



2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty

Rose M. Scott^{a,*}, Cynthia Fisher^b

^a School of Social Sciences, Humanities, and Arts, University of California Merced, 5200 North Lake Road, Merced, CA 95343, United States

^b Department of Psychology, University of Illinois, 603 E. Daniel St., Champaign, IL 61820, United States

ARTICLE INFO

Article history:

Received 9 December 2010

Revised 22 August 2011

Accepted 19 October 2011

Available online 20 November 2011

Keywords:

Language acquisition

Verb learning

Statistical learning

ABSTRACT

Recent evidence shows that children can use cross-situational statistics to learn new object labels under referential ambiguity (e.g., Smith & Yu, 2008). Such evidence has been interpreted as support for proposals that statistical information about word-referent co-occurrence plays a powerful role in word learning. But object labels represent only a fraction of the vocabulary children acquire, and arguably represent the simplest case of word learning based on observations of world scenes. Here we extended the study of cross-situational word learning to a new segment of the vocabulary, action verbs, to permit a stronger test of the role of statistical information in word learning. In two experiments, on each trial 2.5-year-olds encountered two novel intransitive (e.g., “She’s pimming!”; Experiment 1) or transitive verbs (e.g., “She’s pimming her toy!”; Experiment 2) while viewing two action events. The consistency with which each verb accompanied each action provided the only source of information about the intended referent of each verb. The 2.5-year-olds used cross-situational consistency in verb learning, but also showed significant limits on their ability to do so as the sentences and scenes became slightly more complex. These findings help to define the role of cross-situational observation in word learning.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Ordinary word learning takes place in a complex communicative setting that can lead to daunting referential ambiguity (e.g., Gleitman, 1990). Imagine a child who sees a group of girls playing on a playground. One girl pushes another on a swing, a third girl slides down a slide while shrieking, and a fourth clammers on a jungle-gym. While viewing this scene, the child hears his mother say, “Oh look! She’s pimming!” Given the context, this child has no way of knowing whether *pimming* refers to pushing, swinging, sliding, shrieking, or climbing. Is the child able to learn anything about the meaning of *pim*, given such referential ambiguity?

Researchers have long assumed that children cope with such ambiguity in part by taking into account additional

referential contexts in which the same word occurs (e.g., Fazly, Alishahi, & Stevenson, 2010; Fisher, Hall, Rakowitz, & Gleitman, 1994; Pinker, 1984; Siskind, 1996; Yu & Smith, 2007). Across occurrences of each word, aspects of the observed scene that are not relevant to a word’s meaning should be present less consistently than those that are central to its meaning. If children could detect which scene elements consistently co-occurred with each word across multiple observations, then this would help them to isolate the word’s meaning.

However, evidence suggests that solving this word-learning problem is not quite so simple, especially for verbs. Gillette, Gleitman, Gleitman, and Lederer (1999) devised the “human simulation” paradigm to test whether adults could use cross-situational consistency to identify the referents of words. Adult participants watched video clips of mothers interacting with toddlers. In each clip, the mother uttered a common noun or verb; the original soundtracks were removed and participants heard only a

* Corresponding author. Tel./fax: +1 209 228 4362.

E-mail address: rscott@ucmerced.edu (R.M. Scott).

tone marking the position of the target word. The task was to guess which word the mother had said. A series of such clips for each target word provided opportunities for cross-situational observation. Adults guessed the referents of 45% of the nouns under these circumstances, but correctly identified only 15% of the verbs. Performance with verbs improved dramatically, however, when participants were also given information about the sentence structures in which the verbs occurred (Gillette et al., 1999; Snedeker & Gleitman, 2004). Similar results were obtained with 7-year-old participants in the same kind of task (Piccin & Waxman, 2007). These findings suggest that scene observations are systematically less useful for learning verbs than for learning nouns. The referents of many concrete nouns can be identified via scene observation alone, but efficient verb learning requires support from sentence-structure cues. Further results suggested that the difficulty with learning verbs via scene observation stems from the fact that verb referents are more abstract (i.e. less “imageable”) than noun referents (Gillette et al., 1999), and therefore naturally harder to observe in scenes without linguistic guidance. In line with this hypothesis, the concreteness or imageability of particular words’ meanings predicts those words’ age of acquisition (Ma, Golinkoff, Hirsh-Pasek, McDonough, & Tardif, 2009; McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2008).

Sentence-structure cues can guide verb learning because syntax is systematically related to verb meaning (e.g., Bloom, 1970; Carlson & Tanenhaus, 1988; Dowty, 1991; Levin & Rappaport-Hovav, 2005). For instance, verbs that describe one participant acting on another can occur in transitive sentences, with two noun-phrase arguments (e.g., “She tickled her”), whereas verbs that describe actions involving only one participant tend to be intransitive, with one noun-phrase argument (e.g., “She laughed”). Children exploit such systematic relationships in verb learning: Toddlers infer that a new transitive verb refers to an event involving two participant-roles rather than one (e.g., Fisher, 2002; Naigles, 1990; Naigles & Kako, 1993; Yuan, Fisher, & Snedeker, *in press*), for example, and use transitive word order to map a new verb onto an event in which the subject of the sentence plays an agent’s role (Gertner, Fisher, & Eisengart, 2006). Moreover, many verbs occur in more than one sentence structure. The *set* of sentence structures in which each verb occurs yields more refined information about each verb’s meaning, even to young verb-learners (e.g., Naigles, 1996; Scott & Fisher, 2009; Snedeker & Gleitman, 2004).

However, within the guidelines given by syntactic cues, the recruitment of scene observations remains essential to verb learning. Different sentence structures vary in the number, type, and arrangement of constituents that serve as arguments of the verb; thus sentence-structure cues can only convey aspects of verb meaning relevant to the number and type of participant-roles involved – what has been termed a verb’s *semantic structure* – rather than its *semantic content* (Grimshaw, 1993). For example, sentence structures can tell learners how many participant-roles a new verb’s meaning involves, and provide abstract information about what kinds of roles these might be (e.g., agent, patient, goal), but the verb’s semantic content must

be inferred via observation of relevant events. Observing a verb’s syntactic behavior yields, not a unique verb meaning, but a class of verb meanings that can range broadly in semantic content (Fillmore, 1970; Gleitman, 1990; Levin, 1995; Pinker, 1994). As Gillette et al. (1999) pointed out, “there is no ‘hot syntax’ and ‘cold syntax’ that will differentiate *burn* from *freeze*” (p. 170).

Thus returning to the playground scene with which we began, we see that even with the aid of sentence-structure cues, the learner faces a problem. In our example, the child could use the intransitive sentence context to determine that *pim* describes a noteworthy state or activity of one participant. But which one? In principle, there remain many ways to characterize the role of an event participant. Although a structural constraint has been added, the problem of word mapping remains within those constraints. Ultimately, learners must combine observations across sentences and events to determine both the semantic structure and content of a new verb.

Here we ask whether children can use cross-situational consistency to resolve this residual ambiguity in verb meaning. Exploiting this information source is far from trivial. In order to benefit from cross-situational consistency in word learning, children would have to establish a lexical entry for a word and attach to it information about that word’s potential referents, without knowing which aspects of the scene the word described. They would then have to retrieve this entry when they encountered the word again and modify or update it in response to new referential contexts.

Recent experiments have shown that, despite these challenges, infants can use cross-situational consistency to learn novel *object labels* under referential uncertainty, at least under some circumstances (Smith & Yu, 2008; Yu & Smith, 2011). Such evidence has been interpreted as support for models proposing that statistical information about the co-occurrence of words and referents plays a powerful role in word learning (e.g., Fazly et al., 2010; Siskind, 1996; Yu & Smith, 2007). However, object labels represent only a fraction of the vocabulary and arguably represent the simplest case of word learning from scene observation (e.g., Gillette et al., 1999; Waxman & Lidz, 2006). Here we adapt the same experimental technique to ask whether the use of cross-situational consistency in word learning extends beyond this simplest case. That is, we ask whether young children can use information derived from cross-situational consistency to identify the referents of verbs.

We next review the recent work with nouns and outline what would be required for children to succeed in a similar learning task with verbs. We then present experiments in which 2.5-year-olds encounter novel intransitive (Experiment 1) and transitive verbs (Experiment 2) under referential uncertainty. In these experiments, cross-situational consistency was a necessary source of information about the intended referent of each verb.

Smith and Yu (2008) tested 12- and 14-month-olds’ ability to use cross-situational observation to learn novel nouns. Infants saw a series of training trials in which pictures of two novel objects were presented side by side; each pair of objects was accompanied by two novel labels.

On each trial, it was impossible to tell which novel label went with which novel object. Across trials, however, each of six novel words consistently co-occurred with only one object. The training trials were followed by test trials in which infants heard one label and saw its referent and a distracter object. Infants looked longer at the referent object, suggesting that they had learned the labels by keeping track of which word consistently accompanied each picture.

Smith and Yu's (2008) findings allow us to conclude that 12- to 14-month-olds possess the basic mechanisms required to benefit from cross-situational consistency in learning names for objects: Infants were able to attach some information about potential object referents to each new word's lexical entry, even when the referential context contained more than one possible referent. They were then able to update this information across trials to identify correct object-noun pairings. Here we ask whether these same basic mechanisms also support verb learning. Can children use cross-situational observation to learn verbs? In order to do so, they must overcome several additional challenges that verbs pose.

First, as discussed above, there is strong evidence that observation of the extra-linguistic context alone is less useful for learning verbs than it is for learning nouns (Gillette et al., 1999). However, adults were better able to make use of scene information when provided with information about the sentence structures in which the verbs occurred (Gillette et al., 1999; Snedeker & Gleitman, 2004; Snedeker, Li, & Yuan, 2003). For this reason, we provided children with relevant sentence-context information. We assume that, as with adults, sentence-context information should help to constrain the range of interpretations children consider, thereby facilitating the use of cross-situational consistency within that range. In addition, results from the human simulation paradigm suggest that adults used scene information to identify concrete verbs (e.g., *throw*, *hammer*) much more often than abstract verbs (e.g., *think*, *know*; Gillette et al., 1999; Snedeker & Gleitman, 2004); a similar difference was found between concrete (e.g., *spoon*) and abstract nouns (e.g., *thing*; Gillette et al., 1999). From these results we can conclude that cross-situational observation should be most useful for learning concrete verbs, simply because their referents are relatively easy to observe in scenes. Thus, we focused here on children's ability to use cross-situational observation to learn verbs that describe concrete actions.

