



Cognitive Science (2013) 1–31
Copyright © 2013 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12035

Competitive Processes in Cross-Situational Word Learning

Daniel Yurovsky,^a Chen Yu,^b Linda B. Smith^b

^a*Department of Psychology, Stanford University*

^b*Department of Psychological and Brain Sciences and Program in Cognitive Science, Indiana University*

Received 26 March 2011; received in revised form 9 September 2012; accepted 17 September 2012

Abstract

Cross-situational word learning, like any statistical learning problem, involves tracking the regularities in the environment. However, the information that learners pick up from these regularities is dependent on their learning mechanism. This article investigates the role of one type of mechanism in statistical word learning: competition. Competitive mechanisms would allow learners to find the signal in noisy input and would help to explain the speed with which learners succeed in statistical learning tasks. Because cross-situational word learning provides information at multiple scales—both within and across trials/situations—learners could implement competition at either or both of these scales. A series of four experiments demonstrate that cross-situational learning involves competition at both levels of scale, and that these mechanisms interact to support rapid learning. The impact of both of these mechanisms is considered from the perspective of a process-level understanding of cross-situational learning.

Keywords: Statistical learning; Word learning; Language acquisition

In any statistical learning problem, the learning system is exposed to a stream of input information and tasked with discovering the underlying structure. However, the information in the stream may not all be of equal importance. Consequently, it is to the learner's advantage to employ a discovery process that biases the acquisition of new information in light of its likely informativeness (Billman & Knutson, 1996; Kruschke, 2001; Pearce & Hall, 1980). One way of producing such a process is through competition; a competitive discovery process is one in which evidence in favor of one structure acts as evidence against another structure. Mechanistically, this may be realized through processes in which stronger representations actively inhibit competing representations and thereby make them even weaker. Competition of this sort reduces noise in the input

Correspondence should be sent to Daniel Yurovsky, Department of Psychology, Jordan Hall, Building 01-420, Stanford University, 450 Serra Mall, Stanford, CA 94305. E-mail: yurovsky@stanford.edu

and sharpens the signal, allowing the learning system to more quickly converge on the underlying patterns. Competitive processes have been implicated in multiple linguistic phenomena, including speech segmentation (Norris, McQueen, & Cutler, 1995), visual, and spoken-word processing (Gaskell, 2003; Luce & Pisoni, 1998; Rodd, Gaskell, & Marslen-Wilson, 2004; McClelland & Elman, 1986; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and grammatical parsing (Smolensky, 1996; Vosse & Kempen, 2000). They are also found throughout the cognitive system in vision (Blakemore, Carpenter, & Georgeson, 1970; Masland, 2001), selective attention (Beck & Kastner, 2005; Desimone, 1988), action selection (Cisek, 2007), and decision making (McKinstry, Dale, & Spivey, 2008; Spivey, Dale, Knoblich, & Grosjean, 2010). Thus, it is reasonable to expect that competitive processes also play a role in statistical learning of language. Here, we consider the problem in the context of cross-situational word-referent learning.

In the cross-situational word learning task (Yu & Smith, 2007), participants are exposed to a series of learning trials each consisting of multiple words and referents. Within an individual trial, there is no information about which words map onto which objects (see Fig. 2). However, across multiple trials, co-occurrence frequencies between words and objects provide the statistical information sufficient to determine which word labels which object. Cross-situational word learning is a good experimental paradigm in which to study the role of competitive process in statistical language learning for three reasons. First, cross-situational word learning is an important phenomenon in its own right. There is general agreement that one of the difficult problems in word learning is ambiguity of reference (Gillette, Gleitman, Gleitman, & Lederer, 1999; Markman, 1990; Yu & Smith, 2007). To which of many potential objects in a scene does a label refer? Many studies have documented the processes by which children can reduce ambiguity in a given naming event, using social, pragmatic, and attentional cues. The cross-situational approach to language acquisition is concerned instead with the processes by which ambiguity can be reduced by accruing information *across* instances (Gillette et al., 1999; Siskind, 1996; Yu & Smith, 2007). Because it is a more recent proposal, however, cross-situational word learning is not well understood mechanistically (Medina, Snedeker, Trueswell, & Gleitman, 2011). Understanding, at a process level, how learners hone in on a coherent system of word-referent pairs in this laboratory task may help us to understand how learners in the real world acquire words even when, most often, no single naming event provides certain information.

Second, cross-situational word learning is a specific example of the general problem of statistical learning. Mechanisms that are found to be important for successful statistical word learning are likely to be important for other kinds of human statistical learning. For example, Aslin and colleagues have shown similar processes to operate in both auditory and visual statistical learning (Fiser & Aslin, 2001; Saffran, Aslin, & Newport, 1996), and Kirkham, Slemmer, and Johnson (2002) have argued that the two are instantiations of a common mechanism. Finally, the structure of information presented in cross-situational learning experiments allows for the systematic examination of competition at two levels: within and across trials.

First, since each individual learning trial presents multiple words and multiple objects, mappings between words and objects could compete such that mapping a word onto one object within a trial could inhibit its mapping onto other objects on that same trial. Competition at this *local* level, in which words within a single utterance compete for referential extent, has been a key component in an assortment of models of word learning (Frank, Goodman, & Tenenbaum, 2009; Fazly, Alishahi, & Stevenson, 2010; McMurray, Horst, & Samuelson, 2012; Siskind, 1996; Yu, 2008). In each of these cases, prior knowledge of the words and items determines the relative strength of competing mappings, but the items in competition—the words and objects—are present within a single, local, learning instance. *Local* competition of this kind finds the best within-trial solution. However, *local* competition also allows learners to leverage knowledge of some of mappings on the trial to learn other unknown mappings. For example, if the strength of a word-object mapping $A-a$ inhibits mappings other objects to A or other words to a , then the learner may be better able to find the right mappings for other words and referents.

Second, co-occurrence frequencies between words and objects provide statistical information across the entire set of trials such that evidence concerning elements not present in the current trial may also enter into competition. That is, evidence acquired about a word-object mapping on one trial can influence mapping that word onto other objects even when the original object is not present. We operationally define *global* competition, in contrast to *local* competition, as competition in which both elements of the competing mapping need not be present in the same learning trial. For example, *global*, cross-trial competition would be implicated if previous evidence for a mapping of word A to object a inhibited mapping word A to object b on a trial in which object a was not present. That is, by definition, *global* competition does not require the previously mapped object to be present on subsequent trials for the past learning of the mapping to inhibit new mappings. If the system prefers one-to-one mappings, this kind of cross-trial competition would make for a very powerful learning device in which latent knowledge strongly constrains the possibility of new learning. Such *global* competition could effectively reduce the noise in the learning environment given that words are frequently used in the absence of their referents (Gleitman, 1990). Competition at the *global* level has been conceptualized within Clark's (1987) principle of contrast, which posits that no two words can have identical referential extent. *Global* competition is also implemented in a number of associative models of language acquisition (Fazly et al., 2010; MacWhinney, 1989; Merriman, 1999; Regier, 2005; Yu, 2008), as well as in Bayesian models of word learning (Frank et al., 2009; Xu & Tenenbaum, 2007). In these latter models, the size-principle implements a competition between positing more word-object links to better explain the input and keeping the total lexicon small.

Competition at these two levels need not be mutually exclusive. In fact, some models explicitly include competitive mechanisms that operate at both the *local* and *global* levels (Fazly et al., 2010; Frank et al., 2009; Yu, 2008). By implementing competition at both levels, the models become significantly more robust to noisy input. *Global* competition speeds up the acquisition of individual mappings across time, and *local* competition lets these mappings be used to learn other mappings. However, despite theoretical proposals

about these kinds of competition in word learning, their use in many models of statistical word-referent learning, and their relevance to the mechanisms underlying statistical word-referent learning, they have not been empirically studied in the context of learning the meanings of words across multiple individually ambiguous trials. Accordingly, in the experiments that follow, we present human learners with cross-situational learning tasks designed to investigate how competition operates at each level and how these levels of competition interact.

To these ends, in each of the following experiments, participants were exposed to two types of words: *single* words, which co-occurred six times with only one correct referent, and *double* words, which each co-occurred six times with two different correct referents. Since these two referents shared the same word, the two mappings should be in direct competition. The experiments were designed to systematically manipulate the contexts in which these competing referents appeared—both within and across trials—so as to study *local* and *global* competition. In Experiment 1, both referents of each *double* word always co-occurred with that word within the same learning trial. This structure provides an opportunity for these mappings to compete both *locally*, within a single trial, as well as *globally*, across trials. Impaired learning of such *double* words compared to *single* words provide the first evidence for competition in cross-situational learning and sets the stage for disentangling the roles of *local* and *global* competition in the following experiments. In Experiment 2, the two competing referents were always encountered on separate trials. Since they could not compete *locally*, this experiment established the contribution of *global* competition. Experiment 3 was based on Experiment 2 but varied the order in which the referents of *double* words were encountered to examine how past learning influences *global* competition and to pit *local* and *global* competition against each other. Finally, Experiment 4 used a novel method for measuring trial-to-trial learning in cross-situational learning tasks to simultaneously measure competition at both the *local* and *global* level.

