

# Cross-situational Learning of Foreign Vocabulary Reveals a Possible Noun Bias in Adulthood

Jason D. Ludington  
University of California, Los Angeles, California, USA

**Abstract**—While the debate continues into its third decade over why toddlers across languages tend to acquire nouns quicker than verbs, surprisingly few studies have re-directed this same question at adult. Does such a word class learning bias exist among adult second language learners? In the present study, 48 participants were taught foreign (Hebrew) vocabulary as short phrases in either cross-situational or single-situational learning contexts to test for a noun bias and whether it would be reduced through exposure to multiple contexts. Separate, non-interacting effects were found for word class (nouns better) and the number of contexts used (two better than one). How these results support the referential clarity hypothesis and other theories is discussed.

**Index Terms**—noun bias, word learning, vocabulary acquisition, cross-situational learning, line drawing, fast mapping

## I. INTRODUCTION

Nitsch (1977) experimented on the effect of the number of learning contexts by teaching students novel words and their meanings either by a) definition alone, b) definition with examples within one story context (a cowboy story), or c) definition with examples from several contexts (including the cowboy story). There was a positive effect of varying contexts on retrieval of word meanings in a unique story context—students who learned across several contexts performed best. This was one of the earliest demonstrated advantages of what was to eventually be called cross-situational learning. Since then, cross-situational learning studies have become quite widespread, but I have yet to encounter a study which addresses two important questions posed in this paper: first, what is the demonstrated advantage of cross-situational learning relative to its opposite—learning from within a single situation? Second, does the purported advantage of learning across situations apply equally in the cases of two very important classes of vocabulary targets for foreign language learners—nouns and verbs?

### A. Potential for Noun Bias

Most cross-situational learning studies train participants with targets of a single word class (nouns and adjectives in the studies mentioned above). It therefore remains unknown whether cross-situational learning is equally effective when used to learn verbs as nouns. These are two words classes that language beginners of all ages typically begin with. Thus finding a difference in usefulness of cross-situational learning as a learning method could have important implications for early foreign word learners.

The noun bias is the tendency for children to acquire noun labels more quickly than other class labels. This noun bias has been well-documented in naturalistic studies of children learning their first language (e.g., Bornstein et al., 2004; Gentner, 1982) and even two languages simultaneously (Levey & Cruz, 2003), yet very few language studies have turned an eye toward the possibility of a noun bias among adult language learners. Findings from two studies suggest there could be a noun bias among adults just as among young children. Piccin & Waxman (2007) found that adults and children were more successful at guessing nouns than verbs from beeps which replaced actual words in video dialogues. One shortcoming of the Piccin and Waxman study was that despite measuring and documenting a word class difference in word imageability, this factor was neither controlled nor assessed for its role in accounting for word guessing accuracy, which could have accounted for the noun advantage found. Gillette, Gleitman, Gleitman, and Lederer (1999) tested and found that imageability did in fact mediate the noun bias effect they found. However as these studies were simulations focused on hypothesis formation in fairly ambiguous settings, their results may not completely address whether or not there a noun bias among adults beginning to learn a foreign language. In the present study, the effect of imageability was considered for its potential effect in mediating the effect of word class on word learning.

Naigles (1990) suggested verbs are harder to learn because they are more morphosyntactically complex, and suggested knowledge of syntax could aid vocabulary acquisition. Gillette et al. (1999), in researching possible explanations of the noun bias, manipulated syntax as a cue to word meaning. They found that knowledge of a target word's class indeed helped participants guess word meanings. In the current study, training trials presented at the beginning of the experiment clued participants in to the grammar, though the target language syntax was not made explicit. This allowed us to test whether, when participants have at least some knowledge of a language's syntax, noun learning is equivalent to verb learning.

### B. *Cross-situational Learning*

Akhtar and Montague (1999) showed that 2-, 3-, and 4-year olds could map novel adjective terms to either textures or shapes of objects by presenting exemplars which shared a target feature but varied on non-target features. Smith and Yu (2008) found that even participants as young as 12 and 14 months were able to solve an ambiguity problem using cross-situational learning statistics. In their study, infants heard two labels refer to two objects on each of a series of slides. The labels on any particular slide presented an ambiguity problem—these young participants could not know which labels referred to which objects. However across labeled situations infants grasped which labels referred to which referents because labels and referents co-occurred in perfect correlation but labels co-occurred with incorrect referents more randomly.

From a cognitive perspective, the benefits of learning from multiple rather than single learning contexts or situations can be explained in terms of cue-overload theory (Watkins & Watkins, 1976) and stimulus sampling theory (Estes, 1955). Cue-overload theory says that remembering is more likely when multiple learning cues are provided because the additional cues tip the ratio of cues to targets, favoring retrievability. Stimulus sampling theory says memory is a function of the number of stimuli present at the time of recall; by sampling from a variety of environments during learning, the likelihood that any of those stimuli are present during recall is higher and thus recall is more likely. Thus by learning in multiple contexts, learners link multiple cues from each context to their target, thereby increasing retrieval likelihood when targets need to be recalled. In the present study, cross-situational learning is expected to be a superior learning strategy relative to learning target label meanings from single, repeated contexts. Besides the explanations of advantage for multiple learning situations mentioned above, the cross-situational learning effect may function by reducing ambiguity for how words in an utterance maps onto their referents, as explained in more detail below.

In the present study participants learned words for target elements by seeing and hearing them labeled twice per pair of contexts presented sequentially. These “contexts” were images or still scenes of actors performing actions. In each juxtaposed pair of contexts, a single, common element (either an object or action element) was repeated, affording participants the opportunity to solve the ambiguity problem present in each individual image presentation with its phrase description—a problem that immersion learners are very familiar with. By assuming that common visual elements repeated across image pairs instantiate common words repeated between phrase pairs, learners could apply a mapping strategy not otherwise available when learning from single, repeated context. This would demonstrate a purported advantage of the cross-situational over the single-situational word learning strategy.