Second, encoding referential information under uncertainty is likely to be more difficult for verbs (even concrete verbs) than for nouns. In order to use cross-situational consistency to learn the nouns in Smith and Yu's (2008) experiment, infants first had to encode information about the potential referents that accompanied the words on each trial. In Smith and Yu's study the referents were static pictured shapes. The referents of verbs, however, are more complex. Verbs' meanings are relational (e.g., Gentner, 1978; Gleitman, 1990). They denote the roles participants play in an event, often including action or change over time. In the playground scenario with which we began, for example, a child encounters a new verb in the presence

of multiple dynamic actions enacted by different people (e.g., *swinging*, *climbing*, *sliding*). In order for this child to make use of cross-situational consistency, as did the infants in Smith and Yu's experiments, he must first encode something about the potential event referents that accompanied *pimming*. To date there is no evidence that children can attach *event* referential information to a verb's lexical entry when multiple candidate referent events are present.

Third, assuming that children can encode information about potential event referents under ambiguity, they must then retrieve and update this information based on subsequent observations. The added complexity of event as opposed to object referents might complicate the retrieval and comparison of referential options (e.g., Gentner, 1988). The infants in Smith and Yu's (2008) experiments saw the same pictured object tokens repeated across both the training and test trials. In order to keep track of the co-occurrence patterns, they only needed to remember whether a particular label had occurred with a particular object token in the past. This represents a way in which Smith and Yu's task simplified the real-world problem of noun learning. Children learn names for categories of objects (*animals*, *dogs*, *pencils*), and therefore must abstract over within-category variability to work out the meanings of nouns across multiple labeling events (e.g., Xu & Tenenbaum, 2007). This experimental simplification is (even) less defensible for verb learning. Whereas children do encounter repeated tokens of the same object category in real life (e.g., their own bottle), they see repeated *action* tokens only if the original action was captured on video. Actions vary when instantiated by different actors and objects, and even the same actor is unlikely to perform an action in precisely the same way twice. Therefore, to learn verbs via cross-situational observation, children must be able to compare different event tokens and abstract across irrelevant variation.

The evidence is mixed regarding children's ability to abstract across irrelevant variation when learning verbs. Some studies show that when presented with a series of training exemplars, each of which is clearly labeled with the same verb, children can generalize the new verb to new exemplars of the same action (Piccin & Waxman, in preparation; Waxman, 2009; Waxman, Lidz, Braun, & Lavin, 2009); moreover, children can compare distinct event tokens to draw inferences about verb meaning (e.g., Behrend, 1995; Childers, 2008). For example, Childers (2008) showed 2.5-year-olds a series of training events depicting a consistent manner of action (with varying result) or a consistent result (with varying action); each event was enacted in isolation, and accompanied by the same novel verb (e.g., "I'm dacking it!"). Children were then given a set of objects and asked to enact the action named by the novel verb. Children's enactments tended to preserve elements that had been consistent across training events. This suggests that children assumed that event components that consistently occurred with a verb were central to its meaning, whereas components that varied were not; this is the kind of comparison process that is required to benefit from cross-situational consistency in verb learning (see Akhtar & Montague, 1999, for evidence with adjectives).

However, other studies have shown that young children sometimes fail to identify consistent actions across different event tokens (e.g., Imai, Haryu, & Okada, 2005; Imai et al., 2008). For instance, Maguire, Hirsh-Pasek, Golinkoff, and Brandone (2008) found that 2.5-year-olds who learned a novel verb by seeing four distinct event tokens involving different actors had more difficulty extending that verb to relevant new events than did those trained with a single repeated event token. Taken together, these two sets of studies suggest that while children are capable of detecting consistency in the face of irrelevant variation, this process is difficult for them. Thus, it is possible that children would be unable to detect consistency across varying event tokens when faced with the additional challenge of encoding and comparing events under referential uncertainty.

In sum, the existing evidence for young children's sensitivity to cross-situational consistency in word learning is limited to the simplest case, that of learning labels for objects. Learning verbs, even concrete action verbs, requires learners to encode, retrieve, and compare representations of events rather than objects. The present experiments explored whether, despite these challenges, the basic mechanisms that make cross-situational learning possible can also support verb learning. Given the additional difficulties associated with using cross-situational consistency to learn verbs, we decided to test 2.5-year-olds rather than 12-month-old infants. The results reviewed above suggest that while 2.5-year-olds possess the basic comparison process required for cross-situational verb learning (e.g., Childers, 2008), they also have difficulty detecting consistency across events in the face of irrelevant variation (e.g., Maguire et al., 2008). Thus, this age group provides a good test case for the feasibility of using cross-situational consistency in verb learning.

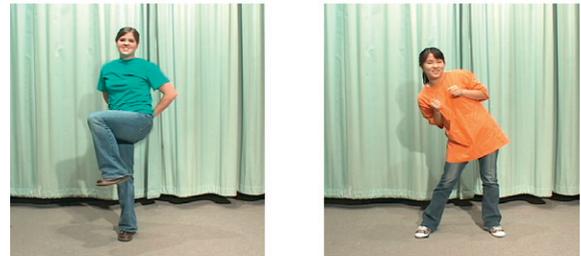
2. Experiment 1

In Experiment 1, we tested whether 2.5-year-old children could (1) attach referential information to novel intransitive verbs when multiple potential referents were present and then (2) use cross-situational consistency to identify each verb's intended referent. We adapted Smith and Yu's (2008) paradigm for use with novel verbs. On each trial, two novel one-participant actions, each performed by a different actor, were presented along with two novel intransitive verbs (e.g., "Look, she's pimming. And she's nading."); Fig. 1). Each potential event-referent was a concrete action (see Table 1 for action descriptions). Children saw multiple tokens of each action, each performed by a different actor. The novel verbs were presented in sentences that identified the new words as intransitive verbs, but provided no information as to which verb described which action. The only cue to each verb's intended referent was the consistency with which it accompanied the same action across trials.

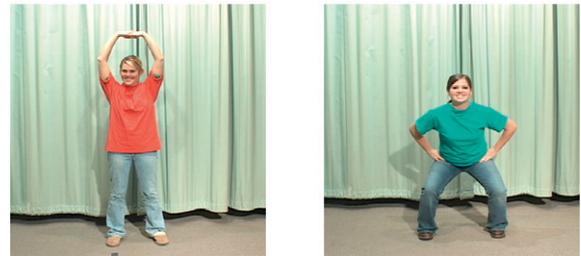
Our design differed from that of Smith and Yu (2008) in several ways. Smith and Yu presented children with six novel object-labels; we reduced the complexity of the task slightly by presenting four novel verbs. Also, whereas Smith and Yu presented novel lexical items in isolation,



Trial 1: "She's pimming. And she's nading."



Trial 2: "She's tazzing. And she's pimming."



Trial 3: "She's nading. And she's rivving."

Fig. 1. Three sample trials from Experiment 1 with accompanying audio.

Table 1

Novel verbs and actions used in Experiment 1.

Verb	Action
Pimming	Tilting torso side to side
Nading	Squatting
Tazzing	Lifting knee
Rivving	Stretching arms over head

we chose to present the novel verbs in sentences. As discussed above, sentences provide information about a verb's semantic structure, thus narrowing the range of potential meanings. This has been shown to facilitate adults' use of scene information in verb learning.

In addition, unlike Smith and Yu (2008), we did not include separate test trials. Instead, we employed a continuous study-test procedure (e.g., Medina, Trueswell, Snedeker, & Gleitman, 2011; Smith, Smith, & Blythe, 2011). Within each trial, we examined where children looked when they heard each novel verb. This allowed us to see how children's interpretation of the verbs emerged over time. We predicted that if children could keep track

of which event consistently occurred with each verb, then as the session progressed they would come to look longer at the matching referent than at the non-matching referent for each verb.

Finally, we sought to explore the relationship between children's productive vocabularies, assessed via parental report, and their performance in this task. Success in our task required children (1) to integrate incoming linguistic information with previously-encountered referential information in order to identify the correct referents for each verb and (2) to shift their attention from one event to the other as they encountered the second verb in each trial. Children with higher productive vocabularies identify and respond to spoken words more quickly and accurately than do children with lower productive vocabularies (Fernald, Perfors, & Marchman, 2006; Fernald, Swingley, & Pinto, 2001; Zangl, Klarman, Thal, Fernald, & Bates, 2005). Given such findings, it seemed plausible that high-vocabulary children might be faster or more successful at identifying each referent, and shifting attention between referents, thereby outperforming low-vocabulary children in this task.

2.1. Method

2.1.1. Participants

Thirty-six 2.5-year-olds participated in the experiment (mean age = 31.2 months, range 30.1–33.1, 18 male, 18 female). All were native speakers of English. An additional 4 children were tested but eliminated because they became fussy (1), failed to complete the experiment (2), or spent over 85% of the experiment looking at one screen (1). Half of the children were assigned to verb order 1 and half to verb order 2 (see Procedure below). Equal numbers of boys and girls were assigned to each verb order. Children's productive vocabularies were measured using the short form of the MacArthur–Bates Communicative Development Inventory, Level 2 (Fenson et al., 2000). Vocabulary scores ranged from 42 to 100 with a median of 80.5. To explore the relationship between vocabulary scores and task performance, we divided the children into high- and low-vocabulary groups by a median split (high-vocabulary group: $n = 18$, median = 94.5, range 81–100; low-vocabulary group: $n = 18$, median = 65.5, range 42–80).

2.1.2. Apparatus

Children sat on a parent's lap facing two 20-in. television screens placed about 30 in. away. The screens were 12 in. apart and about at the children's eye level. Soundtracks were played from a concealed central speaker. A camera hidden between the two screens recorded the children's eye movements during the experiment. Parents wore opaque sunglasses, preventing them from biasing their children's responses.

2.1.3. Materials and procedure

Materials consisted of color videos of four female actors performing novel solo actions. There were four novel actions, each paired with a novel intransitive verb (see Table 1). Each woman performed three of the four novel actions for a total of 12 unique event tokens. Tokens were

presented in synchronized pairs and accompanied by a soundtrack recorded by a native English speaker.