Before turning to the experiments, it is important to lay out what questions we do not intend to answer. First, while we are using two-to-one mappings to investigate the operation of competitive processes in cross-situational word learning, these experiments are not intended to be the definitive word on how children learn homonyms and synonyms. A number of excellent empirical studies have documented the developmental time course of children's difficulties with many-to-one mappings, and potential factors that may mitigate this difficulty (Casenhiser, 2005; Doherty, 2000, 2004; Liittschwager & Markman, 1994). These studies are aimed at elucidating the competitive mechanisms that underpin accumulation of co-occurrence information across multiple ambiguous instances. Understanding the operation and dynamic interaction of these processes is essential in moving toward a process-level account of statistical word learning (Smith, Colunga, & Yoshida, 2010). Second, competitive processes can be implemented in a variety of architectures: in neural networks (McClelland & Elman, 1986), in deductive hypothesis testing systems (Siskind, 1996), in exemplar models (Regier, 2005), in statistical machine learning systems (Fazly et al., 2010; Yu, 2008) and in Bayesian ideal observer models (Frank et al., 2009). They could be innately specified constraints specific to word learning (Clark, 1987; Markman,

1990), or they could be general properties of the cognitive system (Landau, Smith, & Jones, 1988; Regier, 2003). The present experiments do not discriminate among these different architectures or claims. However, what competitive processes in all these approaches have in common is a shared outcome: sharpening the signal present in noisy input, and a shared mathematical operation, normalization, in which strengthening one candidate weakens the others. The unifying criterion is that word-object mappings are acquired not only in parallel but also that the system treats them as dependent. Information about multiple mappings interacts. The empirical question asked in the present experiments is, in our view, prior to model building and differentiation (Yu & Smith, 2012). What roles do *local* and *global* competition play in cross-situational word learning? Although each of type of competition—*local* and *global*—could accelerate cross-situational word learning, the way in which they function and the way they interact implicate different learning mechanisms.

1. Experiment 1

Following the cross-situational learning paradigm (Yu & Smith, 2007), we asked participants to learn multiple word-referent pairs from a sequence of individually ambiguous learning trials. On each trial, learners saw four objects and heard four words but were given no information about which word referred to which object. In contrast to the original experiments, however, each word heard on a trial did not label exactly one of the present objects. Within an individual trial, some words labeled one object, some words labeled two objects, and some words labeled none of the objects. Fig. 1 shows a schematic of the learning situation.

If cross-situational word learning involves a competitive mechanism at either scale—*local* or *global*—learning multiple referents for the same word should be difficult, as mapping the word to one referent should inhibit mapping it onto the other. If competition is *local*, and word-object mappings compete within a single trial, then each referent of a *double* word should directly inhibit the other within each trial. If competition is *global*, then accruing information about one referent of a *double* word on one trial should inhibit

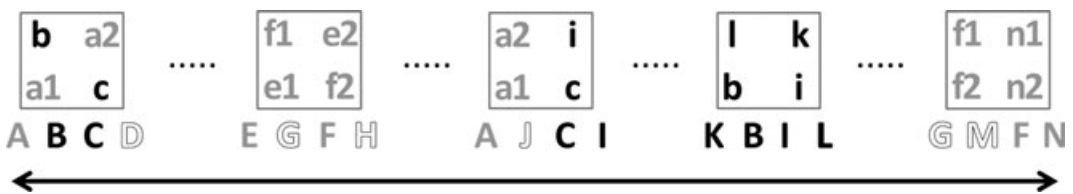


Fig. 1. A schematic of the trial structure of Experiment 1. On each trial, participants encountered four words and four referents, but the number of correct mappings for each word varied by type. Capital letters (e.g., A) indicate words and lower case letters (e.g., a) indicate referents. *Single* words each had one correct mapping per trial (e.g., B–b, C–c), *double* words each had two correct mappings per trial (e.g., A–a1 and A–a2, F–f1 and F–f2) and *noise* words were not mapped to any referent (e.g., D, G). *Single* words and their referents are depicted in black, *double* words and referents in gray, and *noise* words in white.

accruing information about that same word and its other referent on the next trial. Thus, the primary goal of Experiment 1 is to demonstrate that competition occurs in this experimental task by showing impaired learning in the case when there is more than one referent for a word. This is because while competition would affect learning the mappings of each word, it would be beneficial for learning the referents of *single* words, by limiting mappings to other referents, but detrimental for learning referents of *double* words, which requires learning two mappings.

1.1. Method

1.1.1. Participants

Forty-eight undergraduate students at Indiana University received class credit in exchange for volunteering. None had previously participated in any cross-situational learning experiments.

1.1.2. Stimuli and design

The stimuli were trials consisting of pictures of uncommon objects paired with auditorially presented pseudo-words (see Fig. 2). These pseudo-words were computer generated to broadly sample phonotactically probable English forms and were spoken by a synthetic female voice in monotone. In total, 18 unique objects and 18 unique words were produced. The set of 18 words was further divided into six *single* words, six *double* words, and six *noise* words. *Single* words behaved identically to words in other cross-situational learning experiments, co-occurring on each appearance with exactly one correct referent object. In contrast, each *double* word referred to two objects, both of which co-occurred with it on each appearance. In total, *single* words co-occurred six times with their correct referent, and *double* words co-occurred six times with each of their correct referents. Thus, co-occurrence frequency of each correct referent was equated. Finally, *noise* words

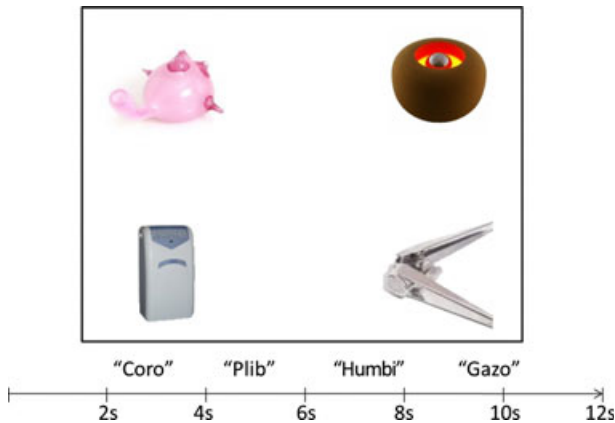


Fig. 2. An example of a single ambiguous learning trial in the cross-situational paradigm. Participants saw four objects and heard four pseudo-words that were mapped onto the objects. Words were assigned to objects randomly, and screen position of objects was independent of the order in which words were spoken.

co-occurred with approximately equal frequency with all objects in the set, and thus they did not map consistently onto any referent. This allowed us to directly compare learning of *double* words to learning of *single* words. *Noise* words were included to produce an equal number of words and referents on each trial, preventing learners from immediately noticing non one-to-one mappings.

Each training trial consisted of four words, heard serially, and four pictures of objects, one in each corner of the screen. One of the words was a correct label for each of the objects. Across the entire set of trials, each of the 18 words and objects appeared six times, resulting in 27 total 4 word \times 4 object training trials. Of these 27 training trials, 2 contained four *single* words, and 14 contained two *single* words, one *double* word, and one *noise* word. The remaining 11 contained two *double* words and two *noise* words. Thus, while each trial always consisted of four words and four objects, the within-trial mapping structure varied considerably from trial to trial and was rarely consistent with one-to-one mapping (2 of 27 trials).

After training, participants were tested for their knowledge of the referents of each word. On each test trial, participants heard one word from the training set and were asked to rank each of four objects in order of their likelihood of being the referent of that word. The set of alternatives for each *single* word consisted of its one correct referent, the referent of a different *single* word, and one correct referent for each of two different *double* words. The set of alternatives for each *double* word consisted of both of its correct referents, the referent of a *single* word, and one referent of a different *double* word. Since *noise* words mapped to no correct referents, the set of alternatives for each noise word consisted of the referent of one *single* word and one referent for each of three different *double* words.

1.1.3. Procedure

Participants were told that they would be seeing pictures of objects and hearing words, and that they should try to determine which referred to which. They then engaged in a training session, followed by a test session. At test, they received the following instructions: “You will now be tested on your knowledge of the word-object mappings. On each testing trial, you will hear one of the words from training and see four objects. Please rank the objects according to their likelihood of matching the word by clicking on them in order from most likely to least likely.” The test would not advance to the next trial until all four objects had been selected.

This method of testing was chosen to balance three important constraints: getting a clean and complete measure of participants’ knowledge, minimizing the chance of biasing them toward or away from one-to-one mappings, and keeping testing reasonably short in length. This ranking procedure provided a better solution than a number of other alternatives. For instance, a simple alternative-forced-choice test would either not provide information for both referents of each *double* word or could potentially bias participants to prefer or disprefer one-to-one mappings. An exhaustive matching test (e.g., “does this sound match this picture”) would require too many trials due to the large number of words and referents. Asking participants to name the pictures would not have these

problems, but it could interact quite differently with learned mappings and would be even more difficult to compare to previous cross-situational word-learning studies.

1.2. Results and discussion

To be given credit for knowing the correct referent for a *single* word, participants were required to rank it as the most likely referent. That is, it had to be their first guess. Test trials for each *double* word contained both of that word's correct referents. If a participant selected either of the correct referents as his or her first guess (e.g., $A—a1\ x\ x\ x$ or $A—a2\ x\ x\ x$), the participant was given credit for knowing at least one correct referent. To get credit for knowing both referents for a *double* word, the participant needed to select both of the referents in either order as guesses one and two (e.g., $A—a1\ a2\ x\ x$ or $A—a2\ a1\ x\ x$). Because these different kinds of words have different chance levels of accuracy, we developed a statistical analysis to correct for this and present these results after standard t-test analyses.

