but they simply referred to this as a "so far unexplained and surprising consequence of added contextual information" (p. 9017). Alternatively, Trueswell et al. (2013) and Roembke and McMurray (2016) analyze error patterns, but these studies utilize adult subjects, which, may limit their relevance for understanding developmental patterns of lexical acquisition. Woodard and colleagues (2016) utilized ambiguous initial trials, but then "switch" conditions in follow up trials (that is, the child's initial selection was not offered as a target), to see if participants could flexibly alter their decision and use other past potential targets to inform their decisions. The authors found that, as the HTM would predict, the children were at chance, implying no memory of past alternative targets. This is the closest the existent body of literature has come to studying error patterns in children. However, since the initial trial didn't have an actual correct answer, children were not truly wrong on the initial trial, only made to *believe* they were wrong. The error patterns of children as they learn novel words has not been adequately studied to date.

Research Question

The present study uses a novel noun-learning task to contribute to the important question of how preschoolers learn new words in complex, ambiguous settings. The research question this study addresses is: on a complex word-learning task, when children select an incorrect foil on a learning trial, are they above chance accuracy on a subsequent extension trial, thus theoretically supporting the ALM? Or, alternatively, are they at chance accuracy, thus theoretically supporting the HTM? In addition, several *post hoc*, exploratory analyses were conducted to provide further insight on the results. Investigating this question may provide valuable evidence to the ongoing debate between the ALM and the HTM, and enrich the field's understanding of word-learning mechanisms in young children.

Methods

Participants. The data considered in this study comes from the normative sample data collection conducted for the development of a new language screening tool, the Quick Interactive Language Screener (QUILS; Brookes Publishing, 2017). The QUILS describes their participants thus: 674 monolingual English-speaking children (352 female) recruited from preschools and Head Start programs in Massachusetts, Pennsylvania, Delaware, Florida, and Nebraska. This sample included 213 three-year-olds, 315 four-year-olds, and 146 five-year-olds. Children's ages ranged from 3.03 to 5.92 years (M = 4.39, SD = 0.66). A majority of the children tested were from low-SES families (76.1%) and the remaining children were from mid-SES families. Socioeconomic status (SES) was determined by maternal education: bachelor's degree or higher was considered middle SES (n = 161); fewer years of education was categorized as low SES (n = 513). Demographic data was available for 43.6% of the participants. Of those who reported this information, 57.8% were White, 31.6% were Black/African American, 8.8% were multiracial, and <1% were Asian or other races. Additionally, 45.9% of parents reported whether or not their child was of Hispanic origin, and of those, 23.3% of children were of Hispanic origin.

Procedure and materials. The QUILS uses touchscreen technology to evaluate a child's receptive language skills in three areas: vocabulary, syntactic abilities, and language acquisition processes. Children were presented with a variety of test items to evaluate all three areas. Children were tested in a quiet room; the audio from the QUILS was delivered to them via headphones while a headphone splitter delivered the same audio to the experimenter. Before the screening began, the experimenter told the children they were going to play some games on a special computer in which a man was going to tell them to touch pictures on a screen. The experimenter demonstrated how to use the touchscreen and told the children to pay close attention to what the man tells them to touch. Three training items were administered to ensure that the child understood the test structure and could successfully operate the touchscreen.

The specific QUILS test items considered in this study were the five novel word-learning items; the intent was for children to correctly match an unfamiliar word with a unfamiliar object. The five novel nouns were *fep*, *pluff*, *merf*, *taf*, and *gelp*. Each of the five test items had two trials: a learning trial and an extension trial.

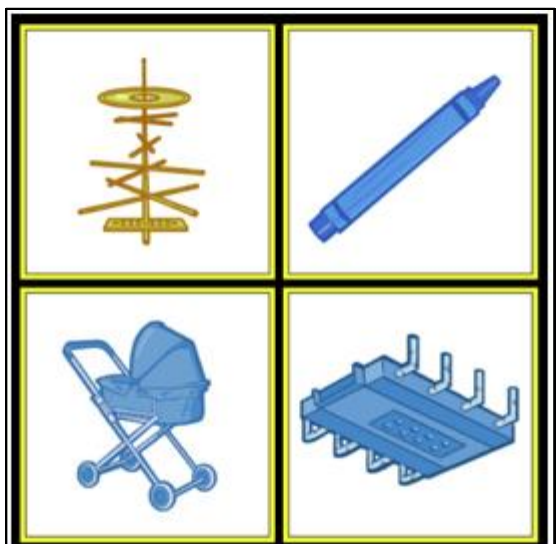


Figure 1. Learning trial stimuli for *fep* item

During the learning trial, children were shown four objects (two novel, two known) and given a nonword (e.g., *fep*). The task instruction included cues about which novel object was correct (e.g., when the child sees Figure 1, they hear, “Look! The fep is blue. Show me the blue fep”). Thus, the learning trial required an integration of mutual exclusivity and a linguistic cue. Unlike much of the past word-learning tasks in the ALM and HTM literature, the initial trial was not completely ambiguous. Children could draw on information other than mutual exclusivity to aid their selection process, much as they do in real-life situations.

The linguistic cues given during the learning trials can be classified into three categories across the five items: one feature cue, two object-relationship cues, and two taxonomic cues. The feature cue is found in the *fep* trial; “the fep is blue.” The cue is a fundamental yet mutable feature of the object. The object-relationship cues are found in the *pluff* and *gelp* trials: “the pluff is on the table” and “the gelp is wearing a hat.” These cues introduce a common object (table, hat) that the children can link with the appropriate target. Finally, the taxonomic cues are found in the *merf* and *taf* trials: “a merf is a kind of bird” and “a taf is a kind of flower.” These cues give the children a semantic category to discern which target is correct.

Additionally, the foils from the learning trial can be classified into two categories: linguistically-aligned foils (two per trial) and a novel object foil (one per trial). The linguistically-aligned foils are those which meet the given linguistic requirement (i.e., feature cue, object-relationship cue, or taxonomic cue), but are known items. The novel object foil is that which meets the mutual exclusivity requirement (it is a novel item), but does not follow the linguistic cue. As an example, on Figure 1, the linguistically-aligned foils are the two known blue objects (buggy, crayon) and the novel object foil is the yellow whirly-gig. All learning trials follow this structure.

Following each learning trial is an extension trial. For each of the five items, children are shown four objects and asked to find the same nonword again (e.g., children see Figure 2 and hear, "Can you show me another fep?"). For these extension trials, the feature and object-relationship properties were removed from the target, and orientation of the target was changed. Taxonomic properties (e.g. "is a kind of bird") remained, due to their nature. Similar to the learning trial, one foil is aligned with the linguistic cue (e.g., the blue mug on the extension trial for *fep*), and at least one foil is a novel object foil (e.g., the robot-like item on the extension trial for *fep*). Note that this trial is called an extension trial rather than a generalization trial because it asks children to extend the knowledge to a new context; the object has changed slightly. A full description of the prompts for all five items and their accompanying visuals is provided in Appendix A.

Figure 2. Extension trial stimuli for *fep* item

Hypotheses

The purpose of this study is not to test the validity of one model in particular, but rather to utilize data from a large sample of preschoolers and evaluate which model best explains the findings. For these data, one way to do this is to consider children's selections on the extension trial *after an error* on the learning trial. The ALM suggests that children maintain multiple hypotheses over time. Thus, it would theoretically predict that children should perform significantly *above-chance accuracy* on the extension trial after an error because they entertain multiple possible referents from the learning trial. After an error, they should recognize – whether implicitly or explicitly – the alternatives that were encountered during the learning trial, thus boosting their accuracy to above chance levels. Conversely, the HTM suggests that children maintain only one hypothesis. It would theoretically predict that children should be *at chance accuracy* on the extension trial after, because they do not entertain other options. The final possible option, significantly *below chance accuracy*, is not predicted by either model and may be indicative of interference from the linguistic cues or the presence of a particularly salient foil.

