Journal of Memory and Language 99 (2018) 62–73

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The company objects keep: Linking referents together during cross-situational word learning

Martin Zettersten [a,*], Erica Wojcik [b], Viridiana L. Benitez [c], Jenny Saffran [a]

[a] University of Wisconsin-Madison, 1202 W Johnson Street, Madison, WI 53706, United States
[b] Skidmore College, 815 North Broadway Street, Saratoga Springs, NY 12866, United States
[c] Arizona State University, 950 S. McAllister Ave, Tempe, AZ 85287, United States

ARTICLE INFO

Article history:
Received 25 May 2017
Revision received 20 October 2017

Keywords:
Cross-situational word learning
Semantic networks
Language
Memory

ABSTRACT

Learning the meanings of words involves not only linking individual words to referents but also building a network of connections among entities in the world, concepts, and words. Previous studies reveal that infants and adults track the statistical co-occurrence of labels and objects across multiple ambiguous training instances to learn words. However, it is less clear whether, given distributional or attentional cues, learners also encode associations among the novel objects. We investigated the consequences of two types of cues that highlighted object-object links in a cross-situational word learning task: distributional structure – how frequently the referents of novel words occurred together – and visual context – whether the referents were seen on matching backgrounds. Across three experiments, we found that in addition to learning novel words, adults formed connections between frequently co-occurring objects. These findings indicate that learners exploit statistical regularities to form multiple types of associations during word learning.

Introduction

One of the central problems faced by observers attempting to learn the words of a novel language is referential ambiguity (Quine, 1960). When a learner hears a novel word, it is likely that a host of candidate referents will be available in the visual environment. Most investigations focused on this problem ask how learners eliminate competing referents to successfully map a label to a single referent (Medina, Snedeker, Trueswell, & Gleitman, 2011; Smith & Yu, 2008; Trueswell, Medina, Hafri, & Gleitman, 2013; Yu & Smith, 2007; Yurovsky, Fricker, Yu, & Smith, 2014). However, learning a word involves more than forming a mapping between a label and an isolated entity. Learners also encode expectations about the types of objects with which a referent is likely to co-occur, and where the word is likely to be encountered (Boyce, Pollatsek, & Rayner, 1989; Miller, 1999; Roy, Frank, Decamp, Miller, & Roy, 2015; Samuelson, Smith, Perry, & Spencer, 2011; Smith, Suanda, & Yu, 2014). While referential ambiguity presents a hurdle for learning label-object mappings, it also provides an opportunity to learn useful information about the contextual structure of the environment, such as which objects are related to one another.

Learners have many strategies at their disposal for solving the problem of referential ambiguity (Akhtar & Tomasello, 2000; Baldwin, 1993; Markman & Wachtel, 1988; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Smith & Thelen, 2003). One proposed strategy entails cross-situational word learning (Smith & Yu, 2008; Yu & Smith, 2007). Although any single encounter with the word "tomato" may be referentially ambiguous, only one consistently occurring entity will emerge as the word's most likely referent across multiple encounters with "tomato" (i.e., a round, squishy, and savory fruit). There is substantial evidence from cross-situational word learning tasks that both infants (Smith & Yu, 2008; Vlach & Johnson, 2013; Vouloumanos & Werker, 2009; Yu & Smith, 2011) and adults (Yu & Smith, 2007; Yurovsky, Yu, & Smith, 2013; Yurovsky et al., 2014) can successfully map labels to objects across multiple ambiguous training instances by using label-referent co-occurrence statistics.

Notably, most cross-situational word learning studies – with a few exceptions (Dautriche & Chemla, 2014; Chen & Yu, 2017; Kachergis, Yu, & Shiffrin, 2009; Roembke & McMurray, 2016) – lack contextual structure. Associations between label-object pairs are the only reliable patterns; relationships among other elements, e.g. between the objects themselves, are intentionally minimized. In natural learning environments, however, any individual instance of a label is immersed in rich contextual information, such as related nouns and verbs, related objects in the environment, or

* Corresponding author.
E-mail address: zettersten@wisc.edu (M. Zettersten).

https://doi.org/10.1016/j.jml.2017.11.001
0749-596X/

visual scenes that connect word utterances across encounters (Hills, 2013; Hills, Maouene, Riordan, & Smith, 2010; Roy et al., 2015). These kinds of contextual structure highlight connections between entities in the environment, such as the associations among objects. Forming connections between related objects, such as tomatoes and lettuce, is crucial to building semantic knowledge (Landauer & Dumais, 1997; Sadeghi, McClelland, & Hoffman, 2015). In the current study, we focus specifically on whether and how learners encode associations between objects from the structure implicit in cross-situational word learning tasks.

Multiple sources of information may lead learners to form object-object associations. One type of cue is the regularity with which objects co-occur in the world (Bar, 2004). Objects are not randomly distributed in the environment, but instead occur in schema-based clusters. When learners hear the word "tomato", they are more likely to be in the presence of some items (e.g., lettuce, onions, and cucumbers) than others (e.g., soccer balls, cleats, and socks). In this sense, the distribution of object co-occurrences in a learner's environment is *skewed* rather than *uniform*: some objects are more likely to occur in each other's company than others.

Another contextual cue that can help learners link objects together is the presence of a visual context shared across similar locations or scenes (Oliva & Torralba, 2007). A similar visual context can link individual objects that are spatially or temporally distant. For example, while tomatoes and lettuce are objects that are sometimes seen together, they also are often seen within the same prototypical visual context: e.g., they both may often appear on a kitchen counter.[1] Shared visual context may aid in linking objects by creating contextual expectations and by guiding attention towards objects in similar contexts: Regularities between the occurrence of objects (such as tomatoes and lettuce) and visual contexts (e.g., the kitchen counter) influences object recognition, such that specific objects come to be linked to specific visual environments (Brockmole, Castelhano, & Henderson, 2006; Oliva & Torralba, 2007; Vlach & Sandhofer, 2011). The kitchen counter therefore begins to activate expectations for both tomatoes and lettuce, and may act as a cue to link objects across different encounters. Simultaneously, shared visual context may also guide attention to objects occurring within a similar visual context. For example, noticing tomatoes on one end of the kitchen counter and lettuce on the other may lead a learner to recognize a relation between the two. Both the co-occurrence of objects and shared visual context are features of the word learning environment that may influence learners' ability to track meaningful links between objects (e.g., Roy et al., 2015; Vlach & Sandhofer, 2011).

Does tracking object-object links help or hurt word learning? On one hand, the fact that tomatoes and lettuce often occur together and frequently share a similar visual context may make the task of word learning even more difficult: the referent for "tomato" may be harder to disambiguate, particularly from the referent for "lettuce", since the two objects frequently co-occur in the presence of each label. A more uniform distribution of potential referents, and more distinctive or variable visual contexts, by contrast, may help the target referent emerge as the most consistent signal across multiple noisy contexts. On the other hand, learning words involves not just learning label-object mappings, but also forming expectations about the contexts in which words occur (Miller, 1999; Saji et al., 2011). From this perspective, it may be useful for a learner to notice regularities beyond a single label-object mapping. Each labeling event is also an opportunity to learn about the company objects (and their labels) keep.