### C. *Predictions*

I predicted single-situational learning would lead to much poorer learning than cross-situational learning for four reasons. First, in single-situational conditions, participants would have had little way of learning the syntax of the target language (noun-verb word order) besides the subtle clues given through feedback on practice trials. In cross-situational conditions, this syntax would have become somewhat apparent after a few trials because the shared element across contexts was always uttered either first or second in phrases; nouns were always uttered first, and verbs were always uttered second. Second, I predicted a disadvantage for single-situational learning because word parsing would have been much less certain (natural intonation provided a clue). For example in the phrase “tinok shotay” participants might have accidentally parsed “tinoksho tay.” Such parsing errors would have led to lower word recognition at test. In cross-situational conditions, however, the phrase “atalef shotay” was followed with the phrase “tinok shotay,” allowing learners to parse “shotay” with greater certainty in each phrase. Third, past learning research has shown that varying the learning context is beneficial to learning (e.g., Nitsch, 1977). And fourth, the presence of a second, labeled instance of each target in new contexts should increase certainty in label-to-referent mapping which should facilitate learning.

Initially I predicted that there should be no noun bias because nouns and verbs were closely matched on factors of word imageability, familiarity, frequency, as well as three measures of image clarity—factors which I found matter to word learnability based on prior work (Ludington, in press). Therefore no word class effect was initially expected. However in the process of collecting data, images were measured for their name agreement (proportion of participants who named each element in learned images consistently with predetermined responses). Based on a detected noun-verb difference in target naming, the revised prediction was that nouns should be learned better than verbs due to their more consistent identification across images.

I predicted an interaction between number of situations and word class. The referential clarity hypothesis (Ludington, in press) suggests noun biases among adult learning are due to more referential ambiguity of verbs than nouns. Thus in the present study noun learning may slightly exceed verb learning in single-situational conditions because nouns tend to be more identifiable in context images (Ludington, in press), and indeed were found more identifiable in the current study. I expected the use of two learning situations per target would eliminate much or all of the ambiguity typical of word learning situations with multiple words and referents, thereby eliminating any word class bias.

By modeling word class with factors that might affect word learnability, the latter factors could be statistically held constant while testing the effect of word class. Some model testing and development was performed in order to better understand how these variables related to word learning.

## II. METHODOLOGY

### A. Participants

Fifty undergraduate participants were recruited from a subject pool of students taking psychology (mostly) or linguistics courses at the University of California, Los Angeles. To be eligible for this study, participants were not to know any more than three words of Hebrew. The first two participants were run only to pilot the procedures; some procedures were altered in response to this dry run, and the data from those two participants were excluded from analysis.

The experiment proper sampled 48 participants: 37 females and 11 males. The average age of participants was 21.5 years (range: 18-40,  $SD=5.0$ ). To assess whether being a native English speaker had any effect on performance, participants were asked to report their language background on paper; 38 participants reported English as their primary language, and 10 reported English was not their primary language. Those who reported English as non-primary also self-reported their language ability in English on a scale of 1 (unable to use the language) – 10 (native fluency). Scores ranged from 6 – 10, and the mean among these 10 participants was 8.00 ( $SD = 1.15$ ).

All participants also reported their knowledge of other languages, and self-rated their proficiency in each reported language, because it was wondered if those with more accumulated language learning experience might have a special advantage in foreign vocabulary acquisition. A count was taken of the number of languages besides English reportedly known; numbers ranged from zero (1 participant) to five (1 participant). To test the hypothesis that more acquired language proficiency might put learners at an advantage in the present task, all self-reported language proficiency scores for all reported languages (except their primary languages which I assumed to be 10 = native) were summed for each participant to form a total foreign language proficiency measure. Participants' average total foreign language proficiency was 8.49 ( $SD = 4.93$ ). A number of participants reported various proficiencies in Spanish (26), French (9), Mandarin Chinese (6), Farsi (4), Hindi (4), Korean (4), Vietnamese (3), Polish (2), Italian (2), Polish (2), German (2), Urdu (2), and a number of other languages (reported once each: Canbese, American Sign Language, Telegu, Amheric, Japanese, Persian, Punjabi, Tunisian Arabic, Arabic, Portuguese, and Nepali). Proficiency in any particular language was not analyzed as no special link was expected to exist between any of these and the target language, Hebrew.

### B. Design

The study was designed as a two-way, mixed factorial experiment. The first independent variable was word class manipulated within subjects at two levels—nouns and verbs. The second independent variable was the number of learning contexts (unique images) presented, manipulated at two levels between subjects: cross-situational and single-situational learning. The dependent variable, recognition, was measured on a binary scale (0=wrong, 1=correct). Outcomes, 24 per participant, were nested within the participant unit of analysis. Sixteen additional predictor variables—characteristics of participants, words, and images—were also tested.

### C. Materials

A consent form, a biographical data form, a laptop computer, and presentation software were used. Following are more detailed descriptions of each.

#### **Consent form.**

All participants signed consent to participate as required by the institution's review board.

#### **Biographical data form.**

A data form collected language background information. One question addressed what the participants' first language was. If not English, another question asked participants to rate their language ability in English on a fluency-scale from 1-10, where 1 meant unable to use any of the language, and 10 meant fluent. A third question asked for other languages the participant knew, and how fluent he or she was in each (using this same fluency scale). Age and sex data were also collected.

#### **Auditory stimuli.**

##### **Language and speaker.**

Hebrew was used as the target language because relatively few people are fluent in this language worldwide, and because I found a Hebrew speaker for hire by which all auditory stimuli were obtained. The speaker was a 38 years old native Hebrew-speaking male who was raised in Israel and moved to the U.S. as an adult. He spoke all words individually and in phrases at a normal speech rate as requested. He spoke all nouns in singular form; verbs were spoken in either masculine or feminine form according to the gender of the illustrated actor performing the verbs in associated image. Phrases were spoken with natural articulation and sentential intonation.