Children were tested with a continuous study-test procedure consisting of 12 trials (see Appendix A for a complete trial sequence). On each trial, two event tokens, each performed by a different actor, were presented simultaneously for 8 s. While viewing the events, children heard two intransitive sentences, each containing a novel verb (see Fig. 1). All sentences were of the form "She's *verbing*." The use of the pronoun subject *she* ensured that each sentence could refer to either event. The first sentence in each trial began with one of a small set of attention-getting words or phrases (e.g., *Look, Wow*); the second sentence began with *And*, and was recorded with contrastive stress on both the subject pronoun and the verb (e.g., "Look, she's pimming. And she's nading!").¹ The stimulus sentences were aligned such that the onset of the first sentence coincided with the onset of the trial, with the onset of the novel verb 1.7 s later on average; the onset of the second sentence began 4 s into the trial, with the onset of the novel verb 1 s later on average. Trials were separated by a 3-s silent blank-screen interval.

The experiment was divided into two blocks of six trials. Each of the 12 event tokens occurred once in the first block and once in the second block. The fact that event tokens were never repeated in the first block of trials enabled us to assess the effect of token variability on children's use of cross-situational consistency. If children required exact repetition of event tokens in order to detect cross-situational consistency, then they should not be able to identify the verbs' referents until they encountered the tokens again in block 2; this would result in better performance in block 2 than in block 1.

Within each block, children saw all possible combinations of actions once (e.g., the tilting-torso event in Table 1 paired once with squatting, once with arm-stretching, and so forth). Event tokens were paired and presented in the same sequence for all participants (see Appendix A). This sequence was randomized with the following constraints: the same actor never appeared on both screens at once, no action occurred more than two trials in a row, and each action appeared an equal number of times on the right and left screens. Across the 12 trials, each action was labeled six times (three times in each block). The action on the left screen was labeled first in half the trials; also, each action was labeled first in half the trials in which it occurred.

We counterbalanced the order in which the two verbs were presented within trials. Children were assigned to one of two verb orders that differed in which verb occurred first. For instance, in trial 1, children in verb order 1 heard "Look, she's pimming. . . And she's nading," whereas children in verb order 2 heard "Look, she's nading. . . And she's pimming." Thus, for each trial, half of the children heard

¹ We used contrastive stress in the second sentence because it made the sentences sound more felicitous to adults. In principle, the stress could have indicated to the children that the first sentence described one event while the second sentence described the other. Four-year-olds have been shown to interpret contrastive stress appropriately (Arnold, 2008); this raises the intriguing possibility that even these younger children (2.5-year-olds) might have used the prosody as a source of contrast information.

one event labeled first, while the other half heard the other event labeled first. Left–right positioning of event tokens was counterbalanced within verb order. This ensured that success in this task could not be due to a simple preference for a particular side, or a strategy such as looking at the left and then the right screen during each trial (as would be expected if children were “reading” from left to right as they heard the sentences). Any such preference or strategy would result in chance performance when averaged across trials.

2.1.4. Analyses

We coded where children looked (left screen, right screen, away), frame by frame from silent video. To assess reliability, eight children’s data were independently coded by a second coder. The two coders agreed on the children’s direction of gaze for 98% of coded video frames.

Within each trial, we analyzed children’s visual fixations within a 2.5-s time window that began at the onset of each novel verb (as each trial contained two novel verbs, there were two 2.5-s analysis windows per trial). This choice of analysis window was based on the timing of the trials. A 2.5-s analysis window ensured that the first verb window always ended before the onset of the second verb window, and the second verb window always ended before the trial ended.

Because of the time it takes to program an eye movement, analyses of visual fixations have traditionally focused on analysis windows that begin 200 ms after word onset (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), or even 300 ms or more in experiments with young children (e.g., Fernald et al., 2006). However, we suspected that in our task children might often initiate meaningful eye movements prior to 200 ms after verb onset. If children inferred that both actions were labeled in each trial, then they might sometimes switch their gaze to the second event as soon as they heard “And she’s...”. For this reason, we chose to use analysis windows that began at the onset of the verb; however, the main results reported below remained the same in analysis windows that began 200 ms after verb onset.

Analysis windows in which the children looked away from the screens for more than 67% of the 2.5-s interval were dropped from the analyses (30 windows, 3.5% of the 864 possible windows). For the remaining 834 analysis windows, we computed children’s looking time to the correct referent for the novel verb (the target event) and the non-matching event (the distracter event).

Children had no cross-situational information about each verb’s meaning the first time they encountered it; therefore we expected children to show no systematic preference for the target event during the first observation window for each verb. In principle, however, children could have used an assumption of mutual exclusivity to infer the meanings of the verbs that were introduced for the first time in the second and third trials (e.g., Golinkoff, Jacques, Hirsh-Pasek, & Nandakumar, 1996; Markman & Wachtel, 1988; Merriman, Marazita, & Jarvis, 1993, 1995). To illustrate, consider the second trial shown in

Fig. 1. Because children already had some evidence that *pimming* referred to the torso-tilting motion, they could have inferred that *tazzing* must refer to the knee-lifting action. However, preliminary analyses of children’s looking times to the target and distracter events on the first observation, averaged across all verbs, showed no preference for the target event, $t < 1$ (target looking-time $M = 1.26$ s, $SD = .50$ s; distracter looking-time $M = 1.20$ s, $SD = .51$ s). This was also true for the verbs introduced for the first time in the second and third trials (i.e. *tazz* and *rivv*), $t(35) = 1.08$, $p = .29$. We therefore omitted the first observation window for each verb and conducted our main analyses on the second through sixth presentations of each verb.

Preliminary analyses revealed no effects of sex or verb order; these factors were not examined further.

2.2. Results and discussion

We begin with descriptive statistics that yield an overview of the time course of children’s looking patterns in the task. In order to succeed in the present task, children needed to look at one event during the first half of the trial, and then switch to the other event upon encountering the second verb in the second half of the trial. This pattern would indicate that children understood the structure of the task, and also reflect emerging identification of the referents of individual verbs.

The time course of children’s looking patterns, shown in Fig. 2, shows just this predicted pattern. Time is displayed on the x axis in increments of 33 ms (i.e. 1/30th of a second or a single video frame). The y axis shows the proportion of trials in which children fixated each location (verb1 target, verb2 target, away), averaged over participants. At the onset of the trial, the children were equally likely to fixate each of the two events. Approximately 200 ms after the average onset of the first verb, there was an increase in the proportion of fixations to the target referent for the first verb. Looks to the verb1 target persisted throughout the first analysis window. Approximately 100 ms after the average onset of the second verb, children shifted their attention to the target referent for the second verb. Thus, children focused their attention on the verb1 target during the first analysis window and then shifted their attention to the verb2 target during the second analysis window. This pattern suggests that children understood the structure of the task, and that they benefited from cross-situational consistency.

Table 2 shows the average looking time to the target and distracter events within the 2.5-s analysis window, as well as an average target-advantage score reflecting the difference between them (target looking-time minus distracter looking-time). To test whether children reliably made use of cross-situational information in verb learning, we analyzed children’s looking times using a $2 \times 4 \times 2 \times 2$ mixed-model analysis of variance (ANOVA) with block (1,2), verb (*pim*, *nade*, *rivv*, *tazz*), and event (target, distracter) as within-subject factors and vocabulary group (high, low) as a between-subjects factor. The analysis

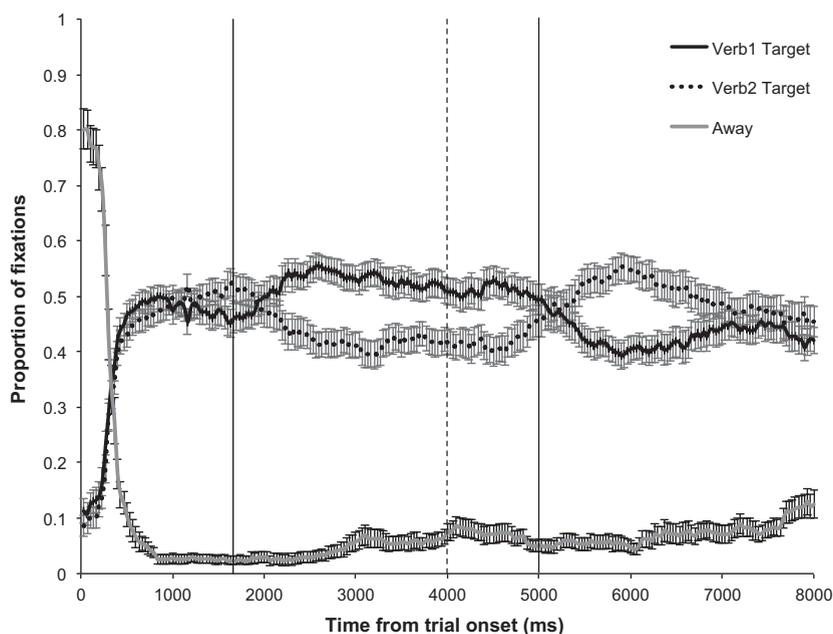


Fig. 2. Proportion of fixations to verb1 target, verb2 target, and away in Experiment 1. Solid vertical lines indicate the average onset of verb1 and verb2. The dashed vertical line indicates the onset of the second sentence. Error bars represent standard errors over participants.

Table 2

Mean (SD) looking times to target and distracter events and target-advantage scores (time to target minus time to distracter) in seconds for Experiments 1 and 2.

	Target event	Distracter event	Target-advantage score
Experiment 1	1.28 (.22)	1.08 (.22)	.21 (.41)
Experiment 2 – high-vocabulary	1.36 (.18)	1.08 (.15)	.28 (.32)
Experiment 2 – low-vocabulary	1.19 (.20)	1.25 (.17)	–.07 (.37)

revealed a significant main effect of event, $F(1,34) = 7.75$, $p = .009$, $\eta^2 = .19$; as shown in Table 2, children looked reliably longer at the target event than at the distracter event. The analysis also revealed a significant main effect of block, $F(1,34) = 6.76$, $p = .014$, $\eta^2 = .17$, which reflects children's tendency to look away more, and thus less at either the target or distracter event, in block 2 (look-away time $M = .17$ s, $SD = .19$ s) than in block 1 ($M = .09$ s, $SD = .12$ s). This analysis revealed no significant effect of verb, $F(3,102) = 1.74$, $p = .16$, or vocabulary group, $F(1,34) = 2.52$, $p = .12$. A marginal interaction of verb and vocabulary-group, $F(3,102) = 2.54$, $p = .06$, $\eta^2 = .07$, suggested that the tendency to look away from both video events varied across verbs and vocabulary group. No other effects were significant, all $F_s < 1$.