Analyses and Results

A priori **analyses and results**

Children's error patterns were analyzed to test these hypotheses. Only children who got the learning trial *incorrect* were considered in the analyses of the extension trial for each item. Recall that 674 children were administered all five items; for each item, children who got the learning trial wrong were isolated for analysis. Since subjects were pulled from the same sample of children, separate analyses were conducted for each item. The number of children who got the learning trial wrong for each item are reflected in the data reported below.

The dependent variable was children's accuracy on the extension trial, calculated as a proportion for each potential response. This variable was compared with chance performance on each item. Two-tailed hypothesis tests for one proportion were conducted to determine if the observed proportion of accuracy differed significantly from the hypothesized proportion of accuracy. Chance for all analyses was set at .25, because there are four possible selections on the extension trial. Limitations with this interpretation of chance are addressed in the discussion.

MedCalc’s test for one proportion calculator (Schoonjans, 2017) was utilized for these analyses. Results of the *a priori* analyses are shown in Table 1 and Figure 3.

Table 1

Extension trial accuracy after error on learning trial

Item	*n*	Observed Proportion (accuracy)	Hypothesized Proportion (chance)	*z* score	*p* value
fep	331	.28	.25	1.30	.19
pluff	393	.21	.25	-1.66	.10
merf	237	.26	.25	0.26	.79
taf	404	.29	.25	1.72	.09
gelp	298	.24	.25	-0.47	.64

These results show that after children selected an inaccurate answer on the learning trial, they were not significantly different from chance accuracy on the extension trial across all five items.

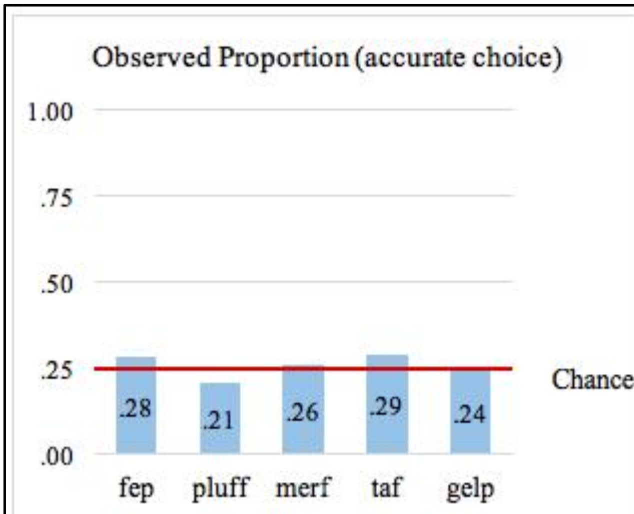


Figure 3 - Extension trial accuracy after an error on the learning trial as compared to chance

***Post hoc* analyses and results**

The *a priori* analyses considered only accuracy on the extension trial. While this approach is informative about children’s learning strategy when it comes to successful repair during word learning, it does not account for instances of failed word learning. Thus, several *post hoc* analyses are justified to observe more holistic patterns of children’s responses on this task to understand the underlying learning mechanisms.

The first set of *post hoc* analyses considered the distribution of answers across all four objects on the extension trial after an error on the learning trial. This analysis utilized the same pool of children as the *a priori* analyses: children who erred on the learning trial. A one by four chi-square goodness-of-fit test was utilized to see if the distribution of answers on the extension trials aligned with chance (placed at .25) or if the distribution is likely not due to chance. Unlike the *a priori* analyses, these analyses determined if children have an at-chance *distribution* of answers, not just at chance *accuracy*. These analyses were run using a one-way chi-square calculator (VassarStats, 2017). Results are reported in Table 2 and they reveal that the distribution of answers across all five items on the extension trial (after children erred on the learning trial) is statistically significantly different from chance.

The second set of *post hoc* analyses provided insight on *why* the distribution of answers is very likely not due to chance. Observation of the raw selection frequencies for each of the four objects on the extension trial (see Appendix B) clearly revealed that children most frequently selected the linguistically-aligned foil. For example, on the *fep* item, the children who erred on the learning trial most frequently chose the blue foil (the mug) on the extension trial. This is presumably because the preceding learning trial includes the linguistic cue "is blue." To confirm statistical significance of this observation, two-tailed hypothesis tests for one proportion were applied to the linguistically-aligned foil on each of the extension trials. This determined if the observed proportion for this foil differed significantly from the hypothesized proportion of selection, which was set at chance

Table 2

Distribution of answers on extension trial

Item	n	χ^2	p value
fep	331	31.24	<.001*
pluff	393	75.00	<.001*
merf	237	33.57	<.001*
taf	404	37.90	<.001*
gelp	298	40.12	<.001*

* = Statistically significant

(.25). These analyses used the same statistical MedCalc software (Schoonjans, 2017); results displayed in Table 3 and Figure 4.

Table 3

Selection of linguistic foil after error on learning trial.

Item	*n*	Observed Proportion (linguistic foil)	Hypothesized Proportion (chance)	*z* score	*p* value
fep	331	.31	.25	2.32	.02*
pluff	393	.43	.25	8.36	<.001*
merf	237	.40	.25	5.36	<.001*
taf	404	.34	.25	4.25	<.001*
gelp	298	.38	.25	5.15	<.001*

* = Statistically significant

This shows that after children erred on the learning trial, they were very statistically likely to choose the linguistically-aligned foil on the extension trial. This is most probably indicative of interference of the linguistic cue on this word-learning task. Thus, accurate target selection, which is what was analyzed in the *a priori* analyses, was disrupted by the complex structure of the task and is likely not a particularly useful measure of children's learning mechanisms.

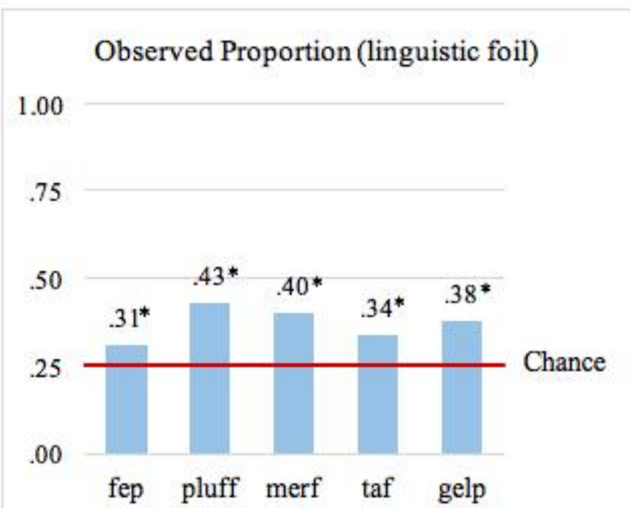


Figure 4 – Selection of linguistic foil on extension trial, after an error on the learning trial, as compared to chance.

The third and final *post hoc* analyses were conducted in an attempt to reduce the interference of the linguistic cue by considering only children who ignored the linguistic cue and chose the novel object foil on the learning trial. For example, from the "fep" item, this analysis only included the children who chose the novel yellow whirly-gig on the learning trial. As another example, from the "merf" item, this analysis included only the children who chose the novel hand truck. This excluded the children who were influenced by the linguistic cue on the initial learning trials; it considered only those who chose the novel object foil. These are the children who successfully recognized that the task required mutual exclusivity, but ignored the presence of the linguistic cue, and so still erred. Their subsequent responses on the extension trial are likely a purer representation of which learning mechanisms children are employing when faced with recognition that they've made an error. A two-tailed hypothesis test for one proportion using the same MedCalc statistical software (Schoonjans, 2017) was applied to the answers from the extension trial to compare these children's accuracy against chance. Results are displayed in Table 4 and Figure 5.