Some Hebrew words were not suitable to use in this study because they were cognates of English words (e.g., *penguin* in Hebrew means penguin). For these I substituted other Hebrew words. For example, I let the Hebrew word for angel, *malachit*, stand in for *penguin*. Four noun substitutions and three verb substitutions were made for target words, each substitution being from another Hebrew word of the same word class. This was done to maintain the target language's status as a completely foreign language to the participants without changing the fact that it was an actual language.

##### **Syntax.**

Hebrew words are normally ordered as subject–verb–object, as in English (Jacobs, 2003). All phrases in this study were produced in a noun-verb order. All verbs were uttered in one gender on each occurrence within the experiment—either in the masculine or feminine form depending on the actor.

#### **Physical attributes.**

Words were recorded using Audacity 1.3 (Beta) (sound recording software). Sound clips were edited to include a (approximately) 100-millisecond lag before speech onset and a 200-millisecond lag after speech offset to ensure that the complete word was uttered and that no soft or subtle word-parts were accidentally cropped during editing. Utterance lengths were measured to the nearest hundredth of a second. Table 1 includes utterance length, as well as means of other stimulus characteristics (described below) of nouns and verbs in this study.

TABLE I.  
NOUN AND VERB STIMULUS CHARACTERISTICS AND SAMPLE SIZES OF MEASUREMENTS

		Nouns		Verbs		Sample size
		Mean	SD	Mean	SD	N
English words	Familiarity	6.56	0.35	6.52	0.45	26
	Imageability	6.78	0.25	6.74	0.25	26
	Frequency	3.56	0.89	3.83	0.99	20
Hebrew words	Utterance length	0.64	0.14	0.75	0.22	48 words
	Phonemic length	5.00	1.38	6.38	2.63	48 words
	Syllabic length	2.33	0.76	2.75	0.99	48 words
Images	Name-isolate	0.91	0.09	0.89	0.13	19 <sup>1</sup>
	Name Agree S images	0.91	0.13	0.76	0.25	10 <sup>1</sup>
	Name Agree F images	0.93	0.08	0.86	0.18	12 <sup>2</sup>
	Goodness of depiction <sup>3</sup>	4.89	0.15	4.82	0.17	20
	Alternative interpretation <sup>3</sup>	4.29	2.07	4.92	2.85	20

Numbers of phonemes and syllables were counted for all auditory stimuli. Although number of syllables did not significantly differ between nouns and verbs,  $p=.11$ , independent samples  $t$  tests showed verbs were generally longer than nouns in phonemic length,  $t(46)=-2.26$ ,  $SE=.61$ ,  $p=.03$ , and had longer utterance lengths,  $t(46)=-2.29$ ,  $SE=.05$ ,  $p=.03$ , (all tests two-tailed). At first glance, a word class difference in word lengths would appear to be of concern as a confounding of word class with length; however this correlation turns out not to be of concern in light of two facts. First, Ludington (in press) worked with nonsense words and found, based on over 7000 observations from 84 participants (96 observations per participant), that target word length as measured in the current study (utterance length, number of phonemes, and syllables) could not predict recognition. Thus no relationship between word length and word learnability was found even with great statistical power. Second, word length was not associated with recognition outcomes in the present study either. Across that and the current studies (using nonsense and Hebrew words, respectively) and with a combined sample of 144, no effect of word length on word learning was found.

#### **Training Images.**

There were two groups of images presented to participants during the training phase of this experiment—those always presented first in trials (labeled “F” images for “first”), and those always presented second in trials (labeled “S” images for “second”). Images were mostly obtained from the internet and some were hand-drawn by research assistants. All images appeared as black and white line drawings, and were easily identifiable as line drawings (as opposed to black-and-white photographs); they were drawn with varying artistic abilities ranging from simple lines and cartoony appearance to detailed renditions with proportions and shading of a realistic nature. S images, also used in a previous study (Ludington, in press), did not qualitatively differ from F images; nor did images obtained from the internet qualitatively differ from those created by research assistants. Name agreement values for S images were measured from the responses of 29 participants sampled in a previous study (unpublished dissertation).

#### **Physical dimensions.**

The heights and widths of F and S images were measured with a computer mouse using a pixel ruler (freeware); these measurements were used to ensure there were no systematic size differences among image stimuli. Heights and widths were summed to yield a composite measure of each image’s size. The composite measurements were tested for size differenced that could potentially bias certain learning conditions. S images required no comparisons because they were seen by all participants regardless of target word class or number of learning situations. F images sharing a noun with S images ( $M=933$  pixels,  $SD=147$ ) were not different in size than F images sharing a verb with S ( $M=991$  pixels,  $SD=161$ ), independent  $t(46)=1.34$ ,  $p=.19$ , demonstrating no word class learning bias owing to image size differences.

#### **Name agreement.**

Name agreement measurements of images were defined as the degree to which participant responses were consistent with target responses; averages are provided in Table 1. Measuring name agreement was theory-motivated: the referential clarity hypothesis specifies that vocabulary learnability is largely a function of how clearly meanings are

<sup>1</sup> These were measurements of isolate images, which were viewed only at test.

<sup>2</sup> Values based on 6 coder average of participants’ data.

<sup>3</sup> Values based on a single coder of each participant’s data.

given in learning situations. As referential clarity is particularly difficult to equate between nouns and verbs, a further practical purpose of measuring name agreement was to statistically account for the influence of name agreement, a proxy for referential clarity, on learning.