Children's significant preference for the target event indicated that, despite the referential ambiguity within each trial, children were able to gather information about the events that co-occurred with each verb and update that information across trials to identify the verb's referent. The fact that children's preference for the target event did not interact with block suggests that children did not require exact repetition of event tokens in order to benefit from cross-situational consistency (as this would have led them to perform better in block 2 than in block 1). In addition,

the lack of an interaction of event and vocabulary group suggests that this preference was similar across vocabulary groups. Follow-up analyses confirmed that vocabulary was not significantly correlated with target-advantage scores, $r = .22$, $p = .19$. Thus, children's vocabulary level did not affect their ability to benefit from cross-situational consistency in this experiment.

Furthermore, the absence of a verb by event interaction suggests that learning about multiple verbs contributed to children's significant preference for the target event. Inspection of the means revealed that the target-advantage scores for all four verbs were positive, showing that the preference for the target over the distracter event was roughly consistent across items. We also examined whether individual children identified multiple verb-referent pairs during this brief task. Following Yu and Smith (2011), we considered a child to have identified a particular verb-referent pair if that child looked longer at the matching event than at the distracter event, averaged across observations 2 through 6 for that verb. Based on this criterion, the median number of verb-referent pairs identified was 3 (of 4). Among the 36 children, 6 identified all 4 pairs, 13 identified 3, 12 identified 2, 4 identified 1, and 1 child identified none of the pairs. This distribution of verbs identified differed significantly from the distribution

expected by chance, $\chi^2(2, N = 36) = 8.98, p = .012^2$. This suggests that 2.5-year-olds managed to sort out which events went with *multiple verbs* in interleaved trials, even though they never received an unambiguous labeling trial for any verb.

Finally, we conducted a separate analysis examining the two analysis windows within each trial. Recall that we omitted the first observation for each verb from our analyses. As a result, for any particular child, there were some verbs that did not appear in both windows within the first block (though all verbs occurred in both windows when averaged across children). This factor could therefore not be included in the main analysis with block and verb. A $2 \times 2 \times 2$ mixed-model ANOVA with analysis window (1,2) and event (target, distracter) as within-subject factors and vocabulary group (high, low) as a between-subject factor again revealed a significant main effect of event, $F(1,34) = 8.56, p = .006, \eta^2 = .20$, reflecting children's reliable identification of the novel verbs' referents. The analysis also revealed a significant main effect of window, $F(1,34) = 4.72, p = .037, \eta^2 = .12$, reflecting children's tendency to look away more during the second window (look-away time $M = .15$ s, $SD = .16$ s) than during the first ($M = .09$ s, $SD = .11$ s) in each trial. There was no significant interaction of window and event, $F < 1$. Thus, as suggested by the switching pattern in the time-course plot in Fig. 2, children's preference for the target event did not differ reliably across the two analysis windows within the trial. No other effects were significant, all F s < 2.70 , all p s $> .11$.

3. Experiment 2

In Experiment 1, 2.5-year-olds used cross-situational consistency to identify the referents of novel intransitive verbs under referential uncertainty. In Experiment 2, we sought to explore the limits of this ability by testing whether children could also identify the referents of transitive verbs under these circumstances. Transitive verbs present several additional sources of potential difficulty for children. First, because transitive sentences contain an additional noun-phrase argument, they might impose a greater language-processing load than the simple intransitive sentences of Experiment 1. Second, transitive verbs are correspondingly more complex in their semantic structures than intransitive verbs, because they involve two participant-roles rather than one. This might complicate the encoding, retrieval, and comparison of referential information in our task. For example, in order to learn a transitive verb, children need to abstract not only across actions performed by different actors, but also across variation in the objects acted upon and in the details of the action when enacted on different objects (cf. Imai et al., 2008; Waxman et al., 2009). This additional variability may make detecting cross-situational consistency more

challenging. Thus, this experiment should provide an initial hint as to how well children's ability to attach referential information to verbs under uncertainty scales up to more difficult learning conditions.

In each trial, children saw two different actors, each performing a novel action on a distinct object. The events were accompanied by two novel transitive verbs (e.g., "Look, she's nading her toy...and she's pimming her toy!"; Fig. 3). Across trials, the actor and object involved in the action varied. As in Experiment 1, the only cue to the verb's referent was the consistent pairing of the verb with a particular action. If children can use cross-situational comparison to identify the referents of transitive verbs under uncertainty, then when presented with a verb, they should look longer at the event that consistently co-occurred with that verb across trials.

3.1. Method

3.1.1. Participants

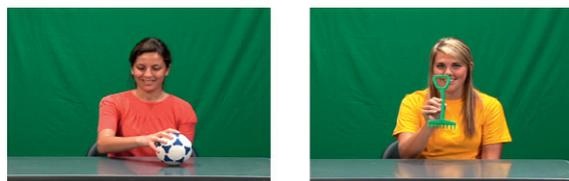
Thirty-six 2.5-year-olds participated (mean = 30.9, range 29.9–33.1, 18 male, 18 female). All were native speakers of English. An additional three children were tested but eliminated because they were fussy (1), distracted (1), or because their overall target-advantage score was more than 3.5 SD from the mean (1). Equal numbers of children were randomly assigned to each of four combinations of experimental list and verb order (see Procedure below). Equal numbers of boys and girls were assigned to each of these four combinations. Vocabulary scores, measured as in Experiment 1, ranged from 15 to 100 with a median of 82.5. Children were again divided into two vocabulary groups based on a median split (high vocabulary group: $n = 18$, median vocabulary score = 98, range 84–100; low-vocabulary group: $n = 18$, median score = 68, range 15–81).

3.1.2. Apparatus

The apparatus was identical to that used in Experiment 1.

3.1.3. Materials and procedure

Materials consisted of color videos of four female actors performing novel actions on one of four different toys (a small stuffed giraffe, a plastic toy rake, a small toy tractor, and a ball); the pairing of actor and toy was consistent across trials. There were four novel actions, each paired with a novel transitive verb. Unlike in Experiment 1, the pairing of verb and action varied across experimental lists



Trial 1: "Look! She's nading her toy...and she's pimming her toy."

Fig. 3. Sample trial from Experiment 2 with accompanying audio.

² This analysis was based on three categories (0 or 1 pairs identified, 2 pairs, 3 or 4 pairs) rather than the five possible categories because the expected values for 20% of the original five categories were less than 5, which renders chi-square significance-testing suspect. The expected values for these three categories were based on the binomial probabilities of each outcome.

(see Table 3). We added the list factor so that if we were to find systematic differences between items, we would be able to determine whether these resulted from idiosyncrasies of particular words or of particular actions.

Each actor performed three of the four novel actions for a total of 12 unique event tokens. Event tokens were presented in synchronized pairs accompanied by a soundtrack recorded by a native English speaker. Left–right positioning of event tokens was counterbalanced within verb order and list. As in Experiment 1, 4 novel verbs were presented in a continuous study–test procedure consisting of 12 ambiguous trials, each involving two actions and two verbs (for a complete sequence, see Appendix A). Each of the 12 unique event tokens was presented once in the first block of 6 trials, and then again in the second block of six trials; thus, again, comparing performance across blocks allowed us to assess children’s ability to abstract across actors and objects.

The procedure was identical to that of Experiment 1 with three exceptions. First, the continuous study–test procedure was preceded by two display trials. On the first display trial, children saw videos of two of the female actors, each holding her toy and smiling, and heard a soundtrack calling attention to each actor and her toy (e.g., “Oh look! She has a toy! And she has a toy!”). After a 3-s blank-screen interval, children saw a second display trial in which the other two female actors and their toys were introduced in the same manner. Each display trial lasted 8 s. The purpose of these trials was to familiarize children with the actors, their toys, and the fact that the objects would be referred to as “toys” throughout the experiment. Second, during the continuous study–test phase, children heard the novel verbs used in transitive rather than intransitive sentences. All sentences were of the form “She’s *verbing* her toy”; the second sentence in each pair was recorded with contrastive stress on the subject, verb, and possessive pronoun (e.g., “She’s *pimming* her toy. And she’s nading her toy!”). The use of the pronoun “she” and the generic noun “toy” ensured that the sentences were ambiguous and could refer to either event. Third, in Experiment 2, the stimulus sentences were aligned such that the onset of the first novel verb occurred 1 s into each trial, with the attention-getting phrase at the start of the first sentence (e.g., “Look!”) occurring slightly before trial onset; the onset of the second novel verb occurred 5 s into

each trial, with the onset of the second utterance falling on average 4.1 s into the trial. This change kept the trial timing similar to that of Experiment 1, but simplified the computation of analysis windows across trials.

3.1.4. Analyses

As in Experiment 1, we coded where children looked (left screen, right screen, away) frame by frame from silent video. Sixteen children’s data were independently coded by a second coder. The first and second coders agreed on the children’s direction of gaze for 98% of coded video frames.

Within each trial, we examined children’s fixations within two 2.5-s analysis windows that began at the onset of each novel verb. As in Experiment 1, analysis windows in which the children looked away from the screens for more than 67% of the 2.5-s interval were dropped from the analyses (20 analysis windows, 2.3% of the 864 possible windows). We then calculated looking times to the target and distracter events, as well as target-advantage scores (target looking-time minus distracter looking-time), for each analysis window. Preliminary analyses of children’s looking times on the first observation of each verb again showed no significant preference for the target. In fact, averaged across all four verbs, children demonstrated a significant preference for the distracter on the first observation, $t(35) = -2.22$, $p = .033$, $d = .37$ (target looking-time $M = 1.09$ s, $SD = .44$ s; distracter looking-time $M = 1.41$ s, $SD = .44$ s). This baseline preference for the distracter is difficult to interpret. Given our counterbalancing, it could not be due to a general preference for a particular video, so it is likely to reflect noise from chance fluctuations in children’s preferences at the start of the experiment. For the verbs introduced for the first time in the second and third trials, children did not show a significant preference for either event, $t < 1$.

Preliminary analyses revealed no effects of sex, list, or verb order. These factors were not examined further.