Previous research suggests that adults use information about the relationships between objects to map objects to novel labels. Specifically, learners can use object-object relations to disambiguate the kinds of objects a label might refer to, such as the fact that a label occurred with animal exemplars rather than items from another category (Dautriche & Chemla, 2014). Other studies have shown that skewed distributions in the frequency with which objects co-occur, as well as thematic groupings among co-occurring objects, can influence how adults learn novel label-object mappings (Chen & Yu, 2017; Roembke & McMurray, 2016; Kachergis et al., 2009). However, word learning moments provide opportunities to not only learn about label-object mappings, but also to learn about the relation between entities occurring in the same context. Exploring the set of candidate referents for a word may lead learners to extract contextual regularities, such as which objects often go together. Furthermore, how learners track these additional regularities may affect how they track the label-object mappings. In the current studies, we assessed adults' ability to form associations between objects in addition to learning the referents of novel words, in the absence of explicit instruction to do so.

In each of the following studies, we asked what adults learn about novel object-object associations as they are engaged in cross-situational word learning and how learning these object-object associations affects word learning. On each trial, learners were presented with one word and four novel objects and were asked to pick the object to which the word referred. The correct word-referent pairings were ambiguous within individual trials, but were disambiguated when word-referent pairs were aggregated across trials, as in the typical cross-situational word-learning task design. No feedback was provided during the training trials. During the test trials, we assessed learning of the relationships between objects.

We exposed learners to object-object links during the learning phase in two ways: by manipulating how frequently specific objects occurred together, and by providing a visual context cue that was identical for pairs of objects. We manipulated the distribution of object co-occurrences by creating two types of object co-occurrence distributions: A *uniform* distribution, that is, a condition in which objects co-occurred equally with each other (as in traditional cross-situational word learning studies), and a *skewed* distribution, where each object occurred more frequently with one particular object than with other objects. We manipulated the presence of a visual context cue, a unique background that was identical for some objects but not others, to highlight the links between objects.

In Experiment 1, we tested whether adults could learn object-object structure when both co-occurrences and shared visual context cues highlighted these links. Adults were presented with a skewed distribution in which pairs of objects occurred frequently together and shared identical background images (skewed distribution and visual cue). We assessed adults' learning of both word-object mappings and object-object connections. Experiment 1 was designed to provide a first measure of whether adults can track both word-object links and object-object links when they are highlighted by distributional and visual context cues. In the subsequent experiments, we assessed the distinct contributions of co-occurrence and visual context cues to encoding contextual structure. In Experiment 2, we asked whether the visual context cue alone (uniform distribution and visual cue) was sufficient for participants to learn object-object connections. In Experiment 3, we asked whether the object co-occurrences alone (skewed distribution and no visual cue) were sufficient for participants to learn object-object connections. In all experiments, we assessed whether encoding contextual structure affected word learning. If tracking

[1] Besides providing a visual cue, kitchen counters also evoke a host of semantic and thematic information for a learner that are relevant to learning associations between objects and label-object associations (see Chen & Yu, 2017). Our goal in the current experiments was to isolate the role of visual cues in the absence of semantic information.

object-object links hurts word learning, then we should see trade-offs between encoding object-object links and learning words.

Experiment 1

Experiment 1 was designed to determine whether adults learn object-object pairings in a cross-situational word learning task. Pairs of objects were linked using two types of correlated structure: object-object co-occurrences and visual context cues. Paired objects occurred more frequently together than with other objects (skewed distribution) and shared the same background image (visual context cue). During training, learners were presented with trials where they were asked to choose a referent for a novel label from a set of objects. We measured whether learners correctly linked labels to their referents across training trials. After training, we tested whether learners formed associations between contextually linked objects by asking learners to judge which objects "went together".

We predicted that participants would both successfully map novel labels to the target objects and encode the connection between objects that were linked by the contextual cues during training. Moreover, we hypothesized that there would be a trade-off between learning novel words and encoding object-object associations. Since noticing the object-object relationship may in part result from evaluating the full array of objects on a given trial, leading to less robust label-target associations, we predicted that participants who successfully encoded object-object associations would be worse at mapping the novel label to its target object.

Method

Participants

We recruited 51 participants (22 female; all native speakers of English) through Amazon Mechanical Turk (mean age: 32.8 years, range: 20–69 years). Participants were paid $0.80 and completed the study in an average of 7.0 min ($SD = 2.2$).

Stimuli

The stimuli consisted of eight novel objects (see Fig. 1A) and eight labels (*blick*, *dirg*, *geddle*, *jellup*, *labo*, *manu*, *stip*, *vima*) used in past cross-situational word learning studies (Yu & Smith, 2007). We created two word-object lists by randomly pairing each word with one of the objects. Participants were randomly assigned to one of these pairing lists. Additionally, we used four unique backgrounds (from Vlach & Sandhofer, 2011), and photoshopped each object onto a single background to create the object + visual context images. Importantly, each object had the same background as one other object (see Fig. 1B). Object/background pairings remained consistent across the training trials. During test trials, no backgrounds or surrounding borders were presented (see Fig. 1A).

Design & procedure

Training phase. During training, participants completed 48 cross-situational word learning trials presented in random order. On each trial, participants saw four objects along with the visual prompt "There's a X here. Which one of these is a X?" (see Fig. 2). The participant selected one of the objects to move on to the next trial. No feedback was provided. The specific location of the four objects on the screen was randomized on each trial, so that learners could not link objects based on specific locations or the proximity of objects. Participants could learn which object went with each label by observing the regularity with which the novel objects occurred with a label across trials. Each of the eight labels occurred six times and was always in the presence of one target object, creating a learnable label-object pair. Participants' choices on each trial provided a measure of how well they were learning word-object mappings; there was no separate test measuring word learning.

Manipulating object co-occurrences. We manipulated the regularity with which objects occurred together during the training phase. Across the 48 training trials, each individual object appeared 24 times total – six times as a target and 18 times as a distracter (since there were 192 total object occurrences – 48 trials multiplied by four objects per trial – distributed evenly across eight objects). If object-object co-occurrences were approximately uniform, as in past studies, this would result in object-object pairs occurring together 43% of the time (roughly 10 of 24 trials) on average (see also Experiment 2). Instead, objects in the current study were paired such that each object occurred more frequently with one other object across training trials (see Fig. 3; *skewed* distribution). We manipulated co-occurrence strength such that two object-object pairs occurred together 75% of the time (18 of 24 trials; *high frequency* pairs), while two other object-object pairs occurred together 58% of the time (14 of 24 trials; *moderate frequency* pairs). On trials for which an object was the target object (6 total), its paired object occurred as the distracter on 4 of those trials, in order to ensure that each label was unambiguously associated with only one object across training trials.

Visual context cue. In addition to manipulating object co-occurrences, we also introduced a visual cue to connect objects by giving object pairs identical backgrounds (see Fig. 1B). The objects that occurred frequently together always had the same background image. Thus, the co-occurrence structure and the visual context cue were redundant cues that linked the same object pairs together.