Judgments of agreement between participants' responses and target responses were made in the form of response coding using three codes: 0=incorrect, 0.5=difficult to classify, and 1=consistent. Across developmental phases of this research project, and described in greater detail elsewhere (Ludington, in press; unpublished dissertation), nine different coders participated in making judgments of response consistency, but any one participant's response was coded by either one coder or six coders. Three different response sets (S images, F images, and isolate images) from three phases of the current research were coded by a three different but overlapping sets of coders. Krippendorff's alpha, a measure of inter-rater agreement, was used to measure inter-rater reliability (Krippendorff, 2004) among each set of coders among respective response sets. A code-book was developed and shared with all coders which contained examples and rationales and a few general suggestions for coding responses: to be accepting of morphological variety; to take responses as correct when a noun or verb appears as word-class ambiguous (e.g., vacuum could be a verb or noun); to accept close synonyms as correct, and more distant synonyms as partially correct (.5), but words at other hierarchical category levels (super- or sub-categories) as either wrong or partially correct depending on how distant the relation seems (e.g., "bird" for duck might be .5; "animal" for duck should be 0). To avoid non-independence of judgments, separate Krippendorff's alphas were calculated between coders for each participant's response set (1 alpha per participant), and alphas were averaged over participants to provide a representative sample of inter-rater reliability.

S image name agreement values were measured in prior work (Ludington, in press) using responses from 10 participants (in a sample similar to the present sample in many respects). Their responses were coded by six coders, and coding reliability was measured. Alpha reliability was calculated for each participant; the mean alpha for these 10 participants was .66.

Participants of the present study were asked to name F images they had not viewed during their experiment proper. To measure name agreement values of F images, I randomly selected half of the participants from among the present participant sample, 12 who named one half (24) of the F images (those sharing a noun with S images), and another 12 who named the other half (24) of those images (those sharing a verb with S images). Each of these 24 participants' responses was coded by one of four coders. Additionally, all four coders coded a common set of responses from six participants (25% of the selected participant sub-sample) in order to establish inter-rater reliability. Krippendorff's alpha was calculated based on these four sets of codes for each of 6 participants' responses separately, and the average alpha was .84.

All images in the testing portion of the present study were "isolate" images (images used to convey a single meaning)—either of nouns or verbs—in isolation. Characteristics of these images were measured as part of a previous study (unpublished dissertation). Name agreement values were measured from the responses of 19 participants; Krippendorff's alpha among 6 coders of each participant's response set was calculated separately by participant so that 19 alphas were generated. The average alpha value was .71.

#### ***Goodness of depiction and number of alternative interpretations.***

Goodness of depiction ratings (scale of 1 (poor) – 5 (very well)) were based on participants' ratings of how well each isolate image represented what it was intended to. Number of alternative interpretations was a count of the number of alternative interpretations offered by all participants who rated each image. Both were measured by the responses from 19 participants in prior, unpublished work using these isolate images.

#### **English words.**

Target words were concrete nouns and verbs from the English language. Subjective measures of word familiarity, word frequency, and word imageability were collected as part of a prior, unpublished study. No noteworthy differences existed in the participant samples between that and the present studies.

#### ***Word familiarity.***

Participants in that unpublished study rated word familiarity using the question, "how familiar is each item to you" by entering a number from 1-7, where 1=completely not, and 7=completely.

#### ***Concept frequency.***

Concept frequency ratings were taken to assess "How often have you encountered these concepts—either directly or in images—over the course of your life?" on the following scale: "1=never, 2=rarely (once every 2 years or less), 3=frequently (once every 6 months or less), 4=quite frequently (once every month), 5=extremely frequently: once every week at least."

#### ***Word imageability.***

The same participants that rated familiarity also were asked to rate *word imageability* on "How easy is it to generate a mental image of each item" using this same scale. Table 1 lists all stimulus characteristics described thus far separately by word class.

#### **The learning program.**

A Toshiba laptop computer (screen size: 19 inches diagonally) was used to present auditory and visual stimuli using Superlab 4.0. Details on how this program file was created follow.

#### **Events.**

A learning event was composed of an image and auditory stimulus (a two-word utterance in Hebrew) presented at nearly the same time. The auditory files purposely included about 100 milliseconds of silence at the beginning of each sound clip so that onsets of sound did not occur simultaneously with the onsets of images. In this way images temporally preceded and succeeded sounds. Events advanced over time at a rate of one event every three seconds.

#### ***Trials.***

Each learning trial was composed of a pair of events; thus each trial lasted 6 seconds. Trials were presented back-to-back as a continuous progression of images throughout each segment.

#### ***Blocks.***

A block was composed of four trials that repeated once in the same order, so eight trials. Four orders were created and used between participants to counterbalance order and sequence effects. Each participant viewed a total of 6 blocks.

#### **Learning conditions.**

Half of the participants were assigned to cross-situational learning, and the other half to single-situational learning. Fig. 1 uses a flow diagram to illustrate two cross-situational learning trials. Participants assigned to cross-situational learning would have seen either the upper pair of images in order or the lower ones, but never both pairs. Only one target element (a noun or a verb) was observed twice between each pair of images. By fixing one element but varying other elements between each pair of images, cross situational learning of the said target element was made viable. With images symbolized as letter bigrams, this learning series, in Hebrew's noun-verb syntax, was AC (first image), AB (second image) for learning a noun across contexts, or DB (first image), AB (second image) for learning a verb across contexts. Underlined letters indicate the common referent between image pairs. Thus while all participants viewed the same second image of each pair, only half viewed the first image as image AC (sharing a noun with the second image), and the other half viewed image DB (sharing a verb with the second image).

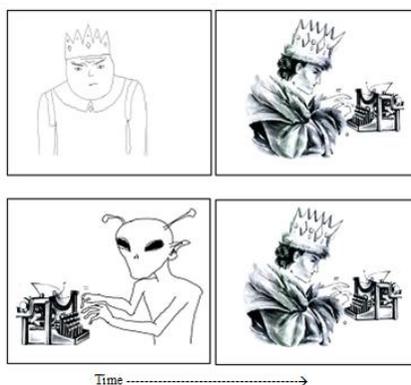


Figure 1. Pairs of context images. Each cross-situational learning trial was composed of a pair of context images linked by a common element (either a noun or a verb). The upper left panel should show an image of a king who is frowning (Hebrew: “melek zoef”) and the upper right panel shows an image of another king who is typing (Hebrew: “melek maklid”); between the pair of images, participants could learn the word for king. The bottom panels show an image of an alien who is typing (Hebrew: “yitsuoer maklid”) and a king who is typing (Hebrew: “melek maklid”; all Hebrew phrases transliterated here by me); their consecutive presentation taught the word for typing. The artistic contributions and adaptations above by Goldie Salimkhan, Kay Lee, and me are reproduced with written permission.