3.2. Results and discussion

We again begin with descriptive analyses characterizing children’s looking patterns across the trial. Fig. 4 shows the time course of children’s looking patterns. As will be discussed below, analyses for Experiment 2 revealed a significant difference between the vocabulary groups. Therefore, fixations are plotted separately for children whose vocabulary scores were above the median (high-vocabulary group; Fig. 4A) and children whose vocabulary scores were below the median (low-vocabulary group; Fig. 4B).

Children in the high-vocabulary group performed similarly to the children in Experiment 1. Within the high-vocabulary group, fixations to the verb1 target began to increase about 800 ms after the onset of the first verb. A modest preference for the verb1 target persisted until about 400 ms before the onset of the second verb (i.e. approximately 500 ms after the average onset of the phrase *and she’s...*), at which point there was a sharp

Table 3
Novel verbs and actions used in Experiment 2, separately by list.

Verb	Action	
	List 1	List 2
Pimming	Sliding in an S on table-top	Spinning on table-top
Nading	Lifting and lowering in one hand	Tossing from hand to hand
Tazzing	Tossing from hand to hand	Lifting and lowering in one hand
Rivving	Spinning on table-top	Sliding in an S on table-top

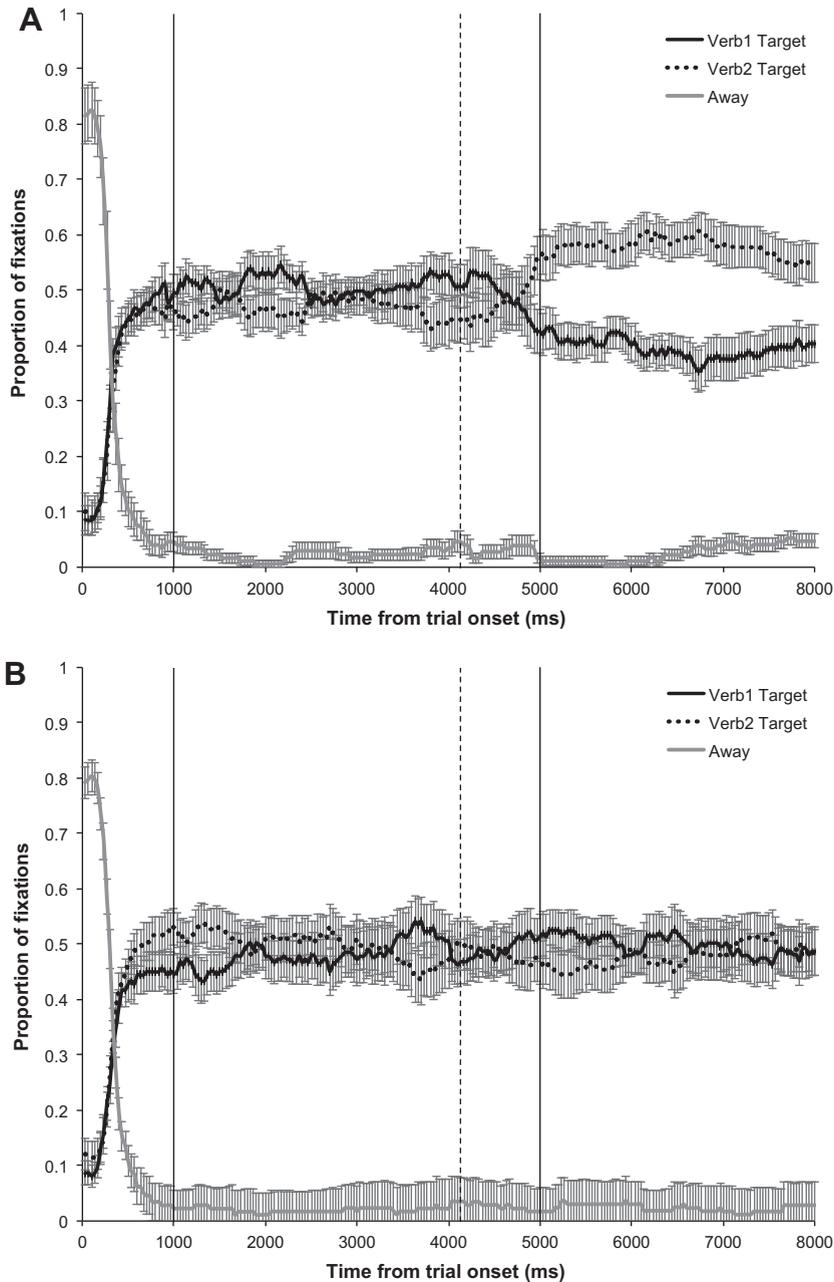


Fig. 4. Proportion of fixations to verb1 target, verb2 target, and away in Experiment 2, separately for high-vocabulary (4A) and low-vocabulary (4B) groups. The solid vertical lines indicate the onset of verb1 and verb2; the dashed vertical line indicates the average onset of the second sentence. Errors bars represent standard errors over participants.

increase in the proportion of fixations to the verb2 target. Children in the high-vocabulary group then maintained a strong preference for the verb2 target throughout the remainder of the trial. Thus, the high-vocabulary children appeared to understand the structure of the task, and to gather information about each verb's referent across trials.

The low-vocabulary group showed a different pattern. As shown in Fig. 4B, the low-vocabulary children did not

show a clear preference for the verb1 target during the first analysis window, nor did they switch to the verb2 target during the second analysis window. This pattern yields no evidence that the low-vocabulary children identified the referents of the novel verbs, or even learned that each event was labeled on each trial.

To test whether children reliably made use of cross-situational information in verb learning, we analyzed

children's looking times using a $2 \times 4 \times 2 \times 2$ mixed-model ANOVA with block (1,2), verb (*pim*, *nade*, *rivv*, *tazz*), and event (target, distracter) as within-subject factors and vocabulary group (high, low) as a between-subjects factor. As in Experiment 1, this analysis revealed a significant main effect of event, $F(1,34) = 5.96$, $p = .02$, $\eta^2 = .15$, reflecting a preference for the target event, and a significant main effect of block, $F(1,34) = 12.42$, $p = .001$, $\eta^2 = .27$, reflecting a tendency to look away more in block 2 (look-away time $M = .08$ s, $SD = .10$ s) than in block 1 ($M = .02$ s, $SD = .05$ s). In Experiment 2, however, this analysis also revealed a significant interaction of event and vocabulary group, $F(1,34) = 8.77$, $p = .006$, $\eta^2 = .21$, as well as a significant interaction of event and block, $F(1,34) = 5.69$, $p = .023$, $\eta^2 = .14$, and a significant three-way interaction of event, verb, and vocabulary group, $F(3,102) = 2.73$, $p = .048$, $\eta^2 = .07$. No other effects were significant, all F s < 1.76 , all p s $> .17$.

The presence of a significant event by vocabulary group interaction, as well as the three-way interaction of event, verb, and vocabulary group, suggests that the high- and low-vocabulary groups performed differently in this experiment. We also found a reliable positive correlation between vocabulary and overall target-advantage scores, $r = .38$, $p = .023$. Given these differences, we next conducted separate analyses for the high- and low-vocabulary groups.

3.2.1. High-vocabulary children

The looking times of the high-vocabulary children were analyzed using a $2 \times 4 \times 2$ repeated-measures ANOVA with block (1,2), verb (*pim*, *nade*, *rivv*, *tazz*), and event (target, distracter) as within-subject factors. There was a significant effect of event, $F(1,17) = 18.97$, $p = .0004$. As shown in Table 2, the high-vocabulary children looked reliably longer at the target event than at the distracter event. There were no significant interactions of event with either block, $F(1,17) = 2.35$, $p = .14$, or verb, $F(1,17) = 1.03$, $p = .39$. No other effects were significant, all F s < 2.73 , all p s $> .117$.

Thus the high-vocabulary children showed a pattern very similar to that found in Experiment 1. They demonstrated a reliable preference for the target event, indicating that they successfully used cross-situational consistency to determine the referents of novel transitive verbs. As in Experiment 1, the absence of a block by event interaction suggests that the high-vocabulary children did not require exact repetition of tokens in order to identify the verbs' likely referents. Thus, the high-vocabulary children were able to encode information about candidate referents for transitive verbs, despite variation in both the actor and the object in the events. The absence of a verb by event interaction suggests that multiple verbs contributed to the target preference shown by the high-vocabulary children. Consistent with this, examination of the target-advantage scores for individual verbs showed that all means were positive. As in Experiment 1, we examined the number of verb-referent pairs identified by each child, again counting a pair as identified if the child looked longer

at the matching event than at the distracter event, averaged across observation windows 2 through 6 for that verb. The median number of pairs identified by the high-vocabulary children was 3 (of 4). Of the 18 children, 3 identified all 4 pairs by this criterion, 9 identified 3, 4 identified 2, and 2 identified 1. This distribution of verbs identified was significantly different from the distribution expected by chance, $\chi^2(2, N = 18) = 10.68$, $p = .005$. Thus, these data suggest that the high-vocabulary children identified the referents of multiple interleaved transitive verbs under referential uncertainty.

As in Experiment 1, we also conducted a separate analysis comparing the two windows within the trial. A repeated-measures ANOVA with window (1,2) and event (target, distracter) as within-subject factors revealed a significant main effect of event, $F(1,17) = 13.76$, $p = .002$, $\eta^2 = .45$, and a significant interaction of window and event, $F(1,17) = 4.73$, $p = .044$, $\eta^2 = .22$. The main effect of window was not significant, $F < 1$. The high-vocabulary children's preference for the target event thus differed reliably across the two windows within each trial (window 1: target looking-time $M = 1.26$ s, $SD = .20$ s, distracter looking-time $M = 1.18$ s, $SD = .22$ s; window 2: target looking-time $M = 1.46$ s, $SD = .30$ s, distracter looking-time $M = .99$ s, $SD = .28$ s). This difference is visible in the time-course plot in Fig. 4A: the children more accurately found the target for the second verb in the trial than for the first. The reason for this difference, which did not appear in Experiment 1, is ambiguous. One possibility is that it reflects the increased difficulty of this task. If children were uncertain about the referent of the first verb in the trial, they might spend more time considering the two options. This would result in a weaker looking-time advantage for the target in the first analysis window. The second verb in each trial provides additional constraint (i.e. children now know that one action is *pimming* while the other is *rivving*), which may have contributed to the stronger performance in the second window. Another possible interpretation, however, is that during the first window the children divided their attention between the two events because they were anticipating hearing the second verb. This would reflect learning about the structure of the trial, and would have the secondary effect of reducing the absolute match preference in the first measurement window of each trial. The present data do not allow us to choose with confidence between these two interpretations.