Test phase. Since word learning was tested during the training trials, the test phase was entirely focused on whether participants learned which objects were linked together during the training phase. On each trial, we presented one of the eight objects from training as a target, and three different objects from training as response options. Participants were then asked to choose which object from among the three response options "went with" the target object (see Fig. 2, Test). All objects were presented without the visual context cue, i.e. without a background (as in Fig. 1A), so that the test trials themselves did not contain any cues connecting the objects. Each object was tested once, for a total of eight test trials. The distracter images were chosen such that all objects occurred equally frequently across the eight test trials. The objects were presented in one of four random orders to which participants were randomly assigned.

Results

Word learning

We predicted that participants would learn the label-object pairings, as demonstrated by their overall accuracy on the training trials and by an increase in accuracy with each new occurrence of a particular label. We fit a logistic mixed-effects model predicting participants' trial-by-trial accuracy from the trial number for each label (1–6; centered) (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). We used the lme4 package version 1.1–13 in R to fit all models (Bates & Maechler, 2009).[2] We included by-subject and by-item random intercepts and random slopes for trial number to

[2] All experimental data along with an analysis script and an overview over the analyses in the manuscript are openly available at https://github.com/mzettersten/crossSit.

Fig. 1. (A) unfamiliar object stimuli used in all experiments. (B) Unfamiliar object stimuli with backgrounds (visual context cues) used in Experiment 1 & Experiment 2.

fit a model with the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). In order to test participants' overall performance compared to chance, we applied an offset corresponding to the logit of chance performance (i.e., .25, the probability of being correct on a trial) to the intercept of the model.

Overall, participants chose the correct object on 39.7% of training trials, more frequently than the expected probability of 25% if participants were choosing objects at random ($b = .65$, 95% Wald CI = [.38, .91], $z = 4.82$, $p < .001$). Crucially, participants' accuracy improved as training trial number for a given target label increased, $b = .10$, 95% CI = [.03, .18], $z = 2.61$, $p = .009$ (see Fig. 4A, Skewed condition).

Object-object association memory

We hypothesized that in addition to learning the label-object pairings, participants would recall which objects had occurred together frequently and shared a visual context, as measured by the test phase. We predicted participants' accuracy on a given test trial in a logistic mixed-effects model with an intercept, applying an offset corresponding to the logit of chance performance (.33). The model included a by-subject intercept and a random intercept for item pairs (i.e., trials testing the same object-object pair were treated as the same item to account for the non-independence between paired objects).[3]

Participants' probability of choosing the high co-occurring object (M = 42.6%) reliably differed from chance, $b = .39$, 95% CI = [.19, .60], $z = 3.74$, $p < .001$ (see Fig. 5A, Skewed Condition), suggesting that participants successfully encoded which objects were paired together during training.

Additionally, we explored whether memory for object-object associations varied between high frequency and moderate frequency pairs by adding Pair Type (Moderate vs. High) as a fixed effect in the model, as well as a by-subject random slope for Pair Type. Memory was similar between the two types of co-occurring pairs ($p = .65$), thus we collapsed across object-object co-occurrence pair type in subsequent analyses.

Relationship between word learning & object-object association memory

Next, we asked whether contextual structure led to tradeoffs between learning words and forming object-object associations: Did participants who learned label-object mappings better perform worse at identifying associated object pairs? More difficulty in learning the label-object mappings might lead participants to compare objects more frequently in searching for possible referents, increasing their likelihood of learning object-object connections. To test this question, we investigated whether remembering the corresponding object pair for a given object item (object-object association accuracy) was related to differences in learning that object item's corresponding label (word learning accuracy). We fit a logistic mixed-effects model in which we regressed participants' trial-by-trial accuracy in choosing the target referent during word learning on their object-object association accuracy for the target item (correct or incorrect; centered). We included by-subject and by-item random intercepts, as well as by-subject and by-item random slopes for object-object association memory accuracy.

[3] Results remain qualitatively the same when including a by-item random effect grouped by individual items/objects (rather than object/item pairs).

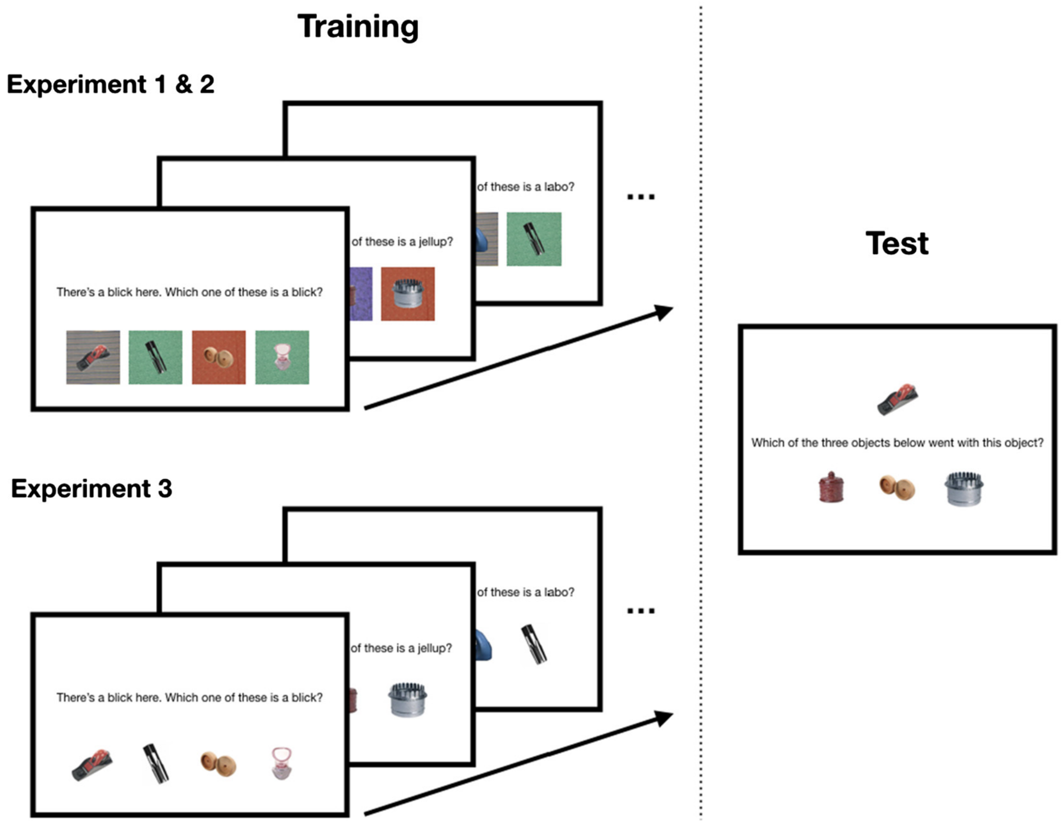

Fig. 2. Trial design for the training and test phase of Experiments 1–3.