Half (12) of the cross-situational learning trials were constructed to teach nouns, and half (12) were constructed to teach verbs. Manipulating a trial to teach a noun versus a verb was a matter of altering which one of two F images to present before an S image. Half of the participants assigned to the cross-situational condition learned either a noun or a verb from each trial, and always learned 12 nouns and 12 verbs across all 24 trials. The other half of the participants learned the opposite word class from each trial as the first group, also learning 12 nouns and 12 verbs. These participant groups all viewed the same S images, but different F images.

Participants assigned to single-situational learning all saw only S images, the same as those assigned to cross-situational learning, but none saw any F images. Instead each S image was seen twice in a row. A few milliseconds of white screen intervened between these images to create the impression of seeing two images rather than one. A sound clip with the Hebrew phrase to describe this image was played at the onset of the image's two presentations. Fig. 2 demonstrates this condition as a flow diagram.

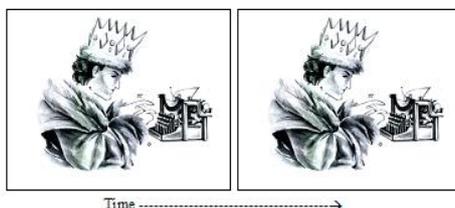


Figure 2. A pair of (identical) context images. Each single-situational learning trial was composed of two of the same image. This image by Goldie Salimkhan is reproduced with written permission.

**Test.**

In the present paradigm word learning might progress gradually in a way not well measured as translated products. A target word's meaning may not lend itself well to translation until it reaches a definition-mature point of development. Maturing hypotheses of word meanings may not be as nuanced as specific word translations into the native tongue, and therefore testing underdeveloped vocabulary with words as targets may not be ideal. Testing with images of learned targets should be preferable because these theoretically depict less nuanced versions of target meanings than words would. Therefore word learning was measured as image recognition. Test images choices were of different exemplars than those viewed during learning to provide a realistic model of word recognition in real world settings.

The learning assessment, a multiple-choice test, was to select each spoken target word's correct meaning from four image choices (chance performance was at 25%). Two of the four image meanings were presented in an S image, as were the other two, as a way to limit performance based merely on associating a target words with its whole S image. There were always two nouns and two verbs presented among the options; this helped to limit performance based merely on having learned the targets' word class.

Participants were tested after a 30-second filler task (to eliminate target contents from working memory) given after each block. The choice to measure learning soon after each block, rather than once at the end of the entire experiment, was motivated by an aim to minimize primacy and recency effects or proactive and retroactive interference effects. Furthermore, recognition performance was anticipated, and indeed found to be, in an ideal range for detecting differences between groups (more items per block or a longer delay between training and test could have caused floor effects, but fewer items or a shorter delay might have caused ceiling effects). Six tests, one given after each of six learning blocks, contained four test items presented in a single order for all participants. Each item tested recognition of one target word from each of the four trials per learning block.

**Targets.**

Among participants assigned to the cross-situational learning condition, only words learned across situations were correct options (participants were not told this). For those assigned to single-situational learning, only one of each trial's word pair was the correct option. For half of these participants either the noun or the verb of each S image was tested, and for the other half of these participants, it was the other of each word pair that was tested, thus controlling the number of items tested across all participants and conditions. The location of the target on the screen was randomly chosen from four possible locations (upper left, upper right, lower left, and lower right) and were counterbalanced across test items to control against participants acquiring a location bias during test performance. The locations of foils were randomly assigned to disguise the relationship between foils and targets.

**Foils.**

Among all test items, the target image of one test trial was presented again as a foil option in one other test trial. In this way every image option was presented twice during testing. No element pair (elements from a single S image) was presented with another element pair more than once during test. This was done to discourage participants from using a process of elimination strategy based on their performance on prior items.

**D. Procedures**

Participants began the experiment by completing a consent form and biographical data sheet. Next the experiment proper was run, followed by an image naming task. Finally, participants were debriefed and thanked, and 1 credit (for one hour of participation) was awarded. The experiment usually lasted about 30 minutes. Following are more detailed descriptions of the experiment proper and the naming task.

**Experiment proper.****Training.**

Instructions presented on screen informed participants that they would be presented with images and short Hebrew phrases describing those images, and that the stimuli would progress at a rate of one image every three seconds. Two learning trials (four images) were presented back-to-back. Then instructions immediately appeared on screen to select the correct target meanings of words just learned. Two test trials were given, a spoken word and four choice options given for each. Participants used the mouse to click on a target for both test training items. Then participants were asked to explain their task. The experimenter corrected any response that was not essentially "to learn the meanings of the words." Then the experiment proper began.

**Learning.**

The learning phase of this experiment began immediately after training was completed. Participants were presented with six learning segments interleaved with filler tasks.

**Filler task.**

After each segment, a filler task was given to participants to work on for 30 seconds to prevent rehearsal of words just learned. Participants were instructed to read a numbered question on a sheet of paper and answer it in writing. If they finished writing their response before 30 seconds had passed, they were instructed to continue writing until 30 seconds had passed, at which point they were stopped from writing. The questions asked were made to be interesting and thought-provoking to gain participants' attention and discourage rehearsal of words just learned. Responses to these questions were not analyzed.

**Testing.**

Following each 30 second filler task, a recognition test was given. Each recognition test, four items long, tested one target from each of the four trials in the just-seen block.