3.2.2. Low-vocabulary children

As shown in Table 2, the performance in the low-vocabulary group was much less systematic than in the high-vocabulary group. A $2 \times 4 \times 2$ repeated-measures ANOVA with block (1,2), verb (*pim*, *nade*, *rivv*, *tazz*), and event (target, distracter) as within-subject factors revealed no main effect of event, $F < 1$. This suggests that, unlike the high-vocabulary children, the low-vocabulary children did not demonstrate a systematic preference for the target event. This analysis did reveal a significant effect of block, $F(1,17) = 10.55$, $p = .005$, $\eta^2 = .38$, indicating that the

low-vocabulary children looked away more during block 2 (look-away time $M = .08$, $SD = .08$) than during block 1 ($M = .01$, $SD = .03$). In addition, there was a marginal interaction of block and event, $F(1, 17) = 3.44$, $p = .081$, $\eta^2 = .17$, and a significant interaction of verb and event, $F(3, 51) = 2.83$, $p = .048$, $\eta^2 = .14$. No other effects were significant, all $F_s < 1$.

One might expect that low-vocabulary children needed more observations, or exact repetition of event tokens, in order to identify the verbs' referents in this more difficult task. If so, then they may have demonstrated evidence of learning in later trials. Further examination of the marginal block by event interaction effect suggests that this was not the case: the low-vocabulary group's performance deteriorated over the course of the experiment, rather than improving. In block 1, they demonstrated a small non-significant preference for the target event (target looking-time $M = 1.32$ s, $SD = .34$ s; distracter looking-time $M = 1.17$ s, $SD = .33$ s), $t < 1$. In block 2 they demonstrated a significant preference for the distracter event (target looking-time $M = 1.10$ s, $SD = .22$ s; distracter looking-time $M = 1.31$ s, $SD = .22$ s), $t(17) = -2.14$, $p = .047$, $d = .50$. Thus the low-vocabulary children performed, if anything, worse in the second block, suggesting that they were unable to identify the verbs' referents in this experiment even when they were provided with repeated event tokens.

The analysis for the low-vocabulary children also revealed a significant interaction of verb by event. Examination of the individual verbs revealed that for 3 of the 4 verbs, low-vocabulary children looked numerically longer at the distracter than at the target event. The one exception to this was *pim*, for which the low-vocabulary children demonstrated a preference for the target event (target looking-time $M = 1.38$ s, $SD = .35$ s; distracter looking-time $M = 1.07$ s, $SD = .34$ s). Given that the pairing of verb and action varied across lists, and preliminary analyses revealed no significant effects of list, the reason for this item effect is not obvious. However, the fact that target-advantage scores were negative for three of the four verbs again reflects the difficulty of this task for the low-vocabulary children. Examination of the data for individual children showed that the median number of verb-referent pairs identified by the low-vocabulary children was 2. No child identified all 4 pairs by this criterion, 3 children identified 3 of the pairs, 8 children identified 2, 5 children identified 1, and 2 children failed to identify any of the pairs. This pattern is not different from the distribution that would be expected by chance, $\chi^2(2, N = 18) = 1.79$, $p = .408$. Again, this suggests that the low-vocabulary children found this task much more difficult than did their peers in Experiment 1.

One possibility is that the low-vocabulary children's difficulty resulted from the increased linguistic demands in Experiment 2. We presented the verbs in transitive sentences on the assumption that this would help children to focus on relevant dimensions of the events (i.e. the action rather than the actor or the object; see Arunachalam & Waxman, 2011; Waxman et al., 2009) and

draw useful inferences about the verbs' likely meanings (e.g., Fisher, 2002; Naigles, 1990). However, presenting the verbs in sentences may also have made the task more difficult for the low-vocabulary children. While they were able to process the intransitive sentences in Experiment 1, the addition of an extra argument to the sentences in Experiment 2 may have overwhelmed their language-processing abilities, making it harder for them to identify each verb. This raises the intriguing possibility that the low-vocabulary children might have succeeded if the verbs were presented in isolation, or if pre-exposure to the novel verbs gave children a head start in the language-processing demands of the task (e.g., Arunachalam & Waxman, 2010; Scott & Fisher, 2009; Swingley, 2007; Yuan & Fisher, 2009). We are examining these questions in ongoing work.

It should be emphasized, however, that while presenting the verbs in sentences may have challenged the sheer word-processing abilities of the low-vocabulary children, this probably reflects challenges children face under typical learning conditions. Although verbs and other relational words do sometimes occur in isolation (e.g., *come*, *go*, *up*; Brent & Siskind, 2001), words typically occur in sentences (Aslin, Woodward, LaMendola, & Bever, 1996; Naigles & Hoff-Ginsberg, 1995). Moreover, verbs' appearance in appropriate sentence contexts is informative: Two-year-olds use verb transitivity to guide verb interpretation (e.g., Fisher, 2002; Naigles, 1990). The performance of lower-vocabulary children in Experiment 2 thus suggests that young children may, 'in the wild' as well as in our experiment, sometimes be unable to form a new lexical entry for an unknown verb that is encountered in a sentence, and link to that entry relevant event information. In addition, the contrast between the results of Experiments 1 and 2 suggests that differences in average sentence complexity between intransitive and transitive action verbs might influence the efficiency with which beginning word learners can exploit cross-situational information in verb learning.

Alternatively, the low-vocabulary children's difficulty may have resulted from the increased complexity of the events used in this experiment. In addition to ignoring variation in the actor performing the action, children also had to cope with variation in the object acted upon, and in the details of the action when enacted on different objects. It may be that the low-vocabulary children were unable to detect the cross-situational consistency in the face of this additional variability. One way to reduce the difficulty of this learning task, while still approximating typical learning conditions, might be to hold either the actor or the object constant across event tokens for each verb. After all, in real life children see the same actor perform a particular action on multiple objects (e.g., one's own mother drinking from a glass, a mug, and so forth), and they see particular objects acted on in the same way by multiple people (e.g., one's mother throwing a ball, one's father throwing the ball). This reduction in variability might enable low-vocabulary children to use cross-situational consistency to identify

the verbs' referents (e.g., Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008).

Finally, it is also possible that the low-vocabulary children's difficulty was not uniquely linguistic, but instead arose from limitations in other systems that are correlated with vocabulary level. Marchman and Fernald (2008) showed that vocabulary size is positively correlated with working memory, for example. If the low-vocabulary children in this experiment also tended to have lower working-memory capacity, then they may have had difficulty encoding, retrieving, and comparing the more complicated sentences and events they encountered in Experiment 2.

4. General discussion

When children encounter a new verb, the referential scene offers many possible interpretations. Here we investigated whether 2.5-year-olds could use cross-situational observation as one source of information for constraining these interpretations.

In Experiment 1, children readily used cross-situational consistency to identify the referents of intransitive verbs describing solo actions. Across ambiguous trials, when children encountered a verb, they looked toward the action that had consistently co-occurred with that verb. This preference for the target was consistent across blocks, suggesting they were able to tell which action consistently occurred with the verb even when they had to abstract across variation in the actor performing the action.³

Experiment 2 revealed a slightly different pattern for transitive verbs. Across ambiguous trials, children whose vocabulary was above the median looked toward the target referent, and this pattern was again consistent across blocks. Thus, the high vocabulary children were able to use cross-situational consistency to identify the referents for transitive action verbs, even when this required abstracting across actors and objects. In contrast, children with lower vocabularies failed to learn the verbs in Experiment 2. This suggests that the increase in complexity of the sentences and the scenes made it too difficult for them to keep track of the pairings between verbs and actions.

Children's success in these tasks is striking. Setting aside for a moment the lower-vocabulary children in Experiment 2, 2.5-year-old children were able to use cross-situational consistency to identify the referents of both intransitive and transitive action verbs, despite the fact that they never received a single unambiguous labeling trial. Whereas previous experiments have shown that children can use cross-situational consistency to identify the referents of novel object names under referential uncertainty (Smith & Yu, 2008), the present experiments are the first to demonstrate that children can do so with novel verbs. Thus we can conclude that 2.5-year-old chil-

dren possess the basic mechanisms required to exploit cross-situational consistency in verb learning: The children created lexical entries for new verbs, and attached to those entries some information about candidate event referents, without knowing which event the verb described. They then retrieved and updated this information across trials to identify the intended verb-action pairings. They did this quickly, even though the task required them to detect action similarity across event tokens involving different event participants.

These results complement recent evidence that young children can establish an initial lexical entry for a new verb under great referential ambiguity, and attach useful linguistic information to that lexical entry (Arunachalam & Waxman, 2010; Scott & Fisher, 2009; Yuan & Fisher, 2009, 2010). For instance, Yuan and Fisher (2009) showed 28-month-olds videos of two people conversing, using a made-up verb to describe unseen events. The verbs were presented in transitive sentences (e.g., "Anna blicked the baby!" "Really, she blicked the baby?") or in intransitive sentences ("Anna blicked!" "Really, she blicked?"). One or two days later, children heard the verb in isolation ("Find blicking!") while viewing two test videos: a two-participant causal action and a one-participant action. Children who had previously heard the verb used in transitive sentences looked longer at the two-participant event than did children who had heard the verb used in intransitive sentences. These results suggest that even in the absence of any referential information about a verb's meaning, children can establish a lexical entry and attach to it information about the types of sentences that the verb occurs in. Children can then integrate these facts with later scene information in order to identify the verb's referent. Recent extensions of this dialogue-and-test procedure show that younger children also succeed in similar tasks (Arunachalam, Escovar, Hansen, & Waxman, 2011; Yuan & Fisher, 2010), and that 2-year-olds encode and retain not only verb transitivity, but also the characteristics of the nouns that filled the verb's argument slots in the dialogues (e.g., noun animacy; Scott & Fisher, 2009; Yuan, Fisher, Kandhadai, & Fernald, 2011).

These prior findings, taken together with the present results, suggest that young children create and maintain partial representations of verbs in their lexicons that can then be refined over time. These lexical entries are fragmentary, but can include multiple types of information about a new verb: about its linguistic contexts, or, as the present experiments show, about candidate referent events. In either case, this information can be retrieved when the child encounters the same verb again, and used to guide the search for new information about this verb.