Fig. 3 shows average test accuracy for participants in Experiment 1. Participants showed better than chance knowledge of the referents for *single words* [$M_s = 0.454$, $SD_s = 0.264$, $chance = 0.25$, $t(47) = 5.29$, $p < .001$] and not only one [$M_{\geq 1} = 0.698$, $SD_{\geq 1} = 0.210$, $chance = 0.5$, $t(47) = 6.47$, $p < .001$] but both [$M_{both} = .301$, $SD_{both} = .146$, $chance = 0.17$, $t(47) = 6.32$, $p < .001$] referents for *double* words. Thus, relative to chance, participants learned both one-to-one and two-to-one mappings despite the ambiguity in their input.

However, participants were significantly less likely to learn both referents of a *double* word than the one referent of a *single* word [$t(47) = 3.68$, $p < .001$]. As chance-level

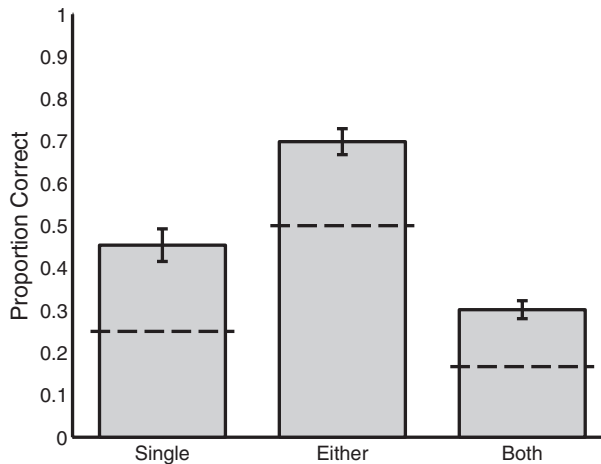


Fig. 3. Accuracy at test for each of the word types in Experiment 1. Error bars represent $\pm SE$. Dotted lines indicate chance levels of performance. Participants learned not only the referents of *single* words but also both referents for *double* words. However, they learned the two referents of *double* words significantly less well than the single referents of *single* words.

performance due to guessing on these two types of tests is different, a direct comparison of absolute accuracy is unfair. We present next the rationale for the analysis we developed to make a fair comparison. The starting assumption is that each correct answer at test could be due to one of two mutually exclusive possibilities: (a) the participant knows the correct mapping for the tested word, or (b) the participant guesses correctly. At test, suppose that a participant gave some number n of correct responses. This participant may have actually known any number of correct mappings less than or equal to n . The remaining correct responses must be explained by guessing. On the basis of the probability of guessing correctly in each test condition, our analysis produces a maximum likelihood estimate of the number of correct responses due to (a) word-object knowledge and the number due to (b) guessing. The number of known mappings can then be converted to a proportion, as in the standard analysis, and compared directly across word types. In all cases, this alternative analysis produced outcomes consistent with the standard t-test analyses of overall performance. Full details for the maximum-likelihood analysis can be found in the Appendix.

This guessing analysis was used to infer the proportion of *single* words for which participants knew the correct referent ($M = 0.39$), and the proportion of *double* words for which participants knew only one ($M = 0.34$) versus both ($M = 0.21$) referents. Because these data need not be normally distributed, we compare *single* word learning to *double* word learning with a non-parametric Mann–Whitney U -test. Participants were significantly less likely to learn both referents of a *double* word than they were to learn the one referent of a *single* word ($z = 2.996, p < .01$). That is, within the experiment, two mappings made up of a single word and two different referents do not act like two independent mappings (two words and two different referents). This means there is competition of some kind.

Experiment 1 thus provides evidence that competition is involved in cross-situational learning. But it does not determine the type of competition. Both of the correct referents of each *double* word were available on each trial, and thus they could have directly inhibited each other through *local* competition. Alternatively, stored information about co-occurrence from prior trials could have resulted in each referent, inhibiting the other on future trials through *global* competition. To help in disentangling these two kinds of competition, consider a sample trial containing words $\{A, B, C, D\}$ and referents $\{a1, a2, b, c\}$. One possibility is that cross-situational learning involves *local* competition in which pairing a word to an object on a given trial reduces the probability of pairing that word with other objects and that object with other words on that trial. In this trial, if A is paired with $a1$, it will be less likely to be paired with $a2$. If it is instead paired with $a2$, the strength of its *local* pairing with $a1$ will be reduced. Thus, on each trial containing A , on average less mapping strength will be accrued for $A-a1$ and $A-a2$ than for $B-b$ or $C-c$. Since $a1$ and $a2$ compete *locally* on each occurrence of A , learners will be less likely to know both *double* pairings ($A-a1, A-a2$) than a given *single* pairing ($B-b$).

If cross-situational learning involves primarily *global* competition, then the key factor might be not potential mappings presented within a trial, but rather all the mappings one has been exposed to across trials. In such a mechanism, the objects within a trial do not interact directly but do so through stored mapping information. Consider again the same

trial: {words: A, B, C, D ; referents: $a1, a2, b, c$ }. Through *global* competition, the previously acquired strength of mapping $A-a1$ will reduce the amount strength stored for $A-a2$ even without a *local* competition between $A-a1$ and $A-a2$, and previous knowledge about mapping $A-a2$ will reduce the amount stored for $A-a1$. Thus, even if no computation relates the words and objects within a trial to each other directly, *global* competition across trials should produce a decrement for learning both *double* mappings ($A-a1, A-a2$) relative to a *single* mapping ($B-b$). Following this rationale, the goal of Experiment 2 is to establish a role for *global* competition.

In Experiment 2, we expose participants to distributional information in which only one of the correct referents of each *double* word is available on each trial. If, for cross-situational learners, out of sight is out of mind, then without *local* competition in the sense of two vying mappings within the same trial, participants should not show impaired learning of double words. If, in contrast, knowledge of a non-present mapping influences the mappings formed within a trial, then learning of the *double* mappings should suffer relative to the *single* ones just as in Experiment 1.

2. Experiment 2

In this experiment, participants were again exposed to both *single* words and *double* words, but each appearance of a *double* word coincided with only one of its correct referents. Co-occurrences with each of the two referents were interleaved randomly over the entire set of training trials (see Fig. 4 for a schematic). If competition operates only at the *local* level, participants in Experiment 2 should have been able to learn the two referents of each *double* word as well as the one referent of each *single* word. In contrast, if there is competition at the *global* level, learning of *double* words should again have been impaired.

2.1. Method

2.1.1. Participants

Forty-eight undergraduate students at Indiana University received class credit for volunteering. None had previously participated in Experiment 1 or any other cross-situational learning experiments.

2.1.2. Stimuli and design

Stimuli for Experiment 2 was similar to those for Experiment 1. Trials were again composed of four pseudo-words and four pictures of objects; 12 words and 18 objects were drawn from the set used in Experiment 1. The 12 words were divided into six *single* words which mapped onto one object referent each and six *double* words which mapped onto two objects each. Trials were presented as before: Four objects were seen in the four corners of the screen, and four words were presented serially from loudspeakers. Each of the *single* words appeared six times, and each of the *double* words appeared 12 times, six times with each of its correct referents. This again produced 27 total 4 word \times 4 object training trials.

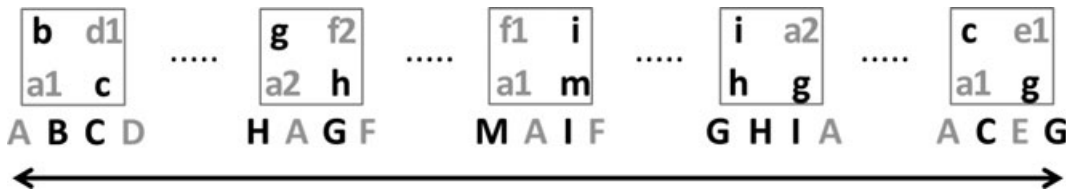


Fig. 4. A schematic of the training structure in Experiment 2. In contrast to Experiment 1, the two referents of each *double* word now appeared randomly on different trials throughout training. Overall, exposure to each correct co-occurrence frequency was identical. *Single* words and their referents are depicted in black; *double* words and referents are depicted in gray. There were no *noise* words in Experiment 2. Capital letters indicate words and lowercase letters indicate referents.

After training, testing proceeded as in Experiment 1. For each word, participants were again asked to rank four possible referents in order of likelihood. The set of alternatives for each *double* word contained both of its correct referents along with two foils; the set of alternatives for each *single* word contained its one correct referent and three foils.

2.1.3. Procedure

Participants were given the same instructions as before—that they would be seeing objects and hearing words and that they should determine to which object each word referred. They then engaged in a training session, followed by a test session. Testing was again assessed via 4-alternative ranking.

2.2. Results and discussion

As in Experiment 1, credit was given for knowledge of the referent of a *single* word if the correct referent was the participant's first choice. Credit for knowing at least one referent of a *double* word was given to participants who selected either of the word's referents as their first choice. Credit for knowing both referents was given only if choices one and two were both correct referents in either order.

Fig. 5 shows test results for Experiment 2. Participants successfully learned referents for *single* words [$M_s = 0.40$, $SD_s = 0.247$, $chance = 0.25$, $t(47) = 4.09$, $p < .001$] and, again, they learned referents for not only one [$M_{\geq 1} = 0.58$, $SD_{\geq 1} = 0.277$, $chance = .5$, $t(47) = 2.08$, $p < .05$] but both [$M_{both} = 0.24$, $SD_{both} = 0.203$, $chance = 0.17$, $t(47) = 2.49$, $p < .05$] referents of the double words. As in Experiment 1, although participants learned all types of mapping at above-chance levels, they were significantly less likely to learn both referents of a *double* word than they were to know the one referent for a *single* word [$t(47) = 3.81$, $p < .001$]. We again applied the guessing analysis to estimate the proportion of *single* words ($M = 0.31$) participants knew, and the proportion of *double* words for which participants knew only one ($M = 0.24$) versus both ($M = 0.15$) referents. As in Experiment 1, participants were significantly less likely to learn both referents of a *double* word than they were to learn the one referent of a *single* word ($z = 3.03$, $p < .01$), a result that indicates competition across trials.