#### Naming Task.

After the experiment was finished, participants were given a context image identification task (to measure name agreement among F image elements). To avoid having participants name the same images seen during the experiment proper (because their prior viewing experience could alter the validity of name agreement as a measure of referential clarity), each participant viewed and named the 24 images they had not seen during the experiment proper. Participants were asked to type two or more words to describe each image (each of which depicted an actor performing an action).

### III. RESULTS

Data were analyzed with logistic regression, a method appropriate for data with a binary outcome variable (successful versus failed recognition). With logistic regression I could see which, if any, factors of targets, images, or participants were predictive of recognition, and could control those factors while analyzing the effects of other factors, if needed. However one disadvantage of using logistic regression in this instance was that model testing with a sample this size (N=48) had to be limited to very small models. In testing measured factors, I developed a small model to describe word learning in the present paradigm.

A number of factors were considered for logistic analysis. I initially tested these 18 factors in individual models. Results and descriptive statistics are displayed in Table 2, below. An alpha criterion of .05 was used for determination of significance. Effect sizes are provided only in more developed models to avoid over- or under-exaggeration at this point.

TABLE II.  
ALL 18 FACTORS TESTED IN INDIVIDUAL MODELS

Factor type	Factors	Mean (SD), or % cases	Wald $\chi^2$	p
Independent	Word class	50.0% nouns	6.57	.01
	Number of situations	50.0% within	22.89	.00
Image	Goodness of depiction	4.87 (.16)	2.47	.12
	Alternative interpretations	4.57 (2.45)	3.62	.06
	Name agreement, F images	0.90 (.14)	0.57	.45
	Name agreement, S images	0.84 (.21)	5.45	.02
	Name agree, Isolate images	0.90 (.11)	0.01	.92
Hebrew Word	Utterance length	0.70 (.19)	0.4	.53
	Syllables	2.54 (.89)	1.53	.22
	Phonemes	5.69 (2.17)	0.05	.82
English Word	Familiarity	6.56 (.39)	1.59	.21
	Imageability	6.76 (.25)	0.51	.48
	Frequency	3.72 (.94)	3.11	.08
Participant	Age	21.53 (5.01)	0.69	.41
	Sex	77.1% female	2.13	.15
	English 1st language	78.7% Eng 1 <sup>st</sup>	1.63	.20
	English proficiency	9.57 (.96)	0.51	.22
	Total proficiency	18.49 (4.93)	3.01	.08

Number of situations was significant, model Wald  $\chi^2(1)=22.89$ ,  $p<.01$ , favoring cross-situational learning over single-situational learning. Word class was also significant, model Wald  $\chi^2(1)=6.57$ ,  $p=.01$ , with a learning advantage for nouns over verbs. To estimate their effect sizes, these models were further developed. Although there were inter-correlations between all three name agreement indices—all  $r > .40$ —only S image name agreement was significantly predictive of outcomes, model Wald  $\chi^2(1)=5.45$ ,  $p=.02$ . Contrary to prediction, the noun bias mentioned above was no less present in cross-situational learning than in single-situational learning conditions, interaction  $p=.93$ .

Having determined which individual factors appeared to affect learning, I modeled them together to confirm each finding with other significant factors held constant. I entered name agreement of S images, word class, and number of situations into a model because these were all found reliable when individually tested. This model was significant, Wald  $\chi^2(3)=29.55$ ,  $p<.001$ , with word class ( $p=.003$ ) and number of situations ( $p<.001$ ) as significant factors; however, name agreement did not add significantly to this model so it was removed.

Next I tested whether the significance of word class found above might just owe its effect to differences between nouns and verbs in name agreement (S images), number of alternative interpretations (which bordered on significance), or imageability (found predictive of word learning in Ludington, in press). To address this, I tested word class with each of these factors, one by one, in two-factor models (to keep models small). First I modeled number of alternative interpretations with word class. Alternative interpretations was not a significant component but word class remained significant,  $p=.02$ , suggesting even when number of alternative interpretations was controlled, word class continued to explain performance. Next I did the same thing with imageability. Word class remained significant,  $p=.01$ , but imageability did not, suggesting imageability could not account for performance beyond word class, but word class could account for outcomes even when imageability was controlled. Finally I tested word class and name agreement of S images. Although both were significant predictors when tested individually, neither factor was reliable when modeled

together. This result suggests that some of the common variance shared between each variable and outcomes was also shared between these variables (i.e., they were confounded). Word class was slightly more reliable, so it was moved forward into the final model.

The final model arrived at was the most reliable model of word learning for its size. The model was significant, Model Wald  $\chi^2(2)=27.37$ ,  $p<.001$ , with word class ( $p=.01$ ) and number of situations ( $p<.001$ ) as reliable model components. This model indicated the number of situations presented had a large effect, nearly tripling the odds of recognition (odds ratio = 2.82) when two situations were presented as when only one was. The model indicated the effect of word class was relatively small such that nouns stood 1.39 times greater odds of recognition than verbs, all other things being equal. Fig. 3 translates these model-specified effect sizes into word recognition probabilities by experimental condition.