A similar picture of extended and piecemeal lexical learning applies across the lexicon (see Gelman & Brandone, 2010; Swingley, 2010, for discussion). Carey and Bartlett (1978) made this point in their description of their famous "chromium" experiment. Children used a single very informative exposure to a new color term to learn something about this word, such as that it was a color word. This knowledge was long-lived, affecting children's

³ In fact, a separate *t*-test on block 1 performance alone revealed a significant preference for the target over distracter event, both for the children in Experiment 1, $t(35)=2.81$, $p=.008$, $d=.47$, and the high-vocabulary children in Experiment 2, $t(17)=3.11$, $p=.006$, $d=.73$.

responses to the same word a week later, but typically supported only partial success in production and comprehension tests. Carey (1978) termed the initial creation of an incomplete lexical entry *fast mapping*, and contrasted this with the *extended mapping* that would be required to refine knowledge of the full meaning and use of each word over many exposures. The present work adds to an emerging literature investigating the learning mechanisms that permit both the initial fast mapping and the extended refinement of verbs' meaning and use.

Our results also point to some of the limits on children's ability to encode and compare potential event referents for verbs under referential uncertainty. The small increase in complexity of the sentences and scenes used in Experiment 2 overwhelmed the lower vocabulary children, preventing them from learning the verbs in this experiment.

Interestingly, this difficulty occurred despite the fact that our task, like that of Smith and Yu (2008), simplified the learning situation in two key ways. First, we presented children with minimal referential ambiguity: As children heard each novel word, they saw *only* two ongoing events. While each event presumably offered up multiple potential construals (and thus there were in fact more than two potential referents on each trial), this task setting still does not approach the degree of ambiguity found in the playground example with which we began. This is a significant simplification, because recent results show that the efficiency with which adults exploit cross-situational information in noun learning degrades dramatically as referential ambiguity increases (Medina et al., 2011; Smith et al., 2011). Second, the presentation of two referents and two words in each trial could have yielded a helpful source of implicit contrast information. Children were never told that the two verbs they heard on each trial named the two actions they saw, but their behavior suggested that this rapidly became clear to them. The fact that the children in Experiment 1, and the high-vocabulary children in Experiment 2, often began to switch their gaze from one video screen to the other as soon as they heard the beginning of the second sentence ("And she's..."; see Figs. 2 and 4A) suggests that they became sensitive to the structure of the task (see also Waxman et al., 2009, for evidence that 2-year-olds can learn to anticipate events in a structured word-learning task). If so, then this task effectively provided children with a positive exemplar and a contrasting event on each trial: One actor was *pimming* but not *nading*, and the other was *nading* but not *pimming*. Such contrast information facilitates word learning. For example, 3-year-olds failed to extend a newly-learned verb to a new instance of the same action when trained with a single action exemplar (Imai et al., 2005; Piccin & Waxman, in preparation), but succeeded if they were also shown a different event to which they were told the verb did not apply (e.g., *Uh-oh, he's not larping that!*; Piccin & Waxman, in preparation; Waxman, 2009; see Waxman & Klibanoff, 2000, for evidence that contrast also supports the extension of novel adjectives). The implicit contrast in each trial of the current experiments might have given children some of the benefit of

these explicit contrast manipulations, making it easier for them to extend the verbs to new event tokens involving different actors and objects.

Thus, our results suggest that while children can attach event referential information to a verb's lexical entry under referential uncertainty, this procedure may break down even in contexts of minimal ambiguity. In situations of greater referential uncertainty, as might be expected in natural language learning, this mechanism may require support from additional sources of information. This is consistent with findings from the human-simulation paradigm, which suggest that as referential uncertainty increases, even adults have difficulty identifying the referents of words based on scene information alone (Gillette et al., 1999; Medina et al., 2011; Snedeker & Gleitman, 2004; see also Smith et al., 2011). We attempted to support the use of cross-situational consistency by providing children with accompanying sentence information, something that has been shown to help adults make use of scene information in verb learning (Gillette et al., 1999; Snedeker & Gleitman, 2004). The fact that low-vocabulary children nevertheless failed in Experiment 2 suggests that they require further support in order to make use of cross-situational consistency in verb learning.

Ongoing work in our lab suggests that another useful source of information is prior exposure to a new verb's phonological form and syntactic properties (see also Swingley, 2007). Results from a pilot experiment suggest that when slightly younger children, 25-month-olds, were presented with the intransitive action verb materials of Experiment 1, they failed to identify the verbs' referents within this brief task. If, however, the same ambiguous trials were preceded by dialogue videos of two women using the novel verbs in intransitive sentences, 25-month-olds succeeded in identifying the verbs' referents. These preliminary results suggest that even 25-month-olds have the tools necessary to profit from multiple information-sources about verb meaning and use, distributed over multiple exposures to the word.

4.1. Learning mechanisms: statistics versus conjectures

The present results show that children can use cross-situational consistency to learn novel verbs under referential uncertainty. These and other results raise many interesting questions about the mechanisms by which children (or adults) manage to recruit cross-situational consistency in word learning, and especially about how the answers to these questions might differ across different word categories. First, one might ask whether children can encode event referential information at all under greater referential ambiguity – ambiguity that approaches the playground scenario with which we began. Second, one might ask how fully children retain the statistics of their experience, and use them to refine verb meanings over time. The literature offers two classes of predictions regarding these questions.

One possibility, *true cross-situational learning*, is cross-situational comparison as it has traditionally been construed – a process of intersection discovery (Pinker, 1984; Siskind, 1996; Smith & Yu, 2008). According to this

view, when learners encounter a new word, they note whatever candidate referents co-occurred with that word. The next time they encounter the same word, they compare the current set of candidate referents with the previous set, adding new possibilities to the memory set, and updating the probability that any of the previously encoded possibilities are correct for this word. Co-occurrence information could come in the form of simple associations between scene-components and words (e.g., Smith & Yu, 2008), or it could consist of a set of candidate interpretations that children generate using other cues such as syntactic structure or inferences about the speaker's referential intention (e.g., Frank, Goodman, & Tenenbaum, 2009; Siskind, 1996). This class of models of cross-situational comparison makes a key prediction about learning: Learners' inferences about word meaning should closely reflect the statistical structure of their experience. For example, if children get clear evidence that their top-ranked interpretation of a word was incorrect, they should recover from this error smoothly, falling back on a wealth of statistical information about other candidate meanings supported by past experience (e.g., Fazly et al., 2010).

An alternative possibility is that children cannot store and compare anything like all possible sets of potential referents, and thus are unable to engage in true cross-situational learning as described above. Instead, children may engage in a form of *approximate* or *conjecture-based cross-situational learning* (Medina et al., 2011; Smith et al., 2011). When children encounter a new word, they may generate a conjecture or guess about that word's meaning. It is this conjecture that is stored and refined (or falsified) over subsequent observations. Children may also store some amount of information about possible alternative referents, depending on their available resources. If they later discover that their conjecture was wrong (e.g., they encounter the word without the hypothesized referent), then they must form a new conjecture based on the information in the present scene and any information they have retained from previous observations. This mechanism also generates a strong prediction. Because only limited information is assumed to be retained across observations (in some cases, only the learner's conjecture is retained), learners may entertain interpretations of a word that are inconsistent with the previous referential contexts in which that word has occurred.

At present, our results are consistent with both of these mechanisms (as are the results of Smith & Yu, 2008). In our experiments, two candidate event-referents were present on each trial and both events were always labeled; each new verb occurred equally consistently with its target action, and equally rarely with any particular distracter action. Under these circumstances, both mechanisms would predict rapid learning. In order to tease these two mechanisms apart, future research will need to make the statistical evidence for one interpretation of each word less overwhelming, permitting direct investigation of the learner's recovery from error. Some recent experiments examining noun learning by adults have begun to address these questions, asking whether referential options that might have been encoded during initial *unsuccessful* attempts to interpret a new word can later be retrieved and used to

guide recovery from the error (compare Medina et al., 2011; Yurovsky, Fricker, Yu, & Smith, 2010).

One strong conclusion from the present experiments, however, is that as we investigate learning mechanisms for cross-situational word learning, we should not assume that these mechanisms apply uniformly to words of different categories. Modest increases in the linguistic and referential complexity of the learning context, from static objects to dynamic actions, and from isolated words to simple transitive sentences, made many 2.5-year-olds unable to succeed in a word-learning task that even 12- to 14-month-olds can conquer in the simplest case (Smith & Yu, 2008; Yu & Smith, 2011). Similarly, the preliminary results with 25-month-olds mentioned above suggest that slightly younger children failed under these circumstances with intransitive action verbs. These limitations in the use of cross-situational statistics in verb learning are particularly striking given that 2- and 2.5-year-olds are not rank novices at verb learning. Children under 2 years old comprehend verbs (e.g., Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Huttenlocher, Smiley, & Charney, 1983; Naigles & Hoff, 2006), and extend them flexibly in production, using them to describe appropriate actions by different actors on different objects (Naigles, Hoff, & Vear, 2009).

4.2. Conclusions

Several recent studies demonstrate that young children can use cross-situational information to learn novel nouns under referential uncertainty (e.g., Smith & Yu, 2008; Yu & Smith, 2011). The present experiments are the first to ask whether children can use cross-situational observation to learn verbs under similar circumstances. Our results show that children can encode event referents for a new verb, attach them to a newly established lexical entry for that verb, and retrieve this information on subsequent trials to refine (or to confirm or reject) the proposed interpretation of the verb. These results demonstrate that the mechanisms required for cross-situational observation still operate when we move beyond the simplest case of associating static objects with isolated words. However, we also saw limits on children's ability to recruit cross-situational observation to learn verbs, suggesting that access to this information source differs across word categories. Verb learning poses challenges that the learning of object labels does not; these differences have consequences for the mixture of information sources required to support efficient word learning.

Acknowledgements

This research was supported by a predoctoral traineeship from NIMH (1 T32 MH1819990) to Rose Scott, by grants from the NIH (HD054448) and NSF (BCS 06-20257) to Cynthia Fisher, and by the Research Board of the University of Illinois. We would like to thank Renée Baillargeon, Andrei Cimpian, and Katherine Messenger for helpful comments on our manuscript; the staff of the University of Illinois Language Acquisition Lab for their help in data collection; and the parents and infants who participated in this research.