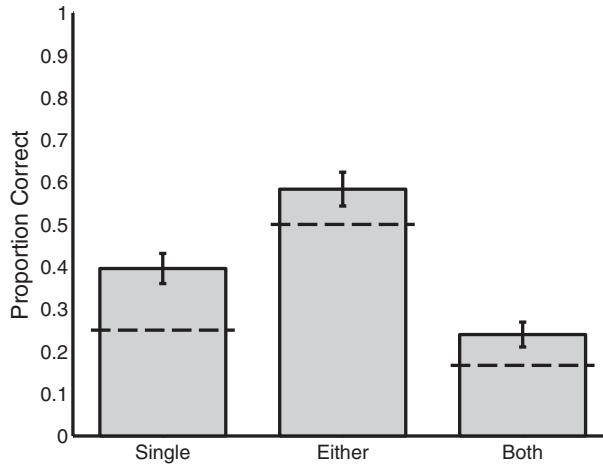


Fig. 5. Accuracy at test for each of the word types in Experiment 2. Error bars represent $\pm SE$. Dotted lines indicate chance levels of performance. Participants learned the referents of *single* words, and both referents for *double* words, but less well than in Experiment 1. Furthermore, they were significantly more likely to learn the one referent of a *single* word than both referents of a *double* word.

The results thus suggest *global* competition. Knowledge of one word-object mapping alters the acquisition of a second referent for the same word even when the first referent is not available within the same trial. It is worth noting that while Experiment 2 had the same number of learning trials as Experiment 1, and learners were exposed to the same word-object co-occurrence information, there were differences in timing. In Experiment 1, learners would have to divide their attention between the two referents for a double word within a single trial. Because in Experiment 2 participants did not have to divide their attention in this way, we might have expected improved learning (Yu & Smith, 2011). However, this was not the case. Instead, learning in Experiment 2 was slightly worse. One possible explanation for this result is that each occurrence of one referent of a *double* word meant a non-occurrence of the other. Participants thus could have treated this non-occurrence as implicit negative evidence (see, e.g., Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). Thus, while implicit negative evidence is treated as deriving from alternative positive evidence in competition models (e.g., Merriman, 1999, see also Xu & Tenenbaum, 2007, for a review), non-occurrence of a word-object mapping could also have a separate additional role in reducing evidence for that mapping. Since assessing this role empirically would require training participants with an unequal number of words and objects, and thus very probably change their learning strategies, we do not pursue it further here. Nonetheless, understanding the independent contributions of competition and non-occurrence would be an interesting question for further research.

In either case, since learning results were no better in Experiment 2 than Experiment 1, *global* competition is likely to be a potent force. This *global* competition could be implemented in one of two ways: either after all information has been acquired, and thus independently of *local* competition (e.g., Frank et al., 2009), or trial-by-trial as

information is being accrued and thus potentially interacting with *local* competition (e.g., Fazly et al., 2010). One way to measure whether *global* competition operates trial-by-trial is through order effects.

If *global* competition occurs on-line from trial-to-trial, then the order in which participants are exposed to the two referents for each *double* word should matter. If *global* competition operates through batch-like statistics at the end of learning (or at test), order effects would not be expected. Accordingly, in Experiment 3, the trials from Experiment 2 were rearranged such that participants received all exposures to one correct referent for each *double* word before seeing the other. If *global* competition occurs at the end of training, then learning results should be identical. If, in contrast, trial-by-trial *global* competition is the primary mechanism of competition, then participants should be significantly more likely to acquire the first referent of a *double* word but significantly less likely to acquire both. Such a pattern was found by Gebhart, Aslin, and Newport (2009) in a similar statistical speech segmentation task.

3. Experiment 3

In Experiment 3, participants again encountered cross-situational learning trials in which half of the words co-occurred with equal frequency (six times) with two correct referents. As in Experiment 2, only one of these correct referents appeared on each trial. However, whereas in Experiment 2 these appearances were randomly interleaved, in Experiment 3 they were strictly ordered (see Fig. 6). Trials in the first half of the experiment presented only one of the two correct referents of each *double* word, and trials in the second half presented exclusively the other correct referent. Thus, participants received all six exposures to one correct referent before ever seeing the other.

If competition is primarily *global*, and occurs only after all training information has been accumulated, there should be no effect of the temporal order of individual trials and performance in Experiment 3 should be identical to Experiment 2. In contrast, if *global* competition emerges trial-by-trial and does not interact with the other local mappings that may be formed within a trial, then one would expect a clear order effect in Experiment 3

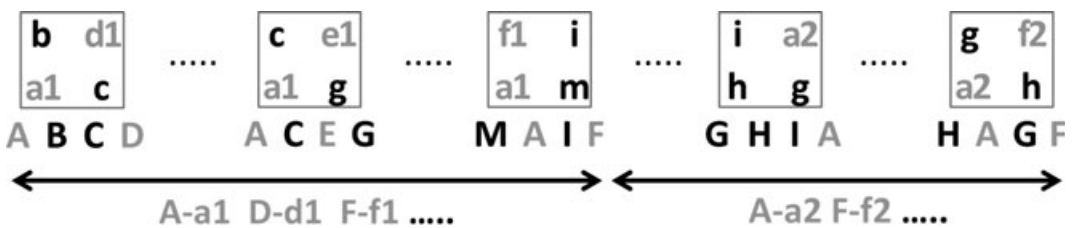


Fig. 6. A schematic of the training structure in Experiment 3. In contrast to Experiment 2, the two referents (called early and late) of each *double* word were separated across training. The first six occurrences of a *double* word appeared with its early referent, and the second six appeared with its late referent. *Single* words and their referents are depicted in black; *double* words and referents are depicted in gray. Capital letters indicate words and lowercase letters indicate referents.

with the first referent being learned better than the second for the *double* words. Thus, both hypotheses about *global* competition predict that *double* words will show a decrement relative to *single* words as found in Experiments 1 and 2, and the key question is whether that decrement is greater for the second learned referent (a result that implicates trial by trial *global* competition) or roughly equivalent (a result that implicates post-learning or batch competition). Predictions are more complicated if there are interactions between *global* and *local* competition. Briefly, *double* words are not the only words presented on each trial and if *local* competition works within a trial for learners to find the one best referent for each word, then *local* within-trial competition in the second half of the experiment could help the learner map *double* words to their correct second referents.

3.1. Method

3.1.1. Participants

Forty-eight undergraduate students at Indiana University received class credit for volunteering. None had previously participated in Experiments 1 or 2 or any other cross-situational learning experiments.

3.1.2. Stimuli, design, and procedure

Stimuli for Experiment 3 were identical to those in Experiment 2 except for the order in which trials were presented. In Experiment 2 the two referents of each *double* word were interleaved randomly across the set of training trials. In contrast, in Experiment 3 they were ordered such that one of the referents co-occurred with the first six appearances of a double word, and the second referent co-occurred with the remaining six appearances. Fig. 6 shows this new training scheme. In the analysis, we will refer to the referent which occurred first as the early referent, and the one which occurred second as the late referent. The testing procedure was also identical to that of Experiment 2.

3.2. Results and discussion

As in previous experiments, knowledge of word-object mapping was assessed via four-alternative ranking. The set of four alternative objects for each word was constructed identically to Experiment 2. Fig. 7A shows mapping accuracy for participants in Experiment 3.

Overall, participants successfully learned referents for *single* words [$M_s = 0.45$, $SD_s = 0.30$, $chance = 0.25$, $t(47) = 4.69$, $p < .001$] and again for not only one [$M_{\geq 1} = 0.73$, $SD_{\geq 1} = 0.24$, $chance = 0.5$, $t(47) = 6.53$, $p < .001$] but both [$M_{both} = 0.40$, $SD_{both} = 0.30$, $chance = 0.17$, $t(47) = 5.42$, $p < .001$] referents for *double* words. However, in contrast to the previous experiments, participants were not significantly less likely to know both referents of a *double* word than they were to know the single referent for a *single* word [$M_s = 0.45$, $M_{both} = 0.40$, $t(47) = 1.53$, *n.s.*]. The guessing analysis again confirmed standard t-test analysis. The analysis described in the Appendix allowed us to infer the proportion of *single* words ($M = 0.36$) participants