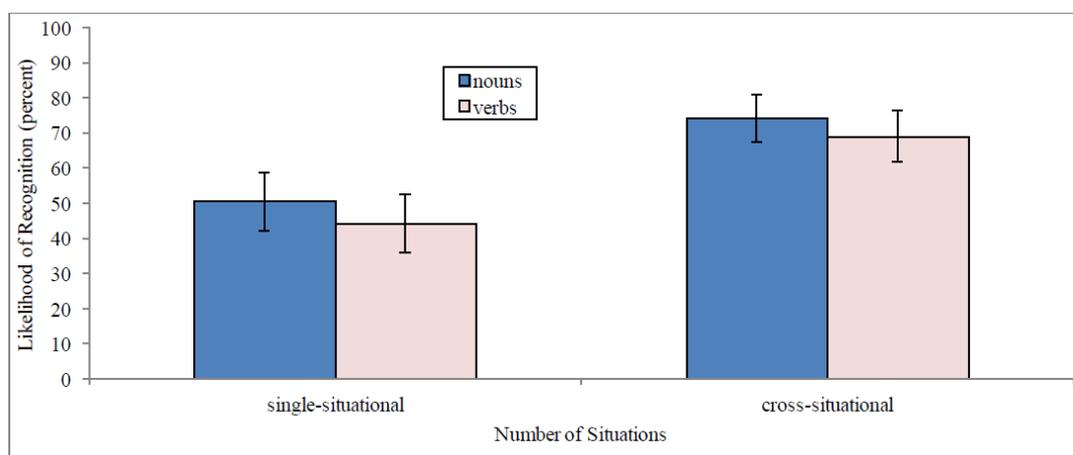


Figure 3. Logistic model-specified effects of number of situations and word class on probability of image recognition. Error bars represent 95% confidence intervals.

#### IV. DISCUSSION

Of eighteen predictors tested, three were significant—number of situations, word class, and name agreement. The large effect of cross-situational learning over single-situational learning agrees with Nitsch (1977), who found that increasing the number of learning contexts (stories) increased word learning. By re-defining “contexts” as line drawings of actors performing actions, the present study demonstrated better word learning when learners were presented with two different, juxtaposed contexts than a single, repeated context. This positive effect of two contexts rather than one could have been due to reduced referential ambiguity based on cross-situational target inference. Alternatively this effect could be mediated by identification of the target language syntax, or even by allowing participants to better parse words in each utterance. Any combination of these reasons could account for the current results. Further study would be needed to untangle these possibilities. However moving beyond this issue, on the results of the current study, incorporating multiple, varied learning situations in language lessons may provide an efficient way for learners to acquire new vocabulary by identifying and thus mapping words to meanings in otherwise-ambiguous mapping contexts. In another application, while review is important for learning, it is hypothesized that review could be made more effective by presenting examples that differ from those already presented.

Although nouns were better learned than verbs in the current study, it was unclear whether the effect of word class owed to name agreement differences between nouns and verbs, or if lower name agreement in verbs was driven by a name-agreement-limiting factor. As several colleagues and one reviewer pointed out, the use of static images as a learning method could handicap verb learning because while nouns can be perceived as stable, motionless entities in the real world, most verbs cannot. Thus teach their meanings through representation in static images might not be a very appropriate or efficient strategy. The greater referential clarity of nouns in images is not unusual: verb images are typically rated as more poorly depicting their intended referents (Kauschke & von Frankenberg, 2008; Ludington, in press), tend to be rated as more complex than noun images (Masterson & Druks, 1998), take longer to name (Humphreys, Riddoch, & Quinlan, 1988), and are not named as well (Kauschke & von Frankenberg; Ludington; present study). Evidently for one or more of these reasons, adults learning foreign vocabulary from still-media appear to be prone to a noun bias. One implication based on this view on the cause of a verb learning handicap is that it could potentially be removed by presenting verb vocabulary in a different medium—perhaps video or real-life—or with clearer illustrations. Illustrations might be made clearer by highlighting intended verb referents with color, motion marks (used in this study), or another visual technique to draw greater visual attention to verb-relevant aspects of images to support more effective verb vocabulary acquisition. This hypothesis should be further studied.

The prediction that the word class effect would hold only in single-situational learning conditions was not supported. The word class effect was fairly constant across levels of the number of situations presented. That cross-situational

learning did not close the learning gap between nouns and verbs suggests that the noun-verb gap is not due to whatever forces improved learning in cross-situational above single-situational conditions. That is, perhaps the observed verb handicap was not due to poorer parsing, syntactic understanding of, or mapping certainty for verbs than nouns. Perhaps verb learning was handicapped by another factor such as poorer name agreement. Future research is needed identify the cause of this observed noun bias among adults.

Name agreement among images presented second (S images) (but not first, or F images) in trials was a significant predictor of word learning. This could be explained as greater participant attention to the second images in each trial because these provided confirmatory evidence of mapping each target word to its meaning. Alternatively, because name agreement among S images (but not F images) was higher for nouns than verbs, its predictive significance could have derived from the significant effect of word class. Finally, name agreement among F images might not have been significant as a predictor because this measure was based on five times fewer response-judgments per word compared to name agreement among S images, and was therefore a more coarse measure of name agreement.

Name agreement S images could account for the word class effect, or word class could account for the name agreement trend (or there could be a third-variable cause). I support the first possibility. Verb may refer to a category of actions with unrelated goals, actors, patients, and instruments (e.g., consider the verb *to brush*). As Gleitman and colleagues noted, there is greater “surface variability in how verbs get realized ... within and across languages” (Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2006, p. 32). Verb situations naturally arise with great variance in terms of the goals, actors, patients, and tools involved, making feature-sharing across situations far less likely. Due to having more variable meanings, people may *learn to not learn* verbs too quickly in order to avoid mistaken mappings. Greater variance in the exemplification of word meanings across situations could make verbs harder to learn than nouns in general, as well as in the current study. Poorer verb name agreement might tap this root cause of less-reliable verb than noun learning or recognition.

The referential clarity (name agreement as the main predictor) hypothesis is but one explanation of the current results. Natural partitions / relational relativity hypothesis (Gentner, 1982) is one of the best-cited explanations of the noun bias in childhood; its application to adult foreign language learning might be considered. The natural partitions aspect of Gentner’s theory could explain why context image name agreement was higher for nouns than for verbs—that nouns are easier to identify because they can be more easily partitioned from their environmental context. The relational relativity aspect of Gentner’s theory, that verbs are difficult to acquire because they label a fragment of a change or action scene not usually “given” by the situation itself, does apply in this case because all verb stimuli were developed from English words, and all participants knew English. Therefore participants’ knowledge of English in this study put them at a relative advantage over first language learners by virtue of their knowing how to segment actions (i.e., how to set semantic boundaries around verb meanings) while viewing images. Thus the natural partitions, but not the relational relativity, aspect of Gentner’s theory lends well to explain the noun bias observed among the present sample of adults.