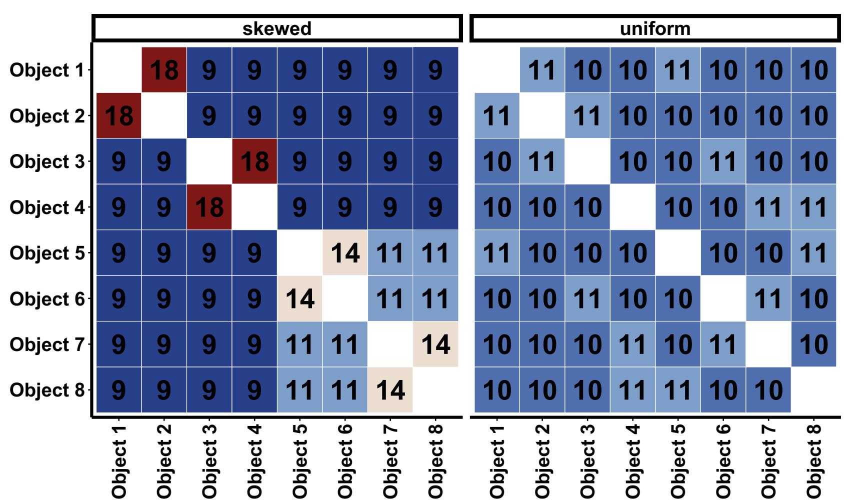

Fig. 3. Object-object co-occurrence matrices for the Skewed and the Uniform Object Co-Occurrence Condition. Color indicates the frequency of co-occurrence along a continuous scale from dark blue (less frequent co-occurrences) to dark red (more frequent co-occurrences). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Participants who correctly chose the high co-occuring object for a given item were *less* accurate on word learning trials for that item, $b = -.31$, 95% CI = [−.60, −.02], $z = -2.12$, $p = .03$ (see Fig. 6A, Skewed Condition), consistent with the hypothesis of a tradeoff between word learning accuracy and object-object association accuracy.

Discussion

In Experiment 1, we demonstrated that participants tracked regularities beyond label-object mappings during cross-situational word learning. In addition to learning novel mappings between unfamiliar objects and labels, participants successfully encoded which objects occurred together frequently, despite no prompting to track this information. Even in a brief, 7-min training phase, participants extracted information about links between objects based on contextual structure (a skewed co-occurrence distribution and a visual context cue) while engaged in a word learning task.

While participants learned both label-object mappings and object-object associations, we also observed an item-level tradeoff between participants' performance on these two tasks. More

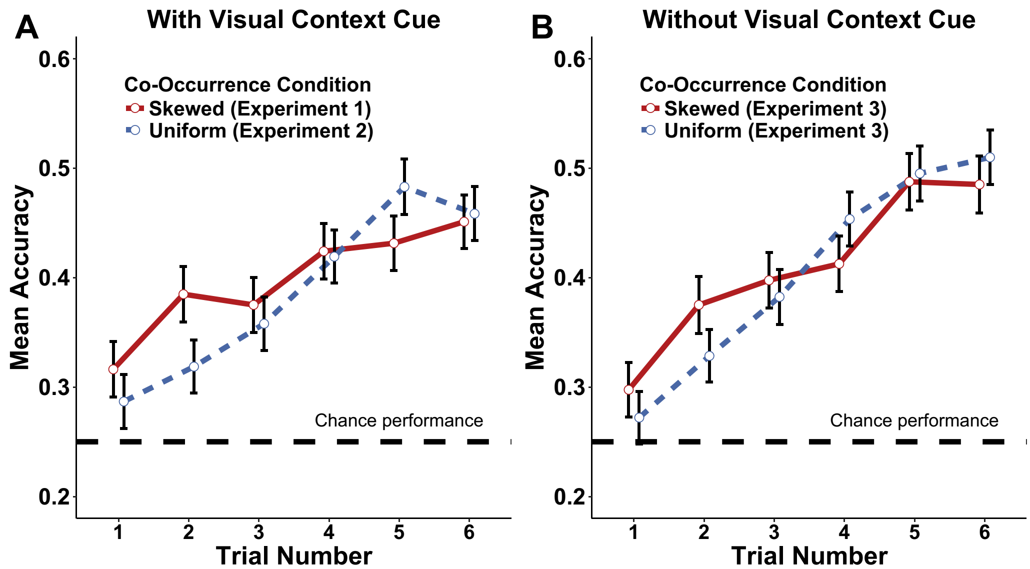

Fig. 4. Word Learning Accuracy across training trials 1–6 for a given label-object pairing: (A) in Experiments 1 and 2 (with visual context cue) and (B) in Experiment 3 (without visual context cue). Error bars represent within-subject SEs (Morey, 2008). Note: Since trial order was fully randomized, participants had sometimes seen several trials by the time they responded to a given item for the first time, meaning they have information to aid in constraining their object choices. We therefore did not expect participants to be completely at chance for the first trial of a given item overall.

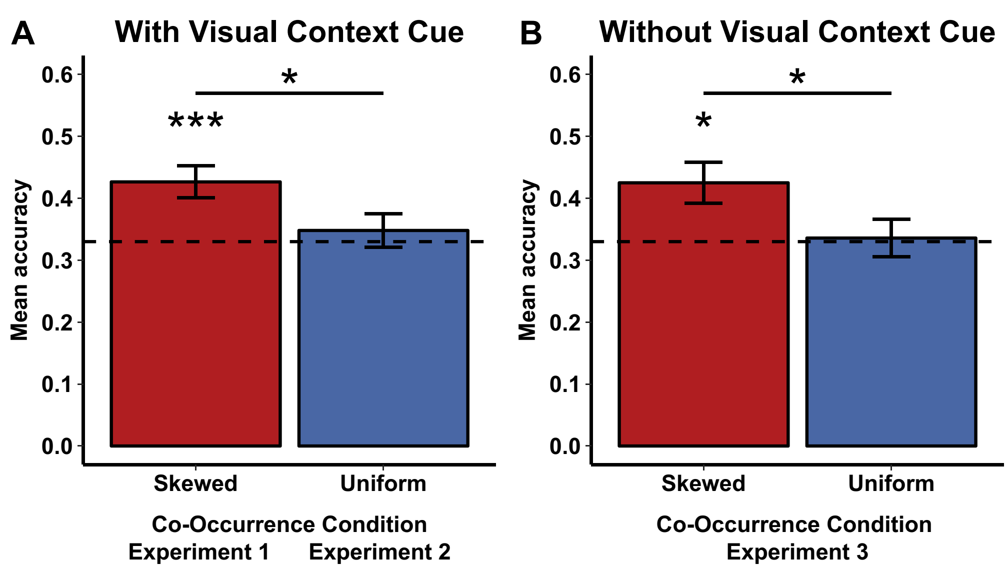

Fig. 5. Object-object association accuracy (A) in Experiment 1 and 2 (with visual context cue) and (B) in Experiment 3 (without visual context cue). Error bars represent +1/−1 SEs. Dashed lines represent chance performance. $^{***}p < .001$; $^{*}p < .05$.

accurately connecting a label with its target object was associated with worse memory of the object that occurred frequently with the target object. This finding suggests that while the contextual structure provided additional information about the novel objects during training, detection of contextual structure was associated with lower accuracy in learning label-object mappings. One possible explanation is that the likelihood of participants encoding object-object associations increased when they had more difficulty in disambiguating which object was the correct referent. That is, the more trials participants needed to identify the target referent for a given label, the more likely they were to attend to a broader array of objects across training trials, which increased the likelihood of encoding associations between frequently co-occurring objects. We will consider possible explanations of this result in the discussion of Experiment 2.