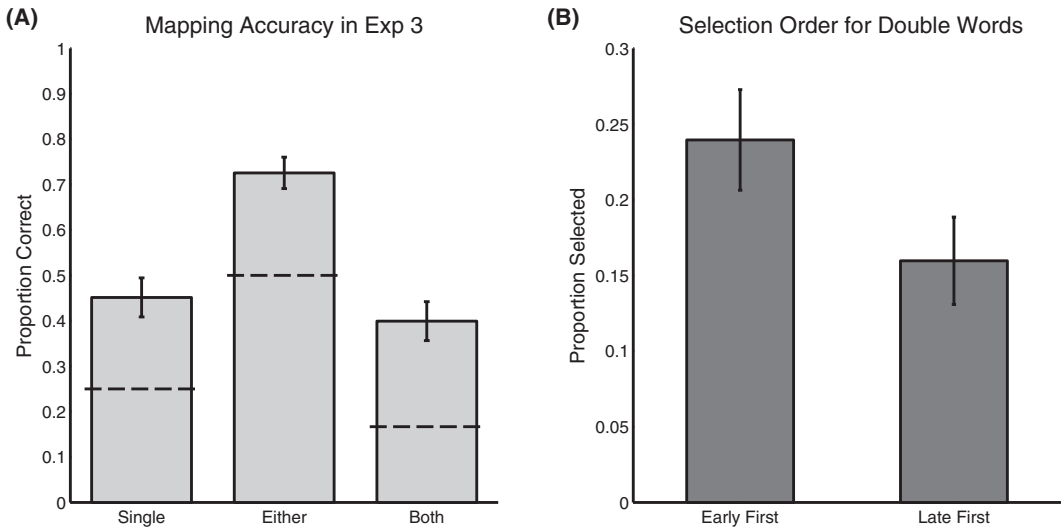


Fig. 7. Test outcomes for Experiment 3. (A) Accuracy at test for each word type. Error bars represent $\pm SE$. Dotted lines indicate chance levels performance. Participants learned the referents of *single* words, and both referents for *double* words. In contrast to participants in Experiments 1 and 2, participants in Experiment 3 were not significantly more likely to know the one referent of a *single* word than both referents of a *double* word. (B) When participants learned both of the referents for a *double* word, they selected them in one of two orders: early first or late first. The bars indicate the proportion of each of these two selection orders. Error bars represent $\pm SE$. Participants were marginally more likely to select the early referent of a *double* word first.

knew, and the proportion of *double* words for which participants knew only one ($M = 0.22$) versus both ($M = 0.34$) referents. A Mann–Whitney U -test showed that participants in Experiment 3 did not learn *single* words more successfully than they learned both referents of *double* words ($z = 0.122$, $n.s.$). Thus, in contrast to previous experiments, the results do not show direct evidence of competition. The results therefore do not support either of the two *global* competition alone hypotheses, both of which predicted some degree of decrement for *double* words versus *single* words (for related results in phonological contrast learning, see also Perrachione, Lee, Ha, & Wong, 2011).

However, although this measure of participants' test accuracies did not show evidence of a competition, there is one more place to look. Since the two referents of each *double* word appeared in a consistent order, the early referent for the first six occurrences and then the late referent for the second six, one might expect them to be treated differently at test. In particular, if *global* competition is applied trial-by-trial as information is accrued, then the early referent should be better learned. Did order matter? When participants correctly selected both referents in their first two guesses ($M_{both} = 0.4$), they were marginally more likely to select the early referent first [$M_e = 0.24$, $SD_e = 0.23$, $M_l = 0.16$, $SD_l = 0.20$, $t(47) = 1.77$, $p = .08$]. These results are shown in Fig. 7B.

Experiments 1 and 2 showed that competition plays a role in cross-situational learning, and that mappings can inhibit each other even when both are not present within a single

trial. That is, there is some form of *global* competition. But the structure of Experiment 3 removed these competitive effects almost entirely. The likely explanation is that *local* competition counteracted the effects of *global* competition from previously learned word-referent pairings. In contrast to Experiments 1 and 2, learners received six exposures to *double-early* mappings in the first half of training. Because of the lack of ambiguity in this *to-be-double* pairing, *local* competition in the first half of training may have allowed participants to acquire significantly more information about the correct mappings of the single words on these trials. Then, while learning *double-late* mappings would be inhibited by *global* competition, it would be supported by *local* competition from the other already well-learned mappings present within a trial.

By this interpretation, cross-situational learning stems from both *global* and *local* competition, and both processes unfold trial-to-trial as information is acquired. *Global* competition protects old mappings from noisy information, and *local* competition leverages prior mapping knowledge to speed the acquisition of new mappings. If we are correct, then we should be able to uncover more direct evidence of competition not only at test but during training. The next experiment was designed to test these predictions. In Experiment 4, participants were exposed to the same information as in Experiment 3, but this time, measures of learning were collected on each trial. If our hypothesis is correct, then these trial-by-trial measures should provide evidence of competition at both levels.

4. Experiment 4

The ranking procedure used to assess learning in the previous experiments provided information only about participants' knowledge at the end of training. The goal of this experiment was to measure learning trial by trial as training unfolds. As testing may interact with training (McClelland, 2006; Roediger & Karpicke, 2006), we chose the indirect approach of asking participants to give their confidence that they knew the correct label for each object. More specifically, at the end of each training trial, a bar containing the numbers 1-10 appeared under each object. Participants were asked to indicate their level of confidence in their knowledge of the correct label for each object on this scale. Since these judgments were made on each trial, analyses could be conducted to determine how knowledge for individual words grew across exposures, and how knowledge of different words interacted within a single trial. From this data, we could then ask more detailed questions. For instance, can a participant's knowledge of some of the words on a trial be used to predict knowledge for others?

4.1. Method

4.1.1. Participants

Forty-eight undergraduate students at Indiana University received class credit for volunteering. None had participated in any of the previous experiments or any other cross-situational learning experiments.

4.1.2. Stimuli and design

Stimuli for Experiment 4 were identical to those in Experiment 3. *Double* words again co-occurred six times with their early referent in the first half of training, and then six times with their late referent in the second half of training. Participants were not made aware that such a switch would occur.

4.1.3. Procedure

Training in Experiment 4 proceeded in exactly the same manner as Experiment 3 except for the introduction of knowledge judgments. At the end of each trial, a bar containing the numbers 1–10 appeared under each of the four objects on the screen. Using this scale, participants were asked to indicate their level of confidence in their knowledge of the correct label for each object. After a judgment was made for all four objects, a “Next” button appeared on the screen. Participants could revise any of their judgments by clicking on a bar again. When participants clicked “Next” button, the next trial began.

Due to an extended time in training, testing in Experiment 4 was also slightly different from the previous experiments. Whereas participants in Experiments 1–3 were asked to make four responses at test—to rank the four alternatives in order of likelihood—participants in Experiment 4 made only one response. They gave only their best guess.

4.2. Results and discussion

Participants were tested for their knowledge of correct referents for each word in the training set by selecting a single best referent from a set of four alternatives. As a result, for *double* words, participants needed to make an active choice between the early and late referent. Before examining participants’ knowledge judgments, we first verify that test results are comparable to previous experiments.

Fig. 8A shows participants’ accuracies at test. As in previous experiments, participants successfully learned the referents for *single* words [$M_s = 0.61$, $SD_s = 0.31$, $chance = 0.25$, $t(47) = 8.07$, $p < .001$] and selected either the early or late referent of each *double* word at levels significantly above chance [$M_d = 0.81$, $chance = 0.5$, $SD_d = 0.22$, $t(47) = 9.57$, $p < .001$]. Between these two referents, participants were significantly more likely to choose the early over the late referent [$M_e = 0.47$, $SD_e = 0.25$, $M_l = 0.34$, $SD_l = 0.22$, $t(47) = 2.18$, $p < .05$, shown in Fig. 8B]. This is the same pattern of results found in Experiment 3, again providing evidence for competition at the *global* level in the preference for early over late referents.

When compared to participants in Experiment 3, those in Experiment 4 were significantly more likely to choose the correct referent of each *single* word ($t(94) = 2.57$, $p < .05$) and were marginally more likely to choose either of the two referents of each *double* word [$t(94) = 1.69$, $p < .1$]. Why did participants perform better in Experiment 4 than in Experiment 3? One possibility is that the addition of the knowledge judgments trial by trial gave participants more time to encode the words and objects presented on each trial. Alternatively, the benefit could have been not just a function of time but also a function of increased engagement in the task which may have resulted in deeper processing

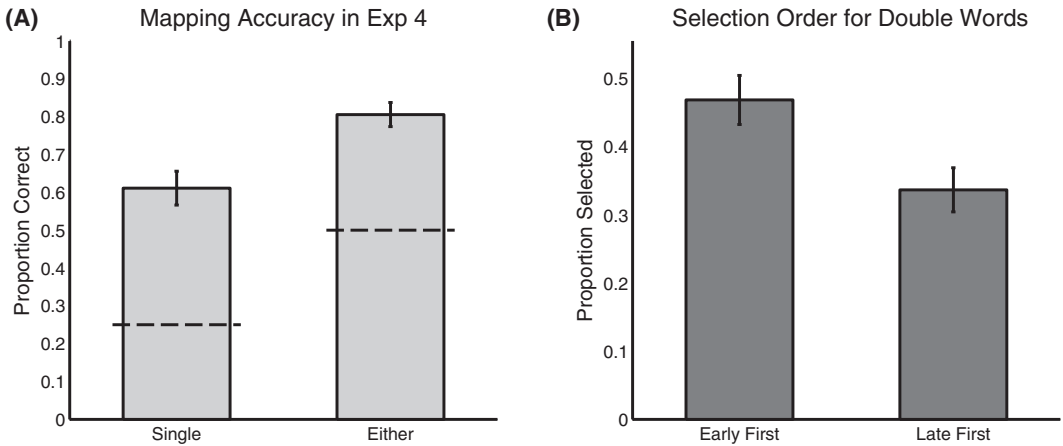


Fig. 8. Test outcomes for Experiment 4. (A) Accuracy at test for each word type. Error bars represent $\pm SE$. Dotted lines indicate chance levels performance. Participants learned the referents of *single* words, and at least one referent for the *double* words. (B) When tested with the *double* words, participants could select either the early or the late referent. The bars indicate the proportion of each of these two selections. Participants were significantly more likely to select the early referent as the correct referent of a *double* word.