Appendix A. Video and audio used in Experiments 1 and 2

A.1. Experiment 1, verb order 1

Verb order 2 was identical except that within each trial, verbs were presented in the opposite order (e.g., Trial 1: “She’s nading. And she’s pimming.”).

	Left screen	Right screen	Audio
Trial 1	Girl A tilting side to side	Girl B squatting	“Look! She’s pimming. And she’s nading.”
Trial 2	Girl C lifting knee	Girl D tilting side to side	“Look! She’s tazzing. And she’s pimming.”
Trial 3	Girl A stretching arms	Girl C squatting	“Oh wow! She’s nading. And she’s rivving.”
Trial 4	Girl D squatting	Girl B lifting knee	“Wow! She’s nading. And she’s tazzing.”
Trial 5	Girl B tilting side to side	Girl C stretching arms	“Wow! She’s rivving. And she’s pimming.”
Trial 6	Girl D stretching arms	Girl A lifting knee	“Oh wow! She’s tazzing. And she’s rivving.”
Trial 7	Girl B lifting knee	Girl C squatting	“Oh look! She’s tazzing. And she’s nading.”
Trial 8	Girl D squatting	Girl A tilting side to side	“Oh wow! She’s nading. And she’s pimming.”
Trial 9	Girl C stretching arms	Girl B tilting side to side	“Look! She’s pimming. And she’s rivving.”
Trial 10	Girl B squatting	Girl A stretching arms	“Look! She’s rivving. And she’s nading.”
Trial 11	Girl D tilting side to side	Girl C lifting knee	“Oh wow! She’s pimming. And she’s tazzing.”
Trial 12	Girl A lifting knee	Girl D stretching arms	“Oh wow! She’s rivving. And she’s tazzing.”

A.2. Experiment 2, list 1, verb order 1

Verb order 2 was identical except that within each trial, verbs were presented in the opposite order (e.g., Trial 1: “She’s nading her toy. And she’s tazzing her toy.”). List 2 used the same videos, but different verbs were paired with each action (see Table 3).

	Left screen	Right screen	Audio
Fam 1	Girl A holding tractor	Girl B holding rake	“Look! She has a toy. And she has a toy.”
Fam 2	Girl C holding ball	Girl D holding giraffe	“Now she has a toy. And she has a toy.”
Trial 1	Girl D tossing giraffe	Girl A lifting tractor	“Now watch! She’s tazzing her toy. And she’s nading her toy.”
Trial 2	Girl B lifting rake	Girl C sliding ball	“Look! She’s nading her toy. And she’s pimming her toy.”
Trial 3	Girl C tossing ball	Girl D spinning giraffe	“Wow! She’s rivving her toy. And she’s tazzing her toy.”
Trial 4	Girl A sliding tractor	Girl B tossing rake	“Look! She’s pimming her toy. And she’s tazzing her toy.”
Trial 5	Girl C lifting ball	Girl A spinning tractor	“Wow! She’s rivving her toy. And she’s nading her toy.”
Trial 6	Girl B spinning rake	Girl D sliding giraffe	“Look! She’s pimming her toy. And she’s rivving her toy.”
Trial 7	Girl A lifting tractor	Girl C tossing ball	“Wow! She’s nading her toy. And she’s tazzing her toy.”
Trial 8	Girl D sliding giraffe	Girl B spinning rake	“Look! She’s rivving her toy. And she’s pimming her toy.”
Trial 9	Girl A spinning tractor	Girl D tossing giraffe	“Look! She’s tazzing her toy. And she’s rivving her toy.”
Trial 10	Girl C sliding ball	Girl B lifting rake	“Wow! She’s pimming her toy. And she’s nading her toy.”
Trial 11	Girl B tossing rake	Girl A sliding tractor	“Look! She’s tazzing her toy. And she’s pimming her toy.”
Trial 12	Girl D spinning giraffe	Girl C lifting ball	“Look! She’s nading her toy. And she’s rivving her toy.”

References

- Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, 19, 347–358.
- Arnold, J. E. (2008). The BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*, 108, 69–99.
- Arunachalam, S., Escovar, E., Hansen, M., & Waxman, S. R. (2011). Verb learning from syntax alone at 21 months. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual boston university conference on language development* (pp. 21–24). Somerville, MA: Cascadia Press.
- Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, 114, 442–446.
- Arunachalam, S., & Waxman, S. R. (2011). Grammatical form and semantic context in word learning. *Language Learning and Development*, 7, 169–184.
- Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Hillsdale, NJ: Erlbaum.
- Behrend, D. (1995). Processes involved in the initial mapping of verb meanings. In M. Tomasello & W. Merriman (Eds.), *Beyond names for things: Children’s acquisition of verbs* (pp. 251–273). Hillsdale, NJ: Erlbaum.
- Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: The M.I.T. Press.

- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33–B44.
- Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, 15, 17–29.
- Carlson, G. N., & Tanenhaus, M. K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Syntax and semantics: Thematic relations* (pp. 263–288). San Diego, CA: Academic Press.
- Childers, J. B. (2008). The structural alignment and comparison of events in verb acquisition. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 681–686). Austin, TX: Cognitive Science Society.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619.
- Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34, 1017–1063.
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the Macarthur communicative development inventories. *Applied Psycholinguistics*, 21, 95–115.
- Fernald, A., Perfors, A. F., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the second year. *Developmental Psychology*, 42, 98–116.
- Fernald, A., Swingle, D., & Pinto, J. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, 72, 1003–1015.
- Fillmore, C. J. (1970). The grammar of hitting and breaking. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 120–133). Waltham, MA: Ginn.
- Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, 5, 56–65.
- Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92, 333–375.
- Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20, 579–585.
- Gelman, S. A., & Brandone, A. C. (2010). Fast-mapping placeholders: Using words to talk about kinds. *Language Learning and Development*, 6, 223–240.
- Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, 49, 988–998.
- Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 59, 47–59.
- Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17, 684–691.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73, 135–176.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 3–55.
- 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, 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, 3101–3119.
- Grimshaw, J. (1993, April). Semantic structure and semantic content: A preliminary note. In *Paper presented at the conference on early cognition and the transition to language*. University of Texas at Austin.
- Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review*, 90, 72–93.
- Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? *Child Development*, 76, 340–355.
- Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., et al. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, 79, 979–1000.
- Levin, B. (1995). Approaches to lexical semantic representation. In D. E. Walker, A. Zampolli, & N. Calzolari (Eds.), *Automating the lexicon: Research and practice in a multilingual environment* (pp. 53–91). Oxford: Oxford University Press.
- Levin, B., & Rappaport-Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press.
- Ma, W., Golinkoff, R. M., Hirsh-Pasek, H., McDonough, C., & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language*, 36, 405–423.
- Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's verb learning and extension. *Developmental Science*, 11, 628–634.
- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, 11, F9–F16.
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20, 121–157.
- McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2008). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science*, 14, 181–189.
- Medina, T. N., Trueswell, J., Snedeker, J., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *PNAS*, 108, 9014–9019.
- Merriman, W. E., Marazita, J., & Jarvis, L. (1993). Four-year-olds' disambiguation of action and object word reference. *Journal of Experimental Child Psychology*, 56, 412–430.
- Merriman, W. E., Marazita, J., & Jarvis, L. (1995). Children's disposition to map new words onto new referents. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 147–184). Hillsdale, NJ: Erlbaum.
- Naigles, L. (1990). Children use syntax to learn verb meaning. *Journal of Child Language*, 17, 357–374.
- Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58, 221–251.
- Naigles, L. R., Hoff, E., & Vear, D. (2009). Flexibility in early verb use: Evidence from a multiple-N diary study. *Monographs of the Society for Research in Child Development*, 74.
- Naigles, L. R., & Hoff, E. (2006). Verbs at the very beginning. In Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 336–363). New York: Oxford University Press.
- Naigles, L., & Hoff-Ginsberg, E. (1995). Input to verb learning: Evidence for the plausibility of syntactic bootstrapping. *Developmental Psychology*, 31, 827–837.
- Naigles, L. R., & Kako, E. T. (1993). First contact in verb acquisition: Defining a role for syntax. *Child Development*, 64, 1665–1687.
- Piccin, T., & Waxman, S. R. (in preparation). Children use contrast and cross-situational exposure to learn verbs.
- Piccin, T., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human-simulation paradigm. *Language Learning and Development*, 3, 295–323.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, 92, 377–410.
- Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, 24, 777–803.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 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, 480–498.
- Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 1558–1568.
- Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In Hall & Waxman (Eds.), *Weaving a lexicon* (pp. 257–294). Cambridge, MA: MIT Press.
- Snedeker, J., Li, P., & Yuan, S. (2003). Cross-cultural differences in the input to early word learning. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 1094–1100). Mahwah, NJ: Erlbaum.
- Swingle, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*, 43, 454–464.
- Swingle, D. (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development*, 6, 179–183.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.

- Waxman, S. (2009, April). Comparison and contrast: How these ubiquitous processes influence the specific task of word learning. In *Paper presented at the Biennial meeting of the Society for Research in Child Development*, Denver, CO.
- Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology*, 36, 571–581.
- Waxman, S. R., & Lidz, J. (2006). Early word learning. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Cognition, perception and language* (6th ed., pp. 299–335). New York: Wiley.
- Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). 24-Month-old infants' interpretation of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, 59, 67–95.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114, 245–272.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18, 414–420.
- Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, 14, 165–180.
- Yuan, S. & Fisher, C. (2010, March). Learning combinatorial properties of verbs from listening. In *Paper presented at the 17th international conference on infant studies*, Baltimore, MD.
- Yuan, S., Fisher, C., & Snedeker, J. (in press). Counting the nouns: Simple structural cues to verb meaning. *Child Development*.
- Yuan, S., & Fisher, C. (2009). "Really? She blinked the baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, 20, 619–626.
- Yuan, S., Fisher, C., Kandhadai, P., & Fernald, A. (2011). You can stipe the pig and nerk the fork: Learning to use verbs to predict nouns. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual Boston University conference on language development* (pp. 665–677). Somerville, MA: Cascadilla Press.
- Yurovsky, D., Fricker, D., Yu, C., & Smith, L. B. (2010). The active role of partial knowledge in cross-situational word learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 2609–2614). Austin, TX: Cognitive Science Society.
- Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development*, 6, 179–208.