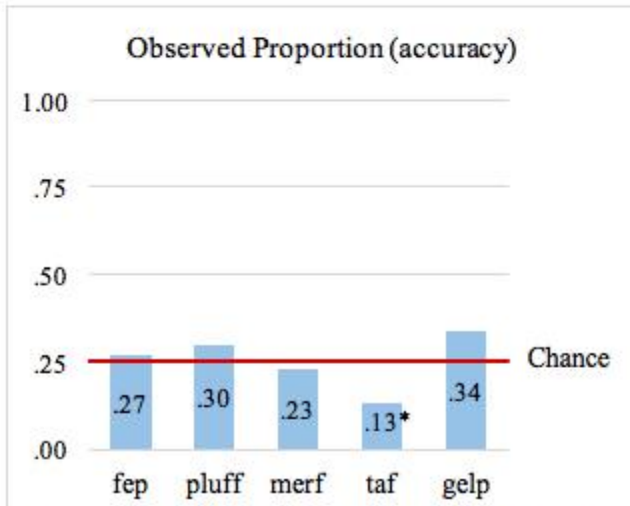


Figure 5 - Extension trial accuracy after selecting the novel foil on the learning trial, as compared to chance

Table 4

Extension trial accuracy after selecting novel foil on learning trial

Item	*n*	Observed Proportion (accuracy)	Hypothesized Proportion (chance)	*z* score	*p* value
fep	30	.27	.25	0.21	.83
pluff	40	.30	.25	0.73	.47
merf	30	.23	.25	-0.21	.83
taf	100	.13	.25	-2.77	.006*
gelp	35	.34	.25	1.27	.20

* = Statistically significant

Note that children who chose the novel object foil on the learning trial are quite infrequent; sample sizes in this *post hoc* analyses are very small and should be interpreted with great caution. The statistical analyses conducted here recommend larger samples sizes be used, so results are exploratory at best. These results show that on four of the five items, children are not significantly different from chance accuracy on the extension trial after choosing the novel object foil on the learning trial. On the *taf* item, they are statistically below chance accuracy.

Discussion

Summary and Implications

The *a priori* analyses in this study considered proportions of children's accuracy after making an error. Specifically, they analyzed all the children who got the learning trial incorrect, regardless of the foil they selected (linguistically-aligned foils or the novel object foil). Accurate responses on the subsequent extension trial were compared with chance to see whether children performed statistically significantly above chance, similar to chance, or below chance. Recall that accuracy statistically above chance would be theoretically indicative of recollection of previous possible targets, which the ALM suggests that children maintain across trials. Accuracy statistically similar to chance would be indicative of a lack of storage of previous possible targets, as the HTM suggests. Accuracy statistically below chance would be predicted by neither model and would likely be the result of interference from a salient foil. Results indicated that for all five items, children's accuracy on the extension trial was not statistically significantly different from chance, regardless of the foil type they chose on the learning trial. These findings align with the HTM and suggest that children likely posit one hypothesis for a novel word-referent pairing when faced with some ambiguity in word-learning settings.

However, while children's selection of the *accurate* target was not statistically different from chance, the first two sets of *post hoc* analyses showed that the *distribution* of answers was not due to chance and that children were most likely to choose the linguistically-aligned foil on the extension trial. Proportions from the raw data also revealed that if children made a linguistically-aligned error on the learning trial (which most children in error did), then they were more likely to make the same error on the extension trial (see Appendix C). For example, on the *fep* item, the children who initially chose the blue crayon or blue buggy mostly went on to choose the blue mug on the extension trial. This pattern of performance suggests that these children failed to recognize the mutual exclusivity requirements for the task; they never chose a novel item. Both the ALM and the HTM models can account for this kind of *incorrect* learning across trials: the ALM would suggest they are cataloging these repeated exposures over time to form a solidified (albeit incorrect) word-referent pairing, and the HTM would suggest that they have already formed an (incorrect) hypothesis, and they are verifying that hypothesis on the extension trial. It is reasonable to suggest that these children believed the trial was about the cue itself; they may never have realized their error, and thus never employed either of the repair strategies that the models predict. This implies heavy influence of the linguistic cue and subsequently justified an additional set of analyses that separately considers children who were not influenced by this cue on the learning trial.

The final analyses, which examined only the children who chose the novel object foil during the learning trial, may be the most informative for contrasting between the ALM and the HTM. Whereas the children who chose the linguistically-aligned foil on the learning trial could do the same on the extension trial, those who chose the novel object foil had to adapt their strategy on the extension trial. Because the novel referent they selected was not present on the extension trial, these children very likely recognized that their selection was incorrect. This recognition of inaccuracy highlights the repair strategy that these children may have adopted to achieve successful word learning. The predictions regarding accuracy would be the same as already stated for the *a priori* analyses: the ALM predicts above chance accuracy and the HTM predicts at chance accuracy. Since they are at chance accuracy for four of the five items, the implication of these results supports the HTM. Together, these findings suggest that children are working with a single hypothesis and may not be recalling other possible unselected answers from the learning trial or cross-checking their knowledge when they get to the extension trial. Instead, they are at chance accuracy when their original conjecture is disconfirmed. However, these results are exploratory due to the small sample sizes.

Children's performance on the *taf* item requires additional consideration. On this item alone, responses were statistically *below* chance – a finding that is predicted by neither model. One explanation for this finding may simply be the nature of that item's design: the novel item on the learning trial for *taf* looks quite visually similar to a novel item on the extension trial for *taf*, and the proportions from the raw data shows that that visually similar foil was the most frequently selected by these children (see Appendix D). However, it may also be indicative of continued interference of the linguistic cue. Or, perhaps it indicates other learning mechanisms employed by children. Since this item also has the largest sample size and the most power in this set of analyses, this unexpected finding is notable.

Overall, across the *a priori* and *post hoc* analyses, results are inconclusive as to which model is best supported. While some analyses appear to lend credibility to the HTM, other data contradict the validity of this trend. Yet, the ambiguity that emerges from these results may itself be the most valuable information: such ambiguity may imply that word learning is not adequately captured by either model. Perhaps, as some researchers have begun to suggest, associative learning and hypothesis testing are *both* driving forces; perhaps they even interact with one another. For example, Romberg and Yu (2014) designed a study in which adult participants were prompted to use a hypothesis testing approach for some trials; they then assessed how this prompting affected statistical associative learning. They proposed various frameworks for how the two approaches may both influence a word learning task, and their results supported interaction between the ALM and the HTM in a variety of ways.

Even more revolutionary, some researchers have proposed that the ALM and the HTM may simply both be different expressions of the *same* underlying mechanism. It is well known that other cognitive processes, such as information selection, retrieval, and storage, are essential parts of learning; these processes are not captured by current word-learning models. The observed instances of associative learning or hypothesis testing collected in the research so far may simply be contrasting examples of a unified set of learning mechanisms that have yet to be cohesively modeled (Yu & L. Smith, 2012; Yu, L. Smith, Klein, & Shiffrin, 2007). This present thesis yields inconclusive evidence; from item to item, there are slightly different trends in children's responses. A hybrid approach, in which children may test hypotheses and also maintain partial information about alternative referents (all while supported by the cognitive processes necessary for learning) could potentially account for the varying patterns seen in this present data. More complex hybrid models may emerge when more is known about the psychological processes that influence information selection, storage, retrieval, and prioritization; all these components undergird word learning (Yu & L. Smith, 2012).