Experiment 1 provides an initial demonstration that contextual structure in the form of object co-occurrences and shared visual background allows participants to form connections between objects while learning novel words. However, it is unclear how these two types of cues (distributional and visual) contributed to participants' learning. One possibility is that participants learn object-object links based on either kind of contextual structure (co-occurrence distribution or visual context cues on their own). Another possibility is that both contextual structures work together to allow participants to form object-object links by providing correlated information (see Morgan, Meier, & Newport, 1987). Finally, just one of these types of structure may be driving participants' object-object connection memory. The goal of Experiments 2 & 3 was to disentangle these different possibilities, while simultaneously investigating the consequences of different kinds

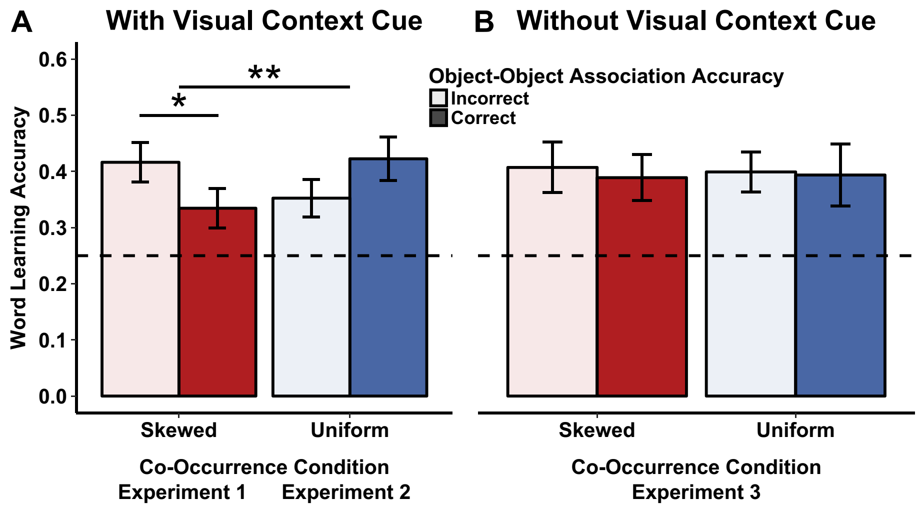

Fig. 6. Item-specific relationship between memory and word learning (A) in Experiment 1 and 2 (with visual context cue) and (B) in Experiment 3 (without visual context cue). Bars represent word learning accuracy for items depending on whether participants were correct (opaque bars) or incorrect (transparent bars) on object-object association memory trials for that item. Error bars represent +1/−1 SEs from the point estimates of the model. Dashed lines represent chance performance. $^{**}p < .01$; $^{*}p < .05$.

of contextual structure on the relationship between object-object association memory and word learning. As a first step, we isolated the effect of visual context cues in Experiment 2, in the absence of object-object co-occurrence structure.

Experiment 2

In Experiment 2, we asked whether participants connect object pairs linked solely by visual context cues, with all objects co-occurring equally frequently. This study was identical to Experiment 1 except that the object co-occurrence structure was *uniform* rather than *skewed*, such that objects occurred together approximately equally frequently (as in traditional cross-situational word learning paradigms; see Fig. 3, *uniform* distribution). Objects were paired only by a visual cue: each object occurred in the same visual context as one other object but, unlike Experiment 1, the objects with the same visual cue (i.e. the same background) did not co-occur more frequently together across the word learning phase.

If learners can link objects during cross-situational word learning based on shared visual context cues alone, we should observe similar accuracy on object-object memory as in Experiment 1. However, if frequent object co-occurrence is crucial to forming connections between objects, we should see reduced memory for object-object associations.

Experiment 2 also allowed us to investigate the source of the tradeoff between learning words and learning connections between objects that we observed in Experiment 1. If this tradeoff can be driven by the presence of either distributional or visual contextual structure, then we should see the same pattern of results that we observed in Experiment 1. However, there may be differences in how these two types of cues generate tradeoffs. In particular, while distributional structure draws learners' attention to paired objects within a trial (potentially leading to worse label-referent disambiguation), forming object-object connections using visual context cues requires learners to track information about individual objects – and their specific contexts – across trials. Without distributional structure heightening connections within trials, learners who are more attentive to individual objects and their background contexts may perform better at word learning *and* at linking objects based on their specific visual context cue. Linking objects based on their context cues requires attending to regularities across trials (rather than within), a strategy that could simultaneously aid word learning, removing the tradeoff observed in Experiment 1. Therefore, in addition to assessing word-object and object-object links individually, we analyzed the relationship between learning words and object-object links.

Method

Participants

We recruited 51 new participants (17 female; all native speakers of English) through Amazon Mechanical Turk (mean age: 35.9 years, range: 19–72 years). Participants were paid \$0.80 and completed the study in an average of 6.9 min ($SD = 2.6$).

Design & procedure

The design was identical to Experiment 1, except for the change in object co-occurrence frequency (see below).

Manipulating object co-occurrences. All objects co-occurred as equally as possible across the learning trials (see Fig. 3; *uniform* object co-occurrences), as in a typical cross-situational word learning task. Each individual object appeared on 24 trials during the word learning phase, and occurred with three other objects on each of these trials. Our goal was to distribute $3 * 24 = 72$ possible occurrences across seven other objects as evenly as possible for each object. Each object occurred at least 10 times with every other object, with two objects occurring 11 times with a given object respectively, creating a near-uniform distribution (see the uniform distribution matrix in Fig. 3). We also balanced how frequently a given object occurred as the distracter with a given target object, such that every object occurred two to three times as a distracter across the six learning trials for a given target object.

Visual context cue. Similar to Experiment 1, we maintained the visual context cue by giving object pairs identical backgrounds. The paired objects in Experiment 2 were associated only through their shared visual context. The eight objects were randomly grouped into four object pairs, with each object in a given pair sharing the same unique background (see Fig. 1B and Fig. 2-Training). For instance, object 1 and object 2 shared the same visual background. However, unlike Experiment 1, these two objects were not more likely to occur together on the same trial than other objects. The fact that paired objects shared the same visual background context during training was therefore the only cue available to participants to aid them in choosing which object "went with" the target object at test.

Results

Word learning

We fit the logistic mixed-effects model used in Experiment 1 to the training trial data. Participants chose the correct object on 38.7% of trials, more frequently than would be expected by chance ($b = .60$, 95% Wald CI = [.30, .90], $z = 3.94$, $p < .001$). Moreover, participants' accuracy improved with each trial for a given target label, $b = .19$, 95% CI = [.12, .25], $z = 5.60$, $p < .001$ (see Fig. 4A, Uniform Condition).