(Craik & Tulving, 1975). Finally, it is possible that the difference in testing itself may have driven the effect. However, the overall pattern is the same as in Experiment 3 and thus the analyses of trial-by-trial judgments over the course of training should be revealing about the cross-trial roles of *global* and *local* competition.

To this end, a first step is to validate the judgment measures by asking whether participants' trial-by-trial knowledge judgments were correlated with learning results at test. If so, they can be used as to infer learners' internal learning state in real-time learning. One simple way to estimate the predictive power of these judgments is to ask whether they predict a correct result at test. We thus analyze these judgments via logistic regression, with test response as the outcome variable, and confidence judgment on each occurrence of an object as predictors. Participants responses at test were coded either 0 (selected incorrect referent) or 1 (selected correct referent). Regression coefficients for this model, seen in Table 1 below, show that participants' judgments began to be significant predictors of test response from the second occurrence of an item.

When tested for their knowledge of the referents for *double* words, participants were significantly more likely to select early referents. When does this effect arise? Fig. 9 shows mean judgment scores provided by participants on each of the six occurrences of each type of referent. While the referents of *single* words and early referents of *double* words grew similarly—with the last occurrence of an early referent even judged marginally better known ($t(47) = 1.90, p = .06$), judgments of late referents were strikingly different. From their second occurrence—the first which is a significant predictor of test behavior, judgments for late referents were significantly lower than those of either other type of referent [$t_{single-late}(47) = 3.21, p = <.01, t_{early-late}(47) = 2.78, p = <.01$]. *Global* competition was thus in effect from the earliest exposures of a late referent. Further,

Table 1

Regression coefficients (Betas) for logistic regression predicting test accuracy on each occurrence

Occurrence	1	2	3	4	5	6
Betas	-.022	0.106***	0.132***	0.170***	0.172***	0.200***

Note. *** = $p < .001$.

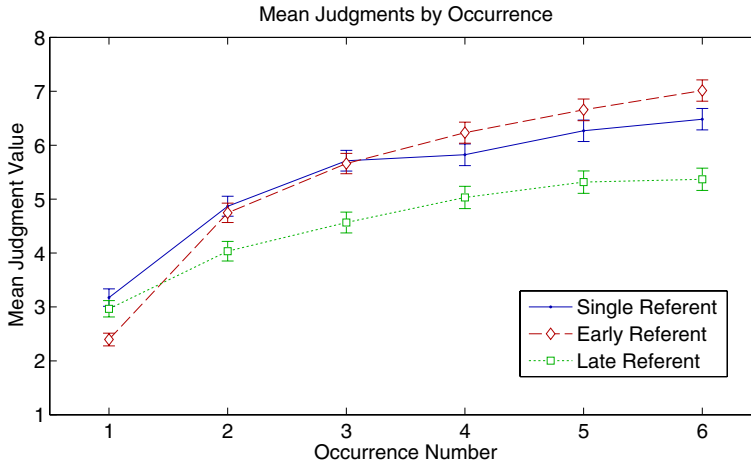


Fig. 9. Mean judgment value given for each referent as a function of occurrence number (1–6) and referent type (single, early, or late). All judgments grew over occurrence number, but single and early judgments were significantly higher than late judgments from their second occurrence.

evidence of competition is seen all the way through to the judgments made on the late referents' final appearances. This means that for any given occurrence of a *double* word and its late referent—for example, the third occurrence—the knowledge judgment was likely to be lower than the comparable (e.g., third) occurrence of that *double* word and its early referent. Thus, prior knowledge inhibited the acquisition of new knowledge through *global* competition.

What about *local* competition? Because judgments were collected on each trial, we can directly test whether *local* competition within each trial may have helped participants in Experiment 4 learn word-object mappings. To this end, we consider a linear regression analysis, in which the outcome to be predicted is the confidence rating assigned to a referent object on any given trial. We first include referent type as a predictor. Then, because knowledge should build on itself, we expect judgments on successive occurrences to be correlated. We thus add judgment rating on the previous occurrence as a further predictor. Finally, if participants are applying *local* competition, then knowledge of these other objects should allow their labels to be ruled out as the potential label of the target object. We thus include an additional regressor—the mean of the judgments made for all three other referents on their previous occurrences. In cases where a referent had not previously occurred, it was assigned a judgment value of 1. As shown in Table 2, all regressors were

Table 2
Coefficients (Betas) for linear regression predicting referent judgment values

Parameter	Beta
Intercept	.858***
Single	.690***
Early	.879***
Previous judgment	.476***
Avg. other ref. judgments	.407***

Note. *** = $p < .001$.

significant predictors. Thus, not only did previous knowledge of a word-referent pairing predict its judgment, previous knowledge of other co-occurring referents did as well.

By way of example, consider a trial containing objects $\{a,b,c,d\}$. Suppose further that this is the fourth occurrence of a , the third occurrence of b , the second occurrence of c , and the fifth occurrence of d . Now suppose we are trying to predict the confidence that a participant had for his or her knowledge of the correct label for object a . What the regression model shows is first, that the confidence assigned to a on its previous (third) occurrence can be used to predict confidence on this occurrence. However, beyond this, the model shows that the confidence assigned to each of the other referents on their previous occurrences (b 's second, c 's third, d 's third) adds information about the value assigned to a . That is, knowledge of some of the referents of each trial increases the information acquired about the mappings of other referents.

Because the first half of training consisted of words that mapped onto object one-to-one—only single and early referents—participants were able to learn some of the *single* words referents. Then, in the second half of training, participants used *local* competition within-trials to reduce uncertainty about the correct labels for late referents. This explains the acquisition of the second referents of *double* words.

Overall, the results provide strong evidence for *global* competition: Learning a word's early referent inhibited the acquisition of that word's late referent from the outset, and continued to do so through all of its occurrences, consistent with a trial-by-trial form of *global* competition. However, participants nonetheless learned two referents for each *double* word. Analysis of the growth in knowledge judgments on each trial suggests that *local* competition—competing mappings within a trial—helped participants learn the second referent. Once some of the *single* word-referent mappings were sufficiently strong, these words would be inhibited as possible candidates for the late referents. Thus, competition at the *local* level was able to compensate for the operation of competition at the *global* level.

5. General discussion

In cross-situational learning experiments, participants are exposed to a series of individually ambiguous learning trials, each containing multiple words and objects.

Nonetheless, across these trials, they successfully learn multiple word-object mappings. One way that humans could do this would be to maintain only a single hypothesis for the referent of each word and to update this hypothesis only when it is disconfirmed (Medina et al., 2011). Such a model is plausible, but it would need significant revision to explain above-chance learning of two-to-one mappings in all four of these experiments, as well as the results of other recent studies of statistical word learning (e.g., Smith, Smith, & Blythe, 2011; Vouloumanos, 2008). An alternative possibility for successful cross-situational word learning would be to track and store all of the co-occurrences on all trials, and then to determine the correct mappings at the end (Frank et al., 2009; Vouloumanos, 2008; Yu, 2008). However, these experiments make such a model unlikely as well. Instead, these results suggest a middle ground: Learners may use past knowledge and the set of words and objects available on each individual trial to determine which subset of the input they store. In this way, they may select a more coherent set of mappings within each trial and consequently, trial-by-trial, reduce the ambiguity in their input. The entire pattern of results provides clear evidence for both *global* competition across trials and also *local* competition within trials. These results thus suggest that a mechanistic understanding of statistical learning will require understanding how within-trial information activates past learning, and how cross-trial learning and within-trial input interact. This level of analysis will be critical to understanding how statistical learning and memory constraints interact to produce what look to be qualitatively different learning strategies as ambiguity scales up (Medina et al., 2011; Smith et al., 2011; Yu & Smith, 2012).

The four experiments in this article contribute to this discussion through three main findings. First, they provide evidence that both *local* and *global* competition are involved in cross-situational learning. That is, the set of mappings formed on a given trial depends both on prior information about the words and objects within the trial, and on prior information about the mappings of the present words with other, non-present, objects. Second, they show that *global* competition is not an operation that is applied merely at decision time but operates trial-by-trial as does *local* competition (Experiments 3 and 4). Finally, they provide evidence that the two mechanisms interact, such that *local* competition can compensate for *global* competition to support the acquisition of two-to-one mappings. Even though *global* competition can be quite potent (Experiment 2), the *local* assignment of word-object links within a trial is a dominant force (Experiments 3 and 4). These results lead naturally to a process-level account of cross-situational learning.

5.1. *Local and global interaction: Toward a processes-level account*

These four experiments implicate a cross-situational word learning process in which words and objects compete for mapping strength (MacWhinney, 1989). Trial-by-trial, the amount of evidence acquired for a particular word-object pairing depends both on the words and objects present, and on memory for their other associations. Here we sketch a conceptual, process-level account of how such a system might function. This is important because the vast majority of recent models of cross-situational learning (e.g., Blythe, Smith, & Smith, 2010; Frank et al., 2009; Fazly et al., 2010; Yu, 2008) have taken a

batch-statistics or corpus-level approach (although, c.f. Horst, McMurray, & Samuelson, 2006) and have not compared their learning mechanisms directly to measures of human performance. The only recent counterexample, from Smith et al. (2011), models a version of cross-situational learning in which the set of candidate referents for each word is completely non-overlapping. Thus, there was no chance for competition in their task.