Limitations

The language acquisition models explored in this study attempt to explain how children track and carry word knowledge over repeated exposures. This study contained only two trials for each item, one for learning and one for extension, and so considering error patterns over time is limited. Since there were only a total of two exposures in this study, there was only one junction for each novel word: the pattern displayed from learning trial to extension trial. It is possible that children's patterns of responses may have changed in informative ways if they were given more exposures to the novel words with varying levels of indeterminacy. Complex repetition trials such as this were not considered in the present study.

This study, unlike much of the previous literature, uses word-learning trials that are not completely indeterminate. Although this more closely approximates children's real world learning environments, it also complicates the data. It is not a pure indeterminacy task. The introduction of the linguistic cue allows children to use other strategies, which are not accounted for in the ALM and the HTM, to correctly learn the novel word. The findings reflect this uncertainty: the contrast of the models as revealed by the *a priori* results are dampened by the interference of the linguistic cues revealed in the *post hoc* analyses. The lack of absolute indeterminacy in this study limits its ability to isolate the individual learning mechanism that the ALM and the HTM hope to capture, so direct comparison between the models is muted.

Still, it is abundantly clear that children do not learn words by employing only a single cue in their everyday lives. As aforementioned, the literature shows that children capitalize on the context (Carey, 1978), the speaker's behavioral and attentional patterns (Baldwin, 1991; L. Smith, 2000), various socio-pragmatic cues (Baldwin, 1993; Tomasello, 2000), and the linguistic and syntactic constraints surrounding a novel word (Gleitman, 1990), to name a few strategies.

The purpose of this research was to understand children's real-world learning capabilities. Thus, increasingly complex learning trials, with foils that complicate ambiguity, were justified despite their impact on pure referential uncertainty.

To partially account for this complex task design, the final post hoc analyses removed children who were influenced by the linguistic cue, which presents with another set of limitations. The children who exhibited this pattern of responding are infrequent, so sample sizes are small and hold little statistical power. Increasing these sample sizes is particularly difficult since they are a reflection of children's free choices, a variable that could not be manipulated with the study design. Thus, although these results clearly contrast the ALM and the HTM, and reflect an advantage towards the HTM, they should be weighed carefully based on their limited power.

Another previously referenced limitation with this study arises from the interpretation of what is considered chance accuracy, particularly in the presence of common object foils. Chance for all analyses was set at .25, because there are four possible selections on the extension trial. This does not factor in children's likely exclusion of common objects that they recognize and know the word for. If chance accuracy was considered to be higher (i.e., .33 or .50, excluding presumably known objects), then the results from this study, across all proportional analyses, would result in below chance accuracy. However, chance was intentionally considered at .25 in these analyses for two main reasons. First, we cannot definitively document that the tested children have experience with the common objects and would rule them out. Second, by it's very nature, chance does not consider other factors: four possible selections means that pure chance is .25. Thus, all choices were given equal weight.

Future directions

The large corpus of data utilized in this study can and should be applied to future research projects that could provide relevant understanding to the mystery of how preschoolers learn novel words. One such future direction would be to consider children's patterns across age. This study analyzed all the participants, ages 3;0-5;11, together, but future research could consider age groups individually. It is possible that since there is such a dramatic difference in language skills across these ages, one model may align better with one age than the other. For example, one argument already posited in this study highlighted how children and adults may utilize different strategies in ambiguous learning situations. Considering how children perform on this novel word-learning task across age may reveal a bias towards one model at a younger age and another model as children's language skills become more advanced.

Additionally, as this corpus of data was collected, reaction times were also documented. Although these data are beyond the scope of the present study, analyzing preschooler's reaction times on the extension trial after an error after the learning trial may also add insight to the crucial consideration of whether children are storing multiple possible referents. Longer reaction times could be indicative of hesitancy, which lends credibility to the idea that they are lingering on previously seen targets. While eye-tracking data has been used in the existent literature (Roembke & McMurray, 2016, L. Smith & Yu, 2008; L. Smith & Yu, 2011; Trueswell et al., 2013), reaction times are a largely unconsidered source of information. Analyzing this variable could be a valuable contribution to the literature.

Finally, future directions may also include using this research for development of clinical interventions for children with language delays. The knowledge presented here, as in much of the word-learning research, represents basic knowledge that has yet to be applied to serve any clinical purposes. The purpose of exploring this basic knowledge is to better understand how word learning works, that clinicians may someday better serve children who have difficulty with word learning. Of course, much more research is required before any clinical applications will be feasible. Currently, research is ongoing to administer this same word-learning task to populations of children who have diagnosed communication delays. This data, once collected, can be contrasted with the data from the typically developing children represented in the current sample in many worthwhile ways. It is possible that there are consistent and meaningful differences in the way typically developing children approach this task versus the way children with communication disorders approach this task. These differences and patterns may eventually inform treatment approaches that are specifically tailored to the learning mechanisms of the populations they intend to benefit.

Conclusion

Beginning in infancy and continuing through childhood, children are miraculously adept word learners, even in the face of ambiguous learning situations. There is an ongoing debate about how children perform the daunting task of learning new words when they are faced with referential indeterminacy. Do they entertain multiple possible referents across trials as the ALM suggests, or do they use a propose-then-verify approach as predicted by the HTM? These two prevailing models have different theoretical predictions for how children should perform after an error, so this study considered preschooler’s responses after an incorrect selection on a word-learning task. This research adds to the existing body of literature that contrasts these models and provides insight into how children may learn novel words. Results showed that while the ALM and the HTM provide excellent starting points, much is still unknown and other hybrid models may better encapsulate children’s strategies. Children’s real-world contexts are vastly more complex than captured by any of the current data, and to truly begin to answer the question of how children learn words in their daily lives, more research is needed.

Works Cited

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: the role of cross-situational learning. *First Language*, *19*, 347–358.

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

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

Baldwin, D. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*(02), 395–418.

Barrett, M. D. (1986). Early semantic representations and early word-usage. In S. A. Kuczaj II, & M. D. Barrett (Eds.), *The development of word meaning: Progress in cognitive development research* (pp. 39-67). Springer-Verl.

Barrett, M. D. (1996). Early lexical development. In Fletcher, P. and MacWhinney, B. (eds.), The Handbook of Child Language (pp. 362-392). Oxford: Blackwell.

Barrett, M. D. (1999). The development of language. Hove, East Sussex, UK: Psychology Press.

Byers-Heinlein, K., & Werker, J. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science, 12*, 815–823.

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

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), 33–44.

Carey, S. (1978). The child as word learner. *Linguistic Theory and Psychological Reality*, eds Halle M, Bresnan J, Miller GA (MIT Press, Cambridge, MA), 264–293.

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

Cartmill, E. A, Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11278–83

Clark, E. (1990). On the pragmatics of contrast. Journal of Child Language, 17, 417–431.

Courage, M. L. (1989). Children's inquiry strategies in referential communication and in the game of twenty questions. *Child Development, 60,* 877–886.

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

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can Infants Map Meaning to Newly Segmented Words? *Psychological Science*, *18*(3), 254–260.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A Probabilistic Computational Model of Cross-Situational Word Learning. *Cognitive Science*, *34*(6), 1017–1063.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning: Research article. *Psychological Science*, *20*(5), 578–585.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, a. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–76.

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*(1), 3–55.

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

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

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87,* 23–34.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: P. H. Brookes.

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

Locke, J. (1690). *An Essay Concerning Human Understanding,* Cleveland, OH: Meridian Books pp. 196

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

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

Markman, E., Wasow, J., & Hansen, M. (2003). Use of the mutual exclusivity assumption by young word learners. Cognitive Psychology, 47, 241–275.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *10*(22), 9014–9019.

Meints, K., Plunkett, K., & Harris, P. L. (1999). When does an ostrich become a bird? the role of typicality in early word comprehension. *Developmental Psychology, 35*(4), 1072-1078.