In order to investigate how co-occurrence structure (uniform vs. skewed) affected learning, we conducted an additional analysis comparing performance in Experiments 1 and 2 (see Fig. 4A). We combined the data from the two experiments and fit a logistic mixed-effects model predicting participants' trial-by-trial accuracy from the trial number for each label (1–6; centered), the co-occurrence type (uniform in Experiment 2 vs. skewed in Experiment 1; centered, −.5, +.5), and their interaction. We included a by-subject random intercept and random slope for trial number, as well as a by-item random intercept and by-item random slope for the interaction term (Barr, 2013).[4] There was a marginal interaction between co-occurrence type and trial number, $b = -.08$, 95% CI = [−.17, .01], $z = -1.73$, $p = .08$. This interaction indicates that the learning curve for uniform co-occurrence structure (Experiment 2) was marginally steeper than for the skewed co-occurrence structure (Experiment 1). However, participants' overall performance did not differ between the skewed (Experiment 1) and the uniform (Experiment 2) co-occurrence structure, $b = .04$, 95% CI = [−.26, .34], $z = .26$, $p = .79$.

Object-object association memory

Unlike Experiment 1, participants were at chance in identifying object pairs (objects that had the same background) from the training phase (Fig. 5A, Uniform condition; M = 34.8%), $b = .04$, 95% CI = [−.20, .28], $z = .33$, $p = .75$.

To investigate whether co-occurrence structure affected if participants tracked which objects occurred together, we compared participants' performance on object-object association tests in Experiments 1 and 2 (Fig. 5A). We regressed participants' trial-by-trial accuracy on co-occurrence type (skewed distribution in Experiment 1 vs. uniform distribution in Experiment 2) in a logistic mixed-effects model with by-subject and by-item-pair random intercepts, as well as a by-item-pair random slope for co-occurrence type. Participants were more likely to choose the target object in Experiment 1 (skewed co-occurrence structure) than in Experiment 2 (uniform co-occurrence structure), $b = .34$, 95% CI = [.03, .65], $z = 2.13$, $p = .03$.

Relationship between word learning & object-object association memory

As in Experiment 1, we investigated whether remembering the object paired with a given item (object-object association accuracy) was related to differences in learning that item's label (word learning accuracy). In contrast to Experiment 1, we found that participants in Experiment 2 who correctly chose the paired object were marginally more accurate on word learning trials for that item, $b = .27$, 95% CI = [−.02, .56], $z = 1.85$, $p = .06$ (Fig. 6A, Uniform Condition).

To further assess differences in the relationship between word learning and forming object-object associations across Experiment 1 and Experiment 2, we compared performance across both experiments using a logistic mixed-effects model in which we regressed participants' trial-by-trial accuracy in choosing the target referent during word learning on their object-object association memory accuracy for the target item, co-occurrence type (uniform vs. skewed), and their interaction (Fig. 6A). We included by-subject and by-item random intercepts, a by-subject random slope for object-object association memory accuracy, and a by-item random slope for the interaction between memory accuracy and co-occurrence type.[5] There was a significant interaction between memory accuracy and co-occurrence type, $b = -.65$, 95% CI = [−1.04, −.25], $z = -3.22$, $p = .001$. In Experiment 1, in which object pairs were cued both by skewed co-occurrence structure and a visual context cue, better object-object association memory was associated with worse word learning. In Experiment 2, in the absence of skewed co-occurrence structure, this pattern was reversed.

Discussion

Experiment 2 clarified the degree to which the two types of contextual structure introduced in Experiment 1 contributed to participants' encoding of object-object associations. Removing the skewed co-occurrence structure from Experiment 1 and including only the visual cue led to worse object-object association accuracy. In fact, participants showed no evidence of learning connections between objects from visual context cues alone, in the absence of co-occurrence structure.

Word learning performance was similar between skewed (Experiment 1) and uniform (Experiment 2) co-occurrence structure, with participants reaching roughly the same level of accuracy by the end of the experiment. This in itself is notable, since participants encountering skewed co-occurrences had a somewhat more difficult task: frequently co-occurring objects in Experiment 1 appeared as a distracter more frequently (4 out of 6 occurrences) than any object in the uniform co-occurrence structure in Experiment 2 (2–3 occurrences), making the correct label-object mapping more ambiguous. However, there was a marginal interaction between trial number and co-occurrence type, suggesting that the contextual structure provided by skewed co-occurrence structure allowed participants to be more accurate at the beginning of the learning phase, perhaps because participants could narrow their choices more quickly. While this interaction should be interpreted with caution, it is consistent with previous findings suggesting an initial boost in accuracy provided by contextual structure (Dautriche & Chemla, 2014).

The co-occurrence structure also influenced the relationship between word learning and object-object memory performance. When object co-occurrences were skewed in Experiment 1, better word learning was associated with worse memory for which object co-occurred frequently with the target object. However, when object co-occurrences were uniform, better word learning was associated with marginally better object-object memory. These results suggest that the two types of contextual structure (distributional cues and visual context cues) have different effects on how word learning is related to object-object learning. Given that the absence of object co-occurrence structure reversed the negative

[4] We originally fit the model with the maximal random effects structure, including by-item random slopes for trial number and co-occurrence type. However, when this model did not converge, we followed the recommendations of Barr (2013), removing the random slopes for lower-order effects but retaining the random slope for the highest-order term to maintain acceptable Type I error rates.

[5] Following the recommendations of Barr (2013), we simplified the maximal random effects structure when the model did not converge, removing the by-item random slopes for memory accuracy and co-occurrence type but retaining the random slope for the highest-order term – corresponding to the main fixed effect of interest (see also footnote 4).

correlation between word learning and object association memory from Experiment 1, one possible interpretation of these results is that the presence of object-co-occurrence structure drove the tradeoffs observed in Experiment 1.

Taken together, Experiments 1 and 2 suggest that skewed co-occurrence structure facilitates learning object-object connections when those objects are also marked by a visual context cue. However, it remains unknown whether skewed co-occurrence structure alone is sufficient to form links between objects. Therefore, in Experiment 3, we investigated how learning proceeds when object co-occurrence structure provides the only link between objects.

Experiment 3

In Experiment 3, we tested whether learners track object co-occurrence regularities in the absence of a correlated visual context cue (see Fig. 2-Training). The design was identical to Experiments 1 and 2, with two exceptions. First, we presented all objects without background images during training (as in Fig. 1A), removing the visual context cue. Second, we manipulated object co-occurrences in two conditions mirroring the co-occurrence structures of Experiments 1 and 2: *Skewed Condition* (identical to the object co-occurrence structure in Experiment 1, without a visual cue context cue) and *Uniform Condition* (identical to the object co-occurrence structure in Experiment 2, without a visual context cue; see Fig. 3). Thus, the only contextual structure present in the Skewed Condition of Experiment 3 was the regularity with which object pairs co-occurred; the Uniform Condition served as a control containing no object-pair regularities. Beyond allowing us to compare Experiments 1 and 2 (with a visual context cue) to the two conditions of Experiment 3 (without a visual context cue), the Uniform Condition control was included to ensure that participants in the Skewed Condition were selecting object pairs at test due to their co-occurrence frequency during training, not due to any visual similarity or other shared characteristic of the paired objects. We predicted that participants would continue to track object-object associations while successfully mapping labels to their target objects in the Skewed Condition.

Removing the background cue also provided a way to disentangle the effects of the visual context cues and co-occurrence structure on the word-learning/object-connection-learning tradeoff. If we continued to see a tradeoff between word learning and object-object association accuracy in the Skewed Condition, as observed in Experiment 1, this would suggest that the object co-occurrence structure alone was responsible for the effect. However, if we did not observe the same tradeoff, this would suggest that the shared visual background cues contributed to the tradeoff observed in Experiment 1.