In the introduction, we noted explicitly that competition—at both levels—could be, and has been, implemented in a variety of model architectures. We again reaffirm that statement here. However, for the sake of clarity of exposition, we will adopt associative terminology to present the conceptual model. We begin by considering what information might be acquired on a trial in the middle of training with four words $\{A,B,C,D\}$ and four referents $\{a,b,c,d\}$. Learning on this trial can be considered as a combination of two processes by which associations in memory are updated (see also Frank et al., 2009; Fazly et al., 2010). In one process, words and objects are sorted out *locally*. The probability that a word and object are paired (e.g., $A-a$) is proportional to the prior strength of that association ($A-a$), and inversely proportional to (a) the strength of the associations between that word and all other objects present on that trial ($A-b$, $A-c$, $A-d$), and (b) the strength of the associations between that object and all of the other words present within the trial ($B-a$, $C-a$, $D-a$).

If the word D and object d have not been previously encountered, *local* competition would produce a bias to pair the two, as all of the other words would inhibit mapping D onto their more likely referents (e.g., $D-a$ would be inhibited by $A-a$, $B-a$, $C-a$; $D-b$ would be inhibited by $A-b$, $B-b$, $C-b$, etc.). *Local* competition can also support the acquisition of a second referent (as in Experiments 3 and 4). If referent b has never been previously encountered, but $A-a$, $C-c$, and $D-d$ are well established, and word B has been previously strongly associated with some other referent x , then, the association of each of the words with b will be inhibited by their associations with the other objects on the trial. However, this inhibition will be much larger for A , C , and D as the trial contains objects with which they are strongly associated. For B , the trial contains only weak associates, and thus it will be most likely to be paired with b .

Via the second process, *global* competition moderates the updates made to associative strengths between the words and objects that have been paired off. In this process, the change in associative strength between a word and object (e.g., $A-a$) is proportional to the strength of the current association in memory and inversely proportional to the other associations in memory for that word (e.g., $A-b$, $A-e$, $A-g$). Mechanistically, this could consist of a single normalized update, or it could be a simple associative update followed by normalization via a different mechanism (see, e.g., Apfelbaum & McMurray, 2011). Either would produce an effect of *global* competition. Importantly, this *global* competition is agnostic to the set of candidate referents on the trial, and all stored associations play a role. *Global* competition is important for stabilizing word-referent pairings in the face of noise in the input. For instance, if A has been strongly associated with a , and a is absent on the current trial, any change in associative strengths between A and other referents will be small because of this previous strong association. *Global* competition explains the results of Experiment 2, in which two referents for the same *double* word are hard to learn

even when they never co-occur. It also explains the decrement in confidence for the second referent of each *double* word shown by participants in Experiment 4.

These two processes produce a highly adaptive word learning system. First, when the system is learning word-object mappings from distributional information that really is produced by one-to-one mappings and occasional noise, the two processes will support each other. *Local* competition will sort out mappings within a trial, and *global* competition will increase the rate at which consistently occurring mappings are formed. However, despite the production of a one-to-one bias at each of the individual stages of competition, the unified system can support the acquisition of two-to-one mappings. This is because the processes are complementary. As seen in Experiment 4, *local* competition can compensate for the action of *global* competition and thus keep the system from failing to learn correct new mappings. Let us now consider several implications of this process-model framework.

5.2. *The importance of local context for statistical learning*

What is striking about the experimental results is the strength of *local* competition relative to *global* competition. In the last two experiments, participants had received all six exposures to each *double* word's early referent before ever encountering its late referent. This should have led to particularly strong *global* competition. Participants' trial-by-trial confidence judgments corroborate this, with early referent judgments being higher than late referent judgments from their earliest occurrences. Nonetheless, participants learned the late referents and did not lose the earlier ones. This highlights the importance of *local* competition in cross-situational word learning, and by extension the importance of understanding what happens in individual learning moments. Evidence from other recent studies of word learning further underscores this point. Onnis, Waterfall, and Edelman (2008) have argued for the importance of local structure in naming events to children in the form of variation sets. Mothers will often highlight the meaning of a single word by using it rapidly in several different sentence constructions, increasing salience through competition. These authors and their colleagues have further argued that this is a general principle that enables learning of structured representations (Goldstein et al., 2010). In a related vein, Perry, Samuelson, Malloy, and Schiffer (2010) have shown that success in word and category learning depends on the local structure of the categories to be learned. When local exemplars vary maximally on dimensions by which they are not organized, *global* learning is accelerated (see also Rost & McMurray, 2009).

Taken together with our competitive framework, these ideas shed light on recent conflicting results in the cross-situational learning paradigm. Several different examinations of many-to-one mapping have yielded mixed results. In some experiments (Ichinco, Frank, & Saxe, 2009), as in our Experiments 1 and 2, participants do not learn multiple mappings well. In other experiments (Kachergis, Yu, & Shiffrin, 2012; Vouloumanos, 2008), as in our Experiment 3, participants succeed in learning multiple mappings. The experiments in this article suggest that these differences in outcomes may lie in the types of competitive processes encouraged by these tasks. For instance, Vouloumanos (2008)

exposed learners to stimuli consisting of one word and one object at a time, with probabilistic co-occurrence relationships expressed only across multiple trials. Thus, there was no opportunity for within-trial *local* competition. In contrast, the present experiments, as well as those described in other explorations of mutual exclusivity in statistical word learning (Ichinco et al., 2009; Kachergis et al., 2012), presented multiple words and objects per trial. The differences in resulting learning rates can be explained by the within-trial and cross-trial structures, and the types of competition they encourage.

5.3. *Representation and process in statistical learning*

In the framework presented in this article, the results are interpreted on the assumption that learners are tracking a simple statistic (e.g., co-occurrence), and the difference in learning results is explained by competitive processes which moderate the stored information. This simple representation is a reasonable starting point, but learners could alternatively represent a different statistical property of their input. For instance, learners could represent a more complex statistic, such as the conditional probabilities between words and objects. Human learners are known to be sensitive to differences in conditional probability in visual (Fiser & Aslin, 2001) and auditory (Aslin, Saffran, & Newport, 1998) statistical learning tasks. However, we believe that such an account based on a complex representation alone would be insufficient on two grounds. First, while tracking conditional probability rather than co-occurrence frequency may be sufficient to explain the results of Experiment 2, it cannot explain the other experiments. Participants showed similar inhibition in Experiment 1 in which conditional probability between *double* words and each of their correct referents was one. It also cannot parsimoniously explain the order effect found in Experiments 3 and 4. Further, one can argue that tracking conditional probability implicitly implements competition at the *global* level. The conditional probability between a word and any given referent is reduced each time that word occurs on that trial without that referent but with other referents. Thus, any complex representation needs to address issues of local and *global* competition to explain the empirical results.

More important, even if a more complex statistic could account for the test results, we believe that it would ultimately be incomplete. The experiments in this article present compelling evidence that cross-situational learning involves important order effects, and that computations about which mapping information to store are made on a trial-to-trial basis. An explanation for the data relying on complex statistical construct would be a move away from a process-level model, and thus have little to say about the trial-by-trial computations. More generally, while questions of process and representation may be treated as orthogonal by normative models (e.g., Frank et al., 2009), they must be intimately connected in process-level models.

5.4. *Mutual exclusivity*

In the phenomenon of disambiguation through mutual exclusivity, children (Markman & Wachtel, 1988; Merriman & Bowman, 1989) or adults (Golinkoff, Hirsh-Pasek, Bailey,

& Wenger, 1992; Halberda, 2006) are first exposed to a consistent mapping between a word and an object. Then, they are shown the same object and a novel object, and asked to find the target of a novel label. Adults and children over 18 months of age overwhelmingly select the novel object as the referent for the novel label. Because such disambiguation is thought to be important to acceleration of word learning, it is considered critical for computational models of word learning (Fazly et al., 2010; Frank et al., 2009; Regier, 2005; Xu & Tenenbaum, 2007; Yu, 2008). Its importance has also inspired a number of competing theoretical accounts: Mutual Exclusivity (Markman, 1990), the principle of Contrast (Clark, 1987), a Pragmatic account (Diesendruck & Markson, 2001), the Novel-Name Nameless-Category principle (Golinkoff et al., 1992), Disjunctive Syllogism (Halberda, 2006), and the size principle (Xu & Tenenbaum, 2007).

While these accounts make different predictions in some cases, we would like to call attention to the point that these accounts posit competition at different levels. Markman's (1990) mutual exclusivity is fundamentally a competitive process at the *global* level: Once a word-object mapping has been established, learners should resist assigning a second label to the same object. It also suggests commitments to trial-by-trial competition. On the other hand, the principle of Contrast and the size principle are competitive processes at the *global* level but do not make a commitment to trial-by-trial competition. Contrast suggests that learners a priori disprefer lexicons in which two labels share the same referential extent, and the size principle suggests that learners a priori prefer lexicons of smaller size. Finally, the pragmatic account and disjunctive syllogism are both processes that operate at the *local* level. On the pragmatic account, a learner expects that the novel label is being produced under Sperber and Wilson's (1986) relevance principle: If the intent of the utterance was to label the familiar object, the speaker would have used the familiar label. Disjunctive syllogism is also about reasoning at the moment of word learning and describes a particular process of elimination used to determine that the novel word should map to the novel label. In contrast to these accounts, the Novel-Name Nameless-Category principle posits that pure novelty, rather than competition with previous mappings, drives the effect.