Meints, K., Plunkett, K., & Harris, P. L. (2008). Eating apples or houseplants? The role of typicality in early comprehension of verbs. *Language and Cognitive Processes, 23*, 434–463.

Merriman, W., & Bowman, L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 4,* 1–123.

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

Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development, 5*(4), 203–234.

Owens, Jr., Robert E. (1996). *Language development: An introduction (fourth edition)*. Boston, MA: Allyn and Bacon.

Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: new evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, *3*(4), 295–323.

Quine, W. V. (1960). *Word and object.* The MIT Press.

Rajan, V., Konishi, H., Ridge, K., Houston, D., Golinkoff, R.M., Hirsh-Pasek, K. & Schwartz, R. (in press) Novel word-object association at 21 months predicts language-specific outcomes in middle school. *Developmental Psychology.*

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language*, *87*, 105–127.

Romberg, A.R. & Yu, C. (2014). Interactions between statistical aggregation and hypothesis testing during word learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928

Schoonjans, F. (2017). MedCalc test for one proportion calculator [Computer software]. Retrieved from https://www.medcalc.org/calc/test_one_proportion.php

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–80.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39–91.

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

Smith, L. B. (2000). Learning how to learn words: an associative crane. *Becoming a Word Learner*, 51–80.

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

Stahl, Stephen A. (1999). *Vocabulary development*. Brookline, MA: Brookline Books.

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.

Tomasello, M., & Farrar, J. (1986). Joint attention and early language. *Child Development, 57,* 1454–1463.

Tomasello, M. (2000). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 111–128). Cambridge University.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–56.

VassarStats (2017). One-Way Chi-Square [Computer software]. Retrieved from http://vassarstats.net/csfit.html

Vogt, P., & Smith, A. D. M. (2005). Learning colour words is slow: A cross-situational learning account. *Behavioral and Brain Sciences, 28,* 509-510.

Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two- and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development*, *12*(3), 252–261.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13-15), 2149–2165.

Yu, C., & Smith, L. B. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. *Psychological Science*, *18*(5), 414–420.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological review*, *119*(1), 21.

Yu, C., Smith, L. B., Klein, K. A., & Shiffrin, R. M. (2007, January). Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same?. In *Proceedings of the Cognitive Science Society* (Vol. 29, No. 29).

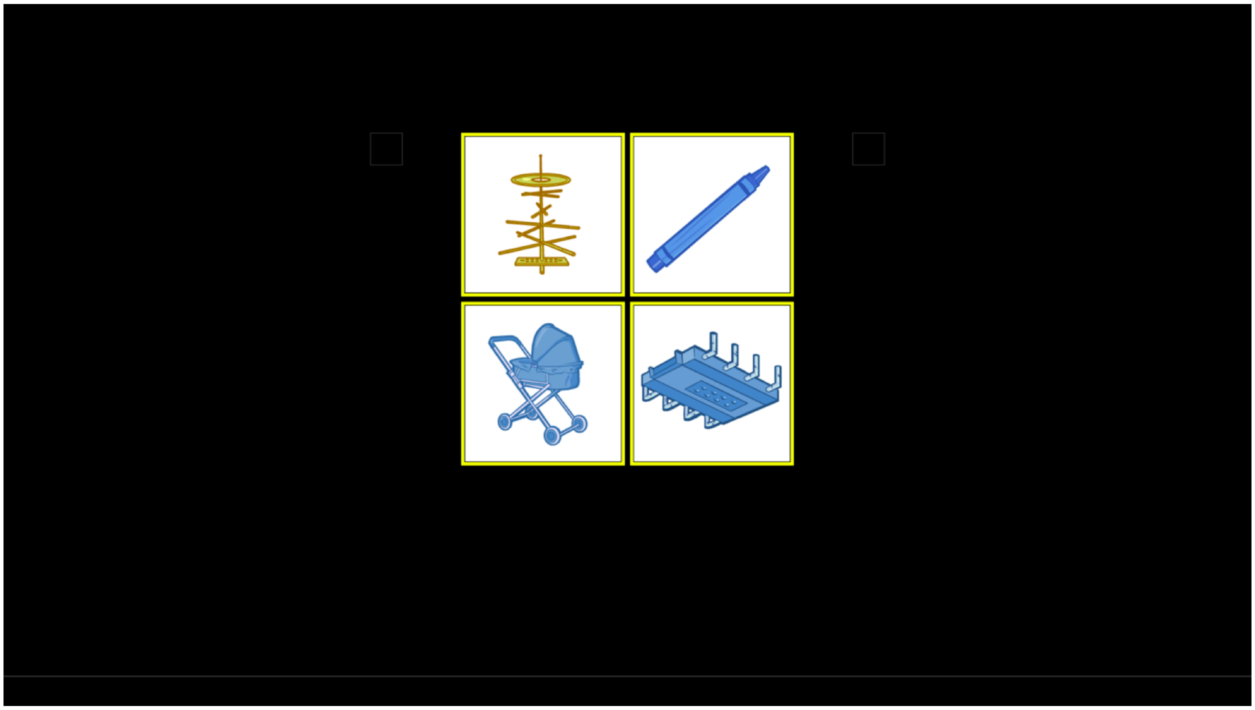
Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin {&} Review*, *21*(1), 1–22.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966.