Method

Participants

We recruited 101 new participants (51 female; 98 native speakers of English) through Amazon Mechanical Turk (mean age: 33.0 years, range: 18–67 years). Compensation was identical to Experiments 1 and 2. Participants were assigned to the *Skewed Condition* (n = 50) or the *Uniform Condition* (n = 51). The average time participants spent on the study was similar to Experiment 1 ($M = 7.1$ min, $SD = 2.3$).

Design & procedure

Training phase. The design was identical to Experiments 1 & 2, except that the backgrounds were removed from each item (see Fig. 2-Training). The object co-occurrence structure in the Skewed Condition was identical to the object co-occurrence structure in Experiment 1: each object occurred more frequently with one other object across training. In the Uniform Condition, no objects were linked together in terms of their co-occurrence, as in Experiment 2.

Test phase. The test phase was identical to Experiments 1 & 2. Participants in both the skewed and the uniform condition completed the same test trials. In the Uniform Condition, participants had no information on which to base their object choice. However, for the purposes of comparing performance between the Uniform and Skewed condition, we term the object that served as a target in the Skewed Condition as the 'target object' in the Uniform Condition as well.

Results

Word learning

To test participants' word learning in Experiment 3, we fit the same logistic mixed-effects model used to compare Experiments 1 and 2, predicting participants' trial-by-trial accuracy from the trial number for each label (1–6; centered), co-occurrence structure condition (uniform vs. skewed; centered, −.5, +.5), and their interaction and including the same random effects structure. Overall, participants mapped novel objects to their respective labels more often than would be expected by chance ($M = 40.8\%$), $b = .70$, 95% CI = [.49, .90], $z = 6.62$, $p < .001$ (see Fig. 4B). Participants were similarly accurate overall in the skewed condition ($M = 40.9\%$) and the uniform condition ($M = 40.7\%$), $b = .02$, 95% CI = [-.26, .30], $z = .17$, $p = .86$. As in Experiments 1 and 2, accuracy improved with each trial for a given target label, $b = .20$, 95% CI = [.15, .24], $z = 9.01$, $p < .001$. The interaction between co-occurrence condition and trial number was not significant, $b = -.07$, 95% CI = [−.15, .02], $z = -1.53$, $p = .13$.

Object-object association memory

We predicted that participants would remember which objects occurred together in the skewed condition, even in the absence of the shared visual background cue. Participants in the (control) uniform condition were expected to perform at chance, since they had encountered no object-pair regularities that could guide their choices. To investigate whether participants tracked object co-occurrences, we fit the same logistic mixed-effects model used to compare Experiments 1 and 2.

Participants were more likely to choose the target object in the skewed condition than in the uniform condition, $b = .42$, 95% CI = [.02, .83], $z = 2.04$, $p = .04$ (see Fig. 5B). In the skewed condition, participants' probability of choosing the high co-occurring object (M = 42.5%) reliably differed from chance, $b = .36$, 95% CI = [.07, .64], $z = 2.45$, $p = .01$. As expected, participants in the uniform condition chose the target object at chance levels (M = 33.6%), $b = -.07$, 95% CI = [−.36, .23], $z = -.44$, $p = .66$.

Relationship between word learning & object-object association memory

Unlike Experiments 1 and 2, there was no evidence of a relationship between word learning and object-object association accuracy (see Fig. 6B). Object-object association accuracy and condition did not interact, $b = -.05$, 95% CI = [−.61, .50], $z = -0.19$, $p = .85$. Object-object association accuracy did not predict word learning performance in either the skewed condition ($b = -.08$, 95% CI = [−.49, .34], $z = -0.36$, $p = .72$) or in the uniform condition ($b = -.02$, 95% CI = [−.43, .38], $z = -0.11$, $p = .91$).

Discussion

In Experiment 3, we found that adults tracked which objects occurred together in a word learning task in the absence of a correlated cue (i.e., a shared visual background) highlighting the co-occurring objects. Additionally, the overall word-learning pattern remained similar across the three experiments. Unlike Experiment 1, we found no evidence for a tradeoff between word learning and object-object association accuracy in the Skewed Condition of Experiment 3. Participants successfully learned both the novel words and the associations among objects. This suggests that, once the correlated visual cues highlighting co-occurring objects were removed, participants' ability to learn the object-object relationships did not affect their ability to learn novel label-object mappings. In the General Discussion, we describe how this effect, together with the results of Experiments 1 and 2, suggest an attentional role for visual context cues.

General discussion

This study was designed to determine whether adults go beyond learning word-object associations to learn object-object links in referentially ambiguous word-learning situations. Across three experiments, we found evidence that adults extracted both label-object associations and object-object connections when contextual structure was introduced in a cross-situational word learning task. When object co-occurrences were skewed, either with (Experiment 1) or without (Experiment 3) visual context cues, learners formed links between frequently co-occurring objects. Visual context cues alone were not sufficient for learners to encode connections between objects (Experiment 2). These findings suggest that adults learn more than just label-object mappings when object co-occurrence regularities – but not visual context cues – are introduced in an ambiguous word learning situation: they also learn about properties of the contextual structure itself, namely which objects are associated with one another.

Contextual regularities such as frequently co-occurring objects are often framed as adding difficulty to the word learning task, increasing ambiguity for a learner attempting to map a label to an object. Our results suggest a different perspective. At least under some circumstances, learners can exploit these contextual regularities to simultaneously learn about label-object mappings and relationships between objects. Participants not only successfully learned novel words despite the challenge presented by frequently co-occurring object pairs, but they also formed associations between objects. Indeed, in Experiment 3, there was no statistical difference in word learning between the skewed and uniform distribution condition.

These results demonstrate adults' ability to track multiple types of statistics concurrently. They also suggest a mechanism by which learners can build up a semantic system, rather than solely word-referent associations. Just as objects occur in clusters in the real world (e.g., tomatoes with other vegetables), the additional co-occurrence structure in the skewed condition contained valuable information about the relationships between objects and the semantic structure of their environment (Sadeghi et al. 2015). Our experiments demonstrate that adult learners can capitalize on this structure, linking referents to novel labels while simultaneously learning about relationships between the referents themselves.

We also found evidence for a tradeoff between word learning and forming associations between objects in the presence of visual contextual cues. In Experiment 1, in the presence of both object co-occurrence and visual context cues, there was a trade-off between successful word learning and remembering object-object associations. In Experiment 2, in the presence of a visual context cue but no object co-occurrence cue, this trend was reversed, such that better word learning was associated with (marginally) better learning of object-object associations. In Experiment 3, when the visual context cue was removed, we observed no connection between word learning and object-object association memory regardless of object-object co-occurrence distribution. This pattern of results suggests that the presence of additional types of contextual structure (visual context cues) moderated trade-offs between word learning and forming object-object associations was.