Some researchers (e.g., Naigles, 1990) have suggested morphosyntactic complexity could account for the noun bias. Verbs tend to allow and require more morphological inflection than nouns across instances and languages. Verbs, in the present study, were uttered either in the masculine or feminine form depending on the actors. It is possible that these inflections account for the found verb disadvantage. The use of inflection on verbs was not manipulated at the individual verb level in the present study, and was not analyzed; therefore this postulation remains speculative.

Finally, it could be that a difference in the target referent itself, and not the accompanying noun or verb in its context, contributed to better word learning in cross-situational over single-situational conditions. I am currently conducting an experiment to address this very question—to look for an effect of changing the target referent across learning repetitions. While people learn categories better by viewing multiple exemplars than by fewer or only one exemplar, a connection between this and word learning is not always drawn. Yet it could be argued words *are* categories—words represent all cases within a meaning-category. A paradigm developed by Roediger and McDermott (1995) from the work of Deese (1959, cited in Roediger & McDermott) demonstrates that word categories are often better remembered than their particular members. In their paradigm a list of words, related by some category, is learned while one centrally-related member is surprisingly not presented in the list. When tested, participants falsely recognize or recall the non-presented member, often with greater certainty than words actually seen. Generalizing from this effect, it would seem that number of related exemplars contributes positively to category learning, and that categories themselves have the potential to be learned even better than their constituents. Perhaps word learning, as a case of category learning, occurs more efficiently from learning a number of different exemplars than from learning a single exemplar repeatedly. It is plausible that exemplars may be remembered more for their category identity than for their exemplar identity.

Cross-situational learning was much more efficient than learning from single, repeated image contexts. This effect might be explained in terms of reducing the number of unknown referents in each image (Greenfield & Alvarez, 1980), increasing the ratio of cues to targets (Watkins & Watkins, 1975), increasing the number of familiar cues present at retrieval (Estes, 1955), revealing the syntactic pattern of the target language to allow meaning induction based on syntax (Gillette et al., 1999), or by allowing the learner to induct categories from different exemplars. This work adds to a growing number of studies showing benefits of cross-situational and contextual variation effects on word learning.

#### ACKNOWLEDGMENT

I would like to thank Patricia Greenfield for her feedback and guidance throughout my PhD work, from which this article was developed. Thanks also to Kay Lee and my many other research assistants for their valuable help, as well as to the Culture, Brain, and Development and the Language and Cognitive Development research labs at UCLA for their feedback at many stages of this project.

## REFERENCES

- [1] Akhtar, N. & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, 19, 347-358. doi: 10.1177/014272379901905703.
- [2] Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S., Pascual, L., Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, 75(4), 1115-1139. doi: 10.1111/j.1467-8624.2004.00729.x.
- [3] Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, 62, 145-154. doi: 10.1037/h0046888.
- [4] Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Technical report no. 257* No. BBN-R-4854.
- [5] Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of lexical acquisition. *Cognition*, 73(2), 135-176.
- [6] Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2006). Hard words. *Language Learning and Development*, 1(1), 23-64.
- [7] Greenfield, P. M. & Alvarez, M. G. (1980). Exploiting nonverbal context to promote the acquisition of word-referent relations in a second language. *Hispanic Journal of Behavioral Sciences*, 2(1), 43-50.
- [8] Humphreys, G. W., Riddoch, M. J. & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, 5(1), 67-104. doi: 10.1080/02643298808252927.
- [9] Jacobs, J. S. (2003). *Hebrew for dummies*. New York, NY: For Dummies.
- [10] Kauschke, C. & von Frankenberg, J. (2008). The differential influence of lexical parameters on naming latencies in German. A study on noun and verb image naming. *Journal of Psycholinguistic Research*, 37, 243-257. doi: 10.1007/s10936-007-9068-5.
- [11] Krippenoff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411-433. doi: 10.1111/j.1468-2958.2004.tb00738.x.
- [12] Levey, S. & Cruz, D. (2003). The first words produced by children in bilingual English / Mandarin Chinese environments. *Communication Disorders Quarterly*, 24(3), 129-136. doi: 10.1177/15257401030240030401.
- [13] Ludington, J. D. (in press). Noun bias in adulthood found to depend on test delay and learning method. *NIDA Journal of Language and Communication*, 19, xx-xx.
- [14] Masterson, J. & Druks, J. (1998). Description of a set of 164 nouns and 102 verbs matched for printed word frequency, familiarity and age-of-acquisition. *Journal of Neurolinguistics*, 11(4), 331-354. doi:10.1016/S0911-6044(98)00023-2.
- [15] Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(2), 357-374. doi: http://dx.doi.org/10.1017/S0305000900013817.
- [16] Nitsch, K. E. (1977). Structuring decontextualized forms of knowledge. Doctoral dissertation: Vanderbilt University, Nashville, TN.
- [17] Piccin, T. B. & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the Human Simulation Paradigm. *Language Learning and Development*, 3(4), 295-323. doi:10.1080/15475440701377535.
- [18] Roediger, H. L. & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21(4), 803-814.
- [19] Smith, L. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 1558-1568. doi:10.1016/j.cognition.2007.06.010.
- [20] Watkins, M. J. & Watkins, O. G. (1976). Cue-overload theory and the method of interpolated attributes. *Bulletin of the Psychonomic Society*, 7(3), 289-291.



**Jason D. Ludington**, born in 1980 in Miami, Florida, completed his Ph.D. in psychology at the University of California in Los Angeles, California in 2012. He is currently teaching and conducting research as a Lecturer in the Faculty of Psychology at Chulalongkorn University in Bangkok, Thailand.