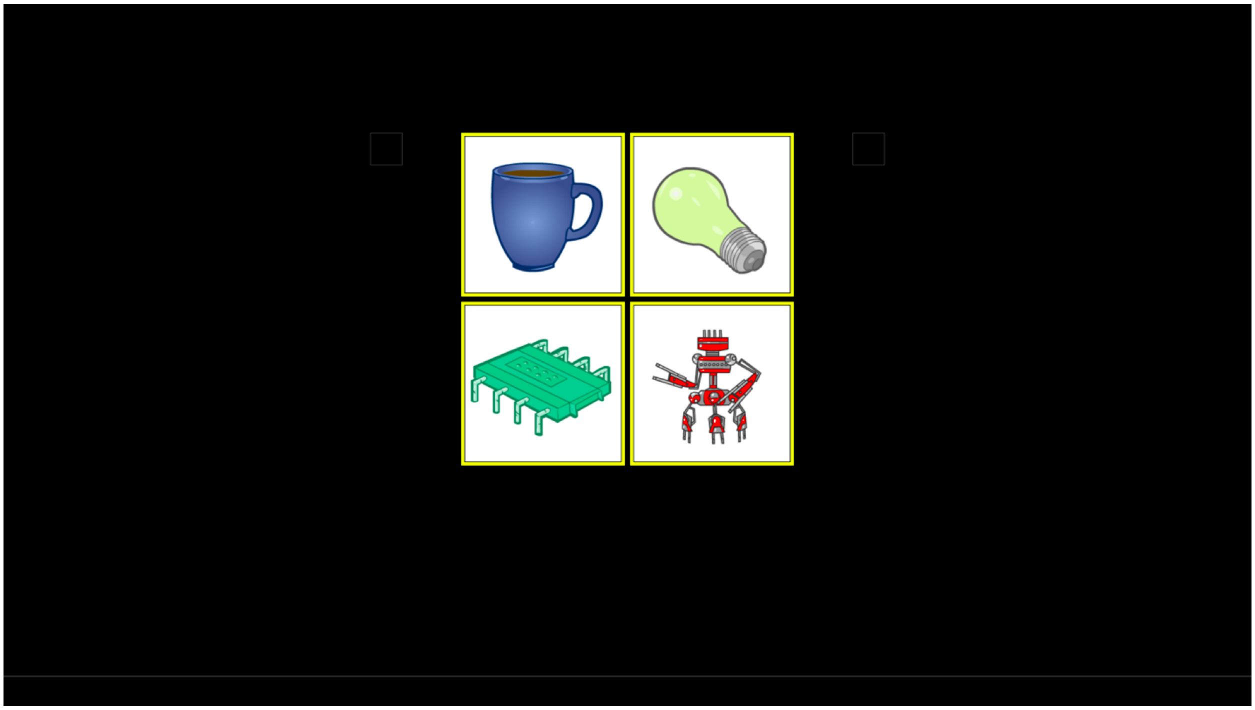
Appendix A

Complete Test Visuals and Accompanying Audio

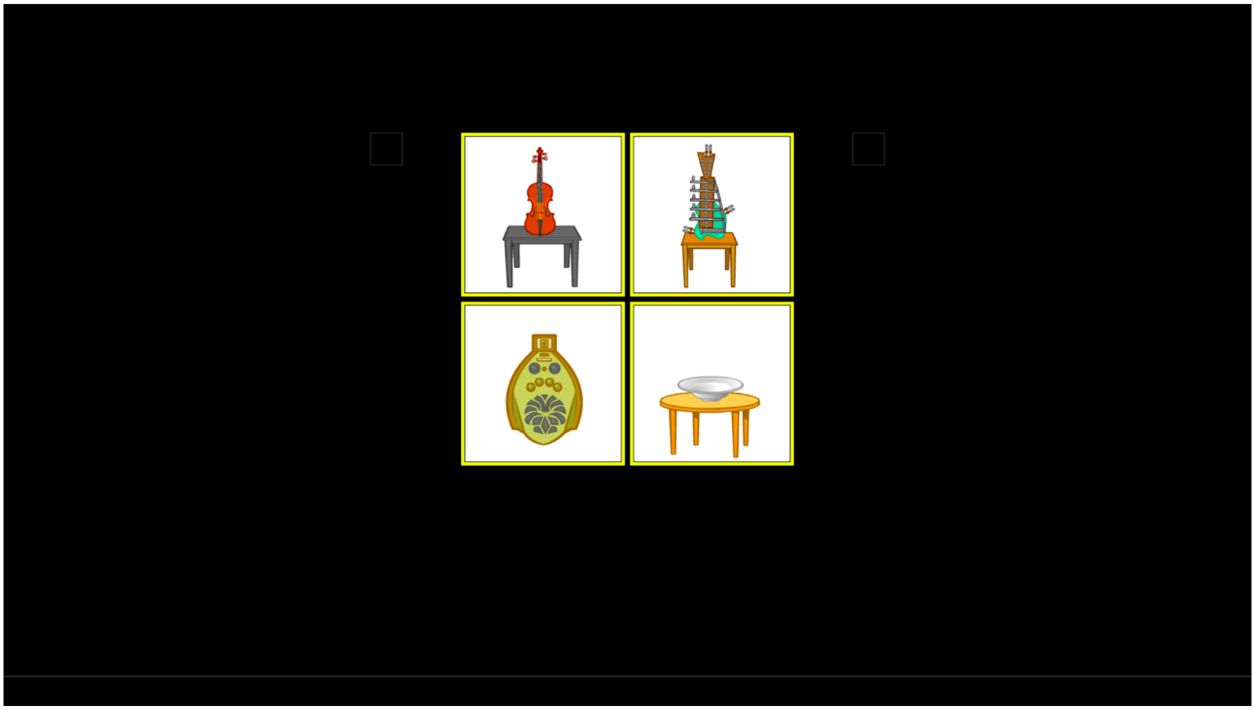
"The fep is blue. Show me the blue fep."



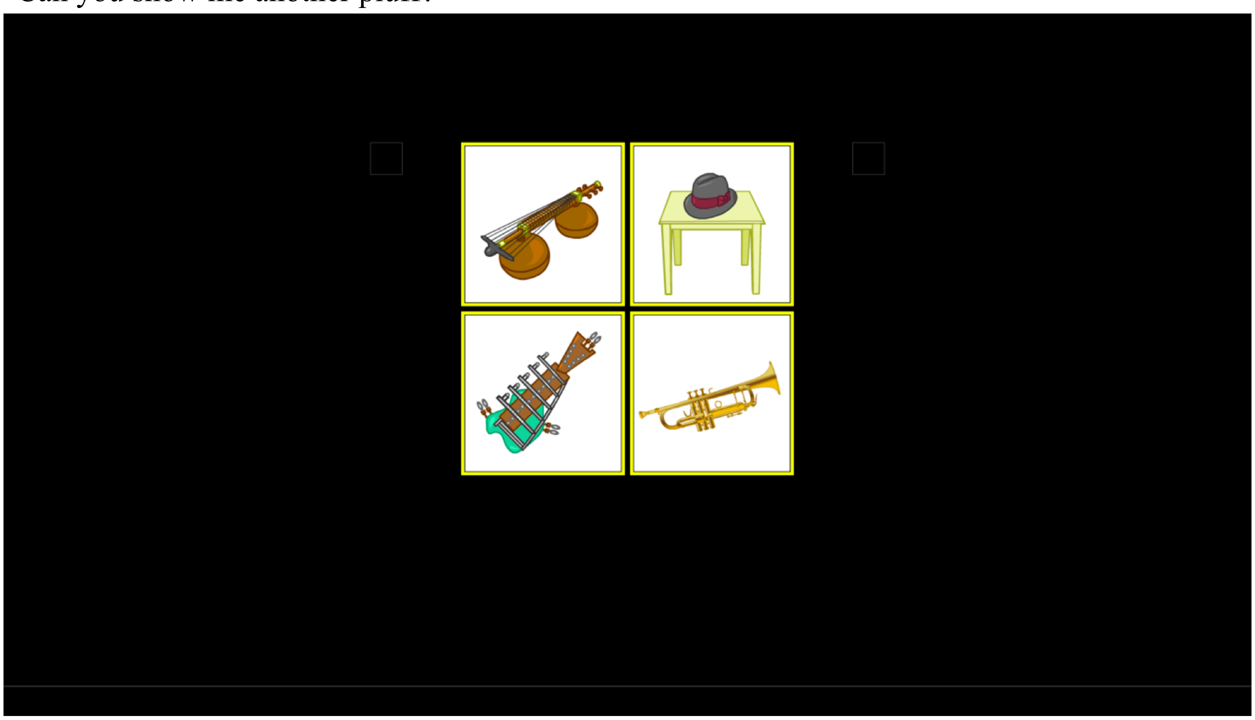
"Can you show me another fep?"