In the framework presented in this article, we suggest that many of these accounts may not be direct rivals. Rather, they address competition at different levels. Disambiguation through mutual exclusivity may involve multiple levels. Understanding how mutual exclusivity operates at a process level (e.g., Halberda, 2006) will require understanding how these different explanations are implemented and how they interact. One potentially fruitful route for further research may consist in examining patterns of recovery from incorrect use of mutual exclusivity. As evidence suggests that mutual exclusivity is not a hard constraint (Merriman & Bowman, 1989; Regier, 2003), the patterns that adults and children show in learning words that do not accord with one-to-one mapping may be useful in determining at which level competition in mutual exclusivity operates.

6. Conclusion

Cross-situational word learning involves the acquisition of word-object by tracking co-occurrence patterns across multiple instances. In this article, we examined the contribution of competition—both at the *global* and *local* levels—to successful learning. While many models include at least one such process in their explanations (Fazly et al., 2010; Frank et al., 2009; Siskind, 1996; Yu, 2008), prior empirical work on non-one-to-one mappings has been done explicitly in the context of unambiguous learning instances (see, e.g., Liittschwager & Markman, 1994). This article presents a systematic empirical analysis of such learning as it plays out, trial-by-trial, in learning words from a set of ambiguous labeling instances. These experiments show that learners bring competitive mechanisms to the table, and do so at multiple levels. These competitive processes reduce the ambiguity involved in tracking word-object mappings, thereby reducing load for learners with cognitive constraints (Kareev, 1995). These kinds of experiments are critical in the context of recent interest in cross-situational approaches to language acquisition. This new approach shifts the focus of questions about language learning from how children disambiguate reference in single learning events to how disambiguation occurs through integration of information over time (Siskind, 1996; Yu & Smith, 2007). It is critical, however, not to lose sight of the earlier insights about in-the-moment processes. The experiments in this article represent a step in this direction, an attempt to connect the statistical learning approach back to the mechanisms that implement its dynamics and underpin the information acquired from each ambiguous naming event.

Although the experiments in this article demonstrate competitive mechanisms by inhibition of learning, precisely these same mechanisms are involved in successful cross-situational learning. They allow learners to filter much of the noise out of the statistics in their input, narrowing in on correct word-object mappings at a faster rate. Identifying these processes, and understanding their interaction, will be critical to constructing satisfying process-level models of cross-situational learning (Smith et al., 2011; Yu & Smith, 2011).

Of course, learning words is not just a matter of mapping words to their unique referents. Rather, the majority of labels refer to whole categories of referents. To better measure and understand competitive processes, these studies sidestep this problem. But real word learners cannot. There is no reason to think that competition must play out only an object-by-object basis. Instead, it is likely that it occurs also at the level of features of both objects and words (Regier, 2005). Understanding competition at multiple scales, and how these mechanisms can interact, can inform statistical learning theories for other linguistic information. How do *local* and *global* competition play out in learning phonemes, segmenting speech, or learning grammatical categories? Answering these questions will be fundamental to developing process-level models of statistical language learning.

Acknowledgments

This study was supported by a National Science Foundation Graduate Research fellowship to DY as well as National Institute of Health grant R01HD056029 to CY. The authors are grateful to Rich Shiffrin for help developing the guessing analysis, and Krystal Klein for her E-Prime template. In addition, the authors thank Greg Cox, Caitlin Fausey, Terry Regier, Bob McMurray, and one anonymous reviewer for their helpful comments on this manuscript, as well as Justin Halberda for comments on an earlier version.

References

- Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*, 1105–1138.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.
- Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*, 1110–1116.
- Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 458–475.
- Blakemore, C., Carpenter, R. H. S., & Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, *228*, 37–39.
- Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, *34*, 620–642.
- Casenhiser, D. M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, *32*, 319–343.
- Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B*, *362*, 1585–1599.
- Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.
- Desimone, R. (1988). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society*, *353*, 1244–1255.
- Diesendruck, G., & Markson, L. (2001). Childrenal overlap: A pragmatic account. *Developmental Psychology*, *37*, 630–641.
- Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, *27*, 367–392.
- Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*, 203–214.
- Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*, 1017–1063.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.
- Frank, M. C., Goodman, N., & Tenenbaum, J. (2009). Using speakers referential intentions to model early cross-situational word learning. *Psychological Science*, *20*, 578–585.

- Gaskell, M. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*, 105–132.
- Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive science*, *33*, 1087–1116.
- 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.
- Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, *14*, 249–258.
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*, 99–108.
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*, 310–344.
- Horst, J. S., McMurray, B., & Samuelson, L. K. (2006). Online processing is essential for learning: Understanding fast mapping and word learning in a dynamic connectionist architecture. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 339–344). Austin, TX: Cognitive Science Society.
- Ichinco, D., Frank, M. C., & Saxe, R. (2009). Cross-situational word learning respects mutual exclusivity. In N. Taatgen, H. van Rijn, J. Nerbonne & L. Schomaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2214–2219). Austin, TX: Cognitive Science Society.
- Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word-referent mappings. *Psychonomic Bulletin & Review*, *19*, 317–324.
- Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition*, *56*, 263–269.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.
- Liittschwager, J. C., & Markman, E. M. (1994). Sixteen- and 24-month-olds use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology*, *30*, 955–968.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*, 1–36.
- MacWhinney, B. (1989). Competition and lexical categorization. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization* (pp. 195–242). New York: Benjamins.
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77.
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.
- Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience*, *4*, 877–886.
- McClelland, J. L. (2006). How far can you go with Hebbian learning, and when does it lead you astray? In Y. Munakata, M. H. Johnson (Eds.), *Processes of change in the brain and cognitive development: Attention and performance XXI* (pp. 33–69). Oxford: Oxford University Press.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 86.
- McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*, 22–24.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Using your lexicon at two timescales: Investigating the interplay of word learning and recognition. *Psychological Review*, *119*, 831–877.

- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*, 9014–9019.
- Merriman, W., & Bowman, L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*.
- Merriman, W. E. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (Ed.), *The emergence of language* (pp. 331–358). Hillsdale, NJ: Erlbaum.
- Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1209–1228.
- Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, *109*, 423–430.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532–552.
- Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*, 461–472.
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*, 1894–1902.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*, 909–957.
- Regier, T. (2003). Emergent constraints on word-learning: A computational perspective. *Trends in Cognitive Sciences*, *7*, 263–268.
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*, 819–865.
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, *28*, 89–104.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 339–349.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- 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., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually-cued attention and early word learning. *Cognitive Science*, *34*, 1287–1314.
- Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, *27*, 720–731.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
- Spivey, M. J., Dale, R., Knoblich, G., & Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? A reply to van der Wel et al. (2009). *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 251–254.
- 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.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.

- Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*, 729–742.
- Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*, 288–297.
- Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language, Learning, and Development*, *4*, 31–62.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420.
- Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*, 165–180.
- Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, *119*, 21–39.

Appendix

Because tests for different types of words (*single* and *double*) had different chance levels of performance due to guessing, comparing absolute accuracy could be misleading. To correct for this, we developed an analysis that breaks test accuracy down into two independent portions: the number of successes due to known word-object mappings (K), and the number of successes due to chance. The number of known word-object mappings can then be compared directly across word types.

Responses due to knowledge are modeled as a deterministic process: if a participant knows the correct answer, he or she always gives it. Guesses are modeled as a draw from a binomial distribution with parameter p set by the number of alternatives from which the participant selects. Each participant produces some number of correct responses n . For any number of correct responses due to knowledge ($k \in 0 \dots n$), the remaining correct responses ($n - k$) must be due to guessing. The number of observed correct responses has some likelihood under each of the different settings of k . Our analysis finds the setting $K = k$ under which the data have the highest likelihood.

For *single* words, accuracy is determined from a single choice among four alternatives. Thus, the binomial distribution for guessing for *single* words has $p = \frac{1}{4}$. For a participant who produces n correct responses out of 6, the maximum likelihood estimate of known mappings K_s is:

$$\begin{aligned}
 K_s &= \operatorname{argmax}_{k_s} P(N = n | K_s = k_s) \\
 &= \operatorname{argmax}_{k_s} \operatorname{binomial}(n - k_s, 6 - k_s, \frac{1}{4})
 \end{aligned}
 \tag{A1}$$

For *double* words, a participant makes two consecutive responses. Zero, one, or both of these responses may be correct. A participant may also know zero, one (K_{d1}), or both (K_{d2}) of the correct mappings for each word. As two of the four alternatives are correct responses, $p = \frac{1}{2}$ for the first response. If the first response is correct, 1 of the remaining 3 choices is

correct. Thus, given a correct answer on the first choice, $p = \frac{1}{3}$ for the second. For a participant who selects the correct first response for n of the six words, and both correct responses for m of the six words, the maximum likelihood estimate for K_{d1} and K_{d2} is:

$$\begin{aligned}
 K_{d1}, K_{d2} &= \operatorname{argmax}_{k_{d1}, k_{d2}} P(N = n, M = m | K_{d1} = k_{d1}, K_{d2} = k_{d2}) \\
 &= \operatorname{argmax}_{k_{d1}, k_{d2}} P(N = n | K_{d1} = k_{d1}, K_{d2} = k_{d2}) \\
 &\quad \times P(M = m | K_{d1} = k_{d1}, K_{d2} = k_{d2}, N = n) \\
 &= \operatorname{argmax}_{k_{d1}, k_{d2}} \operatorname{binomial}\left(n - k_{d1} - k_{d2}, 6 - k_{d1} - k_{d2}, \frac{1}{2}\right) \\
 &\quad \times \operatorname{binomial}\left(m - k_{d2}, n - k_{d2}, \frac{1}{3}\right)
 \end{aligned} \tag{A2}$$