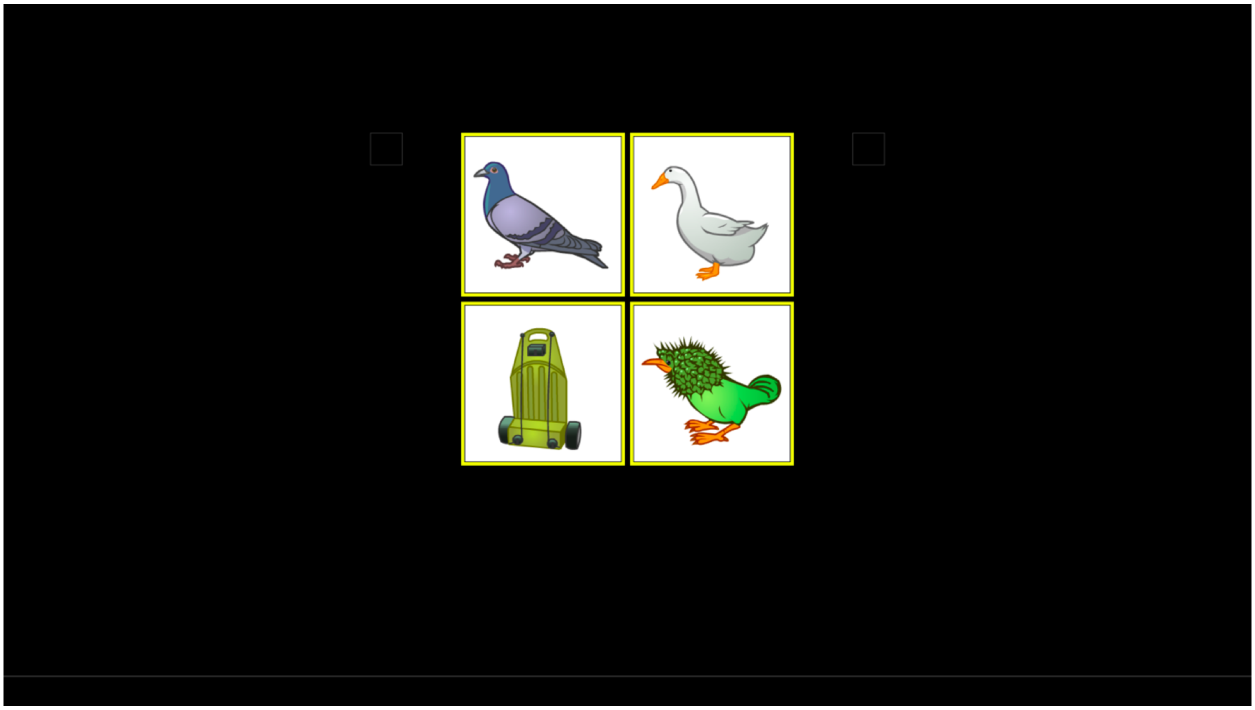
“The pluff is on the table. Show me the pluff on the table.”



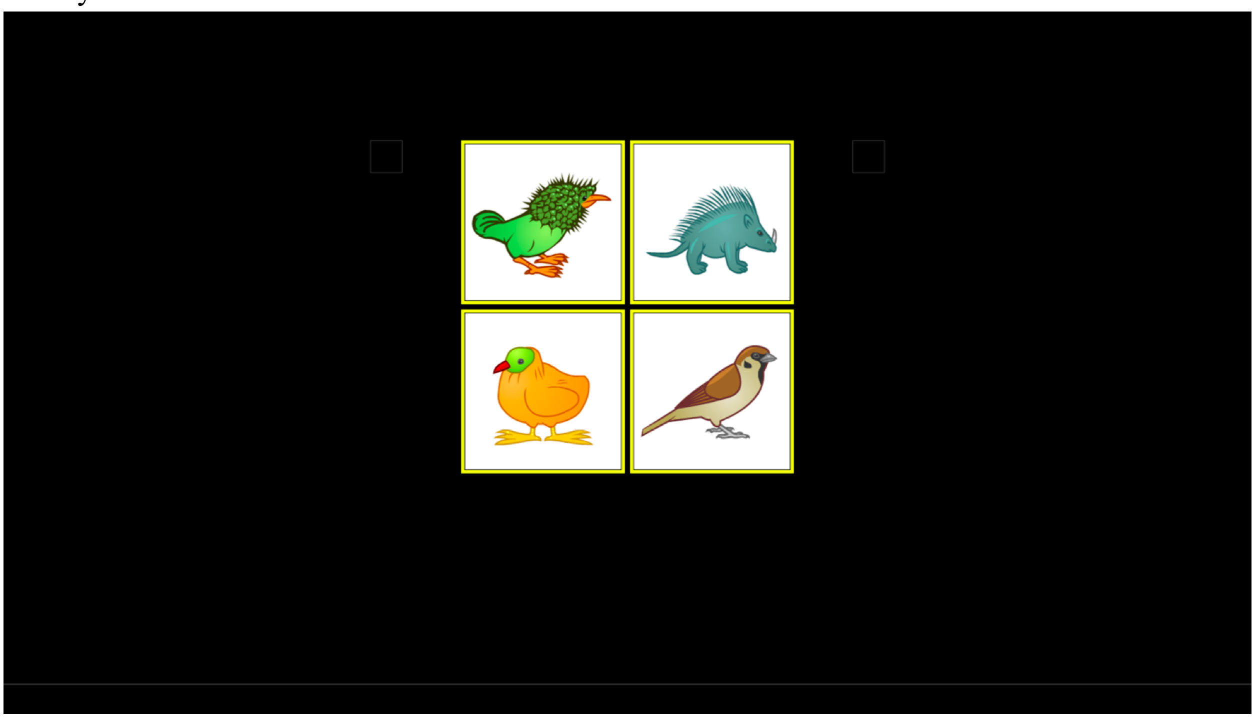
"Can you show me another pluff?”



"A merf is a kind of bird. Show me the merf."

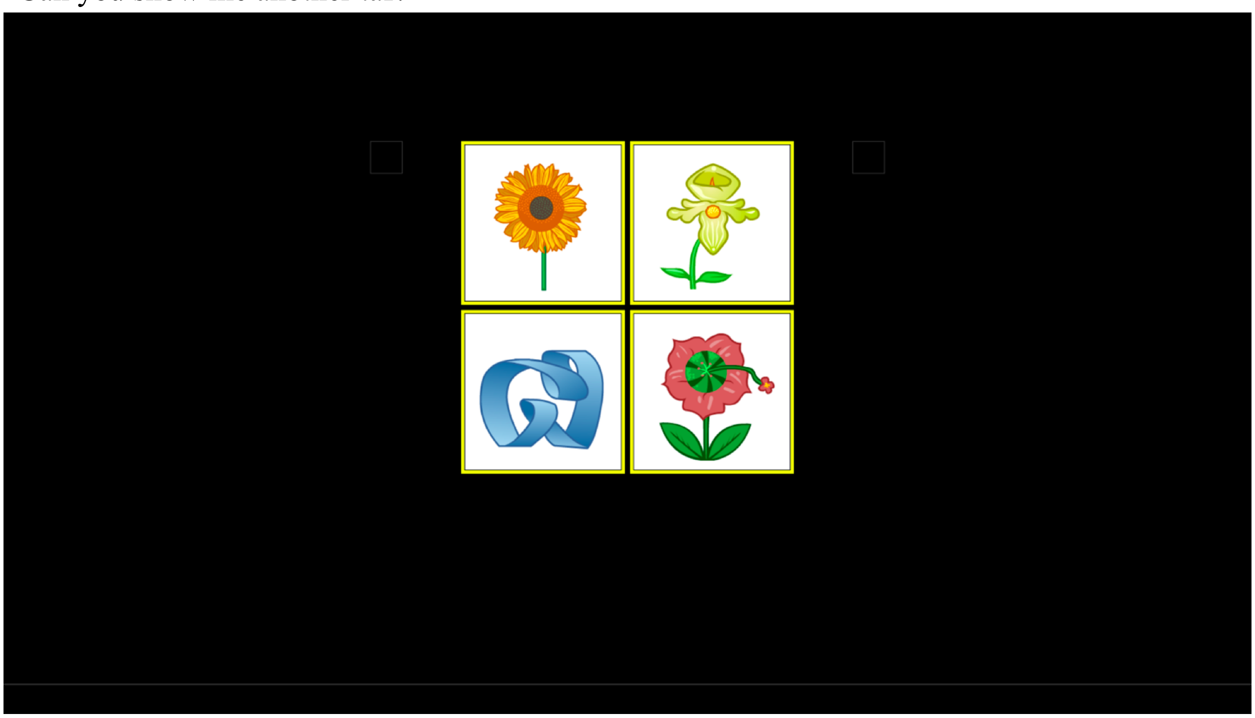


"Can you show me another merf?"

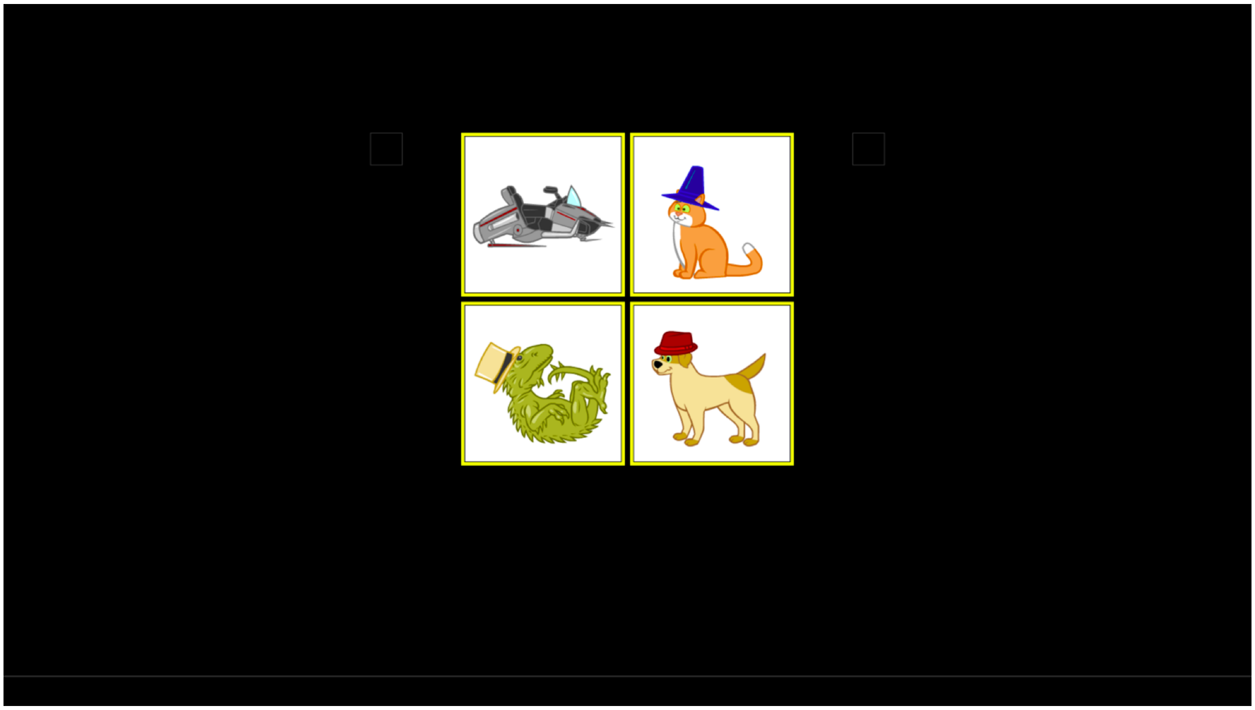


"A taf is a kind of flower. Show me the taf."

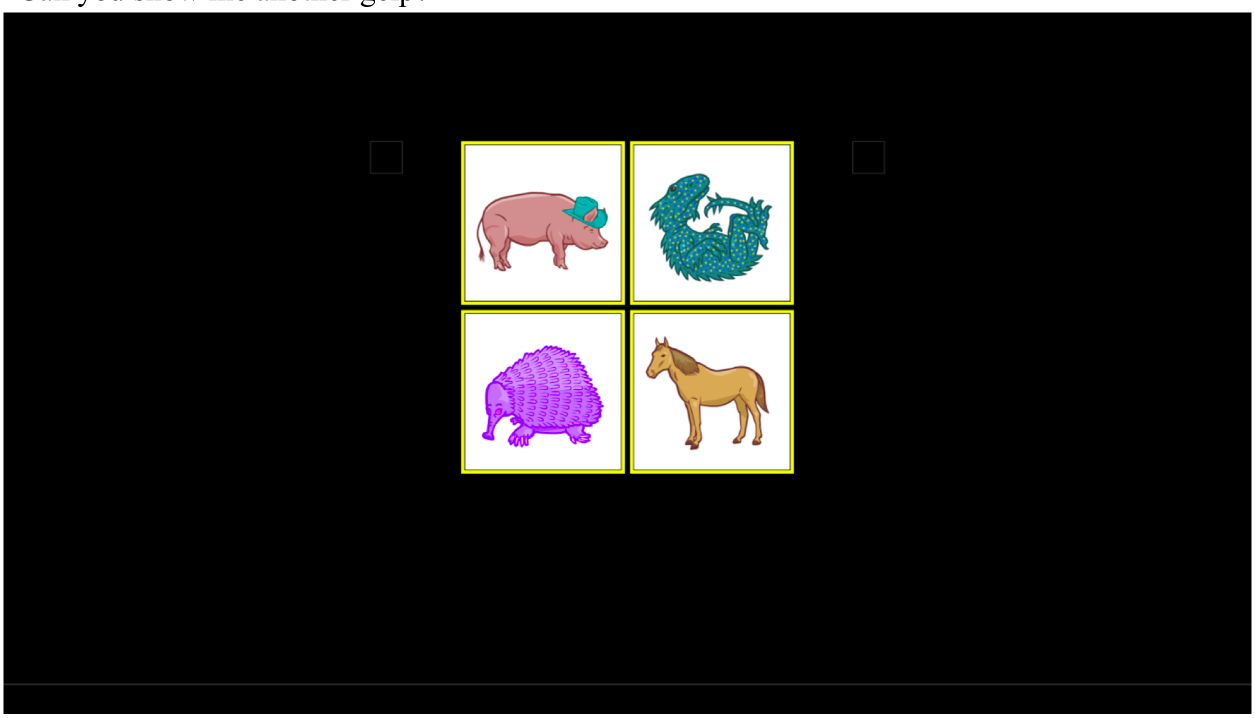
"Can you show me another taf?"



“A gelp is wearing a hat. Show me the gelp with the hat.”



“Can you show me another gelp?”



Appendix B

Raw Distribution of Answers on Extension Trial after Error on Learning Trial

Item	*n*	Choice a: accurate	Choice b: linguistically aligned	Choice c: novel object	Choice d: other*
fep	331	93	101	98	39
pluff	393	84	170	56	83
merf	237	61	95	42	39
taf	404	116	138	96	54
gelp	298	71	113	36	78

*Note: the final foil varies per test item. On the *fep*, *pluff*, and *gelp* items, choice d is a familiar object that is not linguistically aligned with the given cue. On the *merf* and *taf* items, choice d is a novel object that is also linguistically aligned with the given cue.

Appendix C

Raw Distribution of Answers on Extension Trial for Children who Made a Linguistically Aligned Error on Learning Trial

Item	*n*	Choice a: accurate	Choice b: linguistically aligned	Choice c: novel object	Choice d: other*
fep	298	83	97	82	36
pluff	350	67	161	49	73
merf	204	52	90	33	29
taf	300	98	127	38	37
gelp	262	57	106	29	70

*Note: the final foil varies per test item. On the *fep*, *pluff*, and *gelp* items, choice d is a familiar object that is not linguistically aligned with the given cue. On the *merf* and *taf* items, choice d is a novel object that is also linguistically aligned with the given cue.

Appendix D

Raw Distribution of Answers on Extension Trial for Children who Made a Novel Object Error on Learning Trial

Item	*n*	Choice a: accurate	Choice b: linguistically aligned	Choice c: novel object	Choice d: other*
fep	30	8	5	14	3
pluff	40	12	11	7	10
merf	30	7	5	9	9
taf	100	13	13	59**	15
gelp	35	12	7	8	8

*Note: the final foil varies per test item. On the *fep*, *pluff*, and *gelp* items, choice d is a familiar object that is not linguistically aligned with the given cue. On the *merf* and *taf* items, choice d is a novel object that is also linguistically aligned with the given cue.

**Note: The novel object foil on the *taf* extension trial is quite visually similar to the novel object foil on the *taf* learning trial (see Appendix A to view visuals).