Visual context cues may have influenced the relationship between learning label-object and object-object connections because of how they directed attention during the word learning task. Previous research has shown that attention is guided by co-varying visual cues (Chun, 2000; Chun & Jiang, 1999) and associative object knowledge (Moores, Laiti, & Chelazzi, 2003). In our study, visual context cues in the form of object backgrounds may have drawn participants' attention to a narrower set of possible referents for a given label (i.e. to the two objects sharing the same background). When these objects occurred together frequently, as in Experiment 1, accuracy on the word learning task may have been negatively affected because these objects continued to co-occur throughout the task, making them more difficult to distinguish as potential referents. On the other hand, when objects with shared backgrounds did not co-occur together frequently, as in Experiment 2, encoding visual context cues was marginally associated with better word learning, possibly because visual context cues guided attention towards a smaller set of potential referents that were more easily disambiguated on later trials. More broadly, the degree to which coherently co-varying structure supports word learning may depend in part on the presence of distinctive moments that aid in clearly disambiguating potential referents (see Roy et al., 2015). While coherent co-variation may initially help scaffold attention towards the right set of associated objects (Smith et al., 2014), forming label-object mappings may be impeded if co-varying structure makes it difficult to resolve referent ambiguity on later learning trials.

The current results demonstrate adults' ability to extract multiple types of information from object-object and label-object statistical structure across learning trials, supporting previous findings that people track complex information during cross-situational word learning (Dautriche & Chemla, 2014; Roembke & McMurray, 2016; Smith et al., 2014; Yurovsky et al., 2014). While our experiments do not illuminate the specific mechanisms that allowed participants to learn multiple regularities, our findings are predicted by a statistical, incremental account of cross-situational word learning (Smith et al., 2014), whereby participants exploit statistical regularities present in the word learning task. On this theory, learners' ability to incrementally track associations allows them to both encode regularities between labels and objects and between regularly co-occurring objects themselves. However, a hypothesis-testing theory of ambiguous word learning, such as the propose-and-verify account (e.g., Trueswell et al., 2013), could also accommodate our results if traces of earlier word-referent hypotheses are available to learners. If participants can recall objects previously hypothesized to be the referents of a given label, they could use this information during the object-object association recall task. Specifically, co-occurring objects could be indirectly connected via the same target label, since co-occurring objects are more likely to have been hypothesized as potential referents for that label. Future work is needed to understand which of these two mechanisms are responsible for learning object-object associations.

Given the short duration of training (seven minutes on average), the current experiments do not tell us how associative knowledge about relationships between objects and labels unfolds over the course of extended learning experience. Moreover, our test of object-object association memory was a more explicit measure of knowledge than other methods used to assess whether people encode associations between objects (see, e.g., Chun & Jiang, 1999; Roembke & McMurray, 2016). Similarly, the word learning task itself encouraged participants to make explicit choices about label-object mappings, rather than allowing them to more passively absorb statistical regularities between objects and labels (Romberg & Yu, 2015). It is all the more notable that we found robust evidence that participants formed associations between objects over a short exposure phase, even though they were engaged in an explicit word learning task. In future work, it will be important to investigate to what extent infants also rapidly acquire multiple types of associative information across multiple ambiguous and implicit word learning instances.

In the process of learning how words refer to objects in our environment, contextual structure presents both a hurdle to disambiguating word meanings as well as an opportunity to learn more about the relationships between entities in the world. These studies add to a growing language learning literature suggesting that human learners are adept at exploiting the statistical structure present in the learning environment, tracking multiple types of information – label-object mappings and object-object relationships – at the same time (Romberg & Saffran, 2013; Wojcik & Saffran, 2013; Wojcik & Saffran, 2015). This ability to sensitively encode contextual information in the environment may help to explain how we are able to not only map labels to objects across ambiguous situations, but develop rich semantic knowledge at the same time.

Acknowledgments

This research was supported by NSF-GRFP DGE-1256259 awarded to MZ, NSF SPRF 1513834 awarded to VLB, and grants from the NICHD to JRS (R37HD037466) and the Waisman Center (P30HD03352, U54 HD090256). We would like to thank Sumarga (Umay) Suanda for helpful comments on a previous version of this manuscript.

References

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, & N. Akhtar, et al. (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). Oxford: Oxford University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832–843. https://doi.org/10.1037//0012-1649.29.5.832.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*, 617–629. https://doi.org/10.1038/nrn1476.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*, 328. https://doi.org/10.3389/fpsyg.2013.00328.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes.

Boyce, S. J., Pollatsek, a., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology. Human Perception and Performance, 15*, 556–566. https://doi.org/10.1037/0096-1523.15.3.556.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 32*, 699–706. https://doi.org/10.1037/0278-7393.32.4.699.

Chen, C., & Yu, C. (2017). Grounding statistical learning in context: The effects of learning and retrieval contexts on cross-situational word learning. *Psychonomic Bulletin and Review, 24*, 920–926. https://doi.org/10.3758%2Fs13423-016-1163-x.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*, 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10*, 360–366.

Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 892–903. https://doi.org/10.1037/a0035657.

Hills, T. (2013). The company that words keep: Comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language, 40*, 586–604. https://doi.org/10.1017/S0305000912000165.

Hills, T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language, 63*, 259–273. https://doi.org/10.1016/j.jml.2010.06.002.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446. https://doi.org/10.1016/j.jml.2007.11.007.

Kachergis, G., Yu, C., & Shiffrin, R. M (2009). Frequency and contextual diversity effects in cross-situational word learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of Knowledge. *Psychological Review, 104*, 211–240. https://doi.org/10.1037/0033-295X.104.2.211.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 121–157. https://doi.org/10.1016/0010-0285(88)90017-5.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 9014–9019. https://doi.org/10.1073/pnas.1105040108.

Miller, G. A. (1999). On knowing a word. *Annual Review of Psychology, 50*, 1–19. https://doi.org/10.1146/annurev.psych.50.1.1.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6*, 182–189. https://doi.org/10.1038/nn996.

Morey, R. D. (2008). Confidence Intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*, 61–64. https://doi.org/10.20982/tqmp.04.2.p061.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19*, 498–550. https://doi.org/10.1016/0010-0285(87)90017-X.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*, 520–527. https://doi.org/10.1016/j.tics.2007.09.009.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development, 77*, 266–280. https://doi.org/10.1111/j.1467-8624.2006.00869.x.

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

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language, 87*, 105–127. https://doi.org/10.1016/j.jml.2015.09.005.

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science, 37*, 1290–1320. https://doi.org/10.1111/cogs.12050.

Romberg, A. R, & Yu, C. (2015). Interactions between statistical aggregation and hypothesis testing mechanisms during word learning. In E. Grillo & K. Jepson (Eds.), *Proceedings of the 39th annual boston university conference on language development* (pp. 1311–1316). Somerville, MA: Cascadilla Press.

Roy, B. C., Frank, M. C., Decamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *PNAS, 112*, 12663–12668. https://doi.org/10.1073/pnas.1419773112.

Sadeghi, Z., McClelland, J. L., & Hoffman, P. (2015). You shall know an object by the company it keeps: An investigation of semantic representations derived from object co-occurrence in visual scenes. *Neuropsychologia, 76*, 52–61. https://doi.org/10.1016/j.neuropsychologia.2014.08.031.

Saji, N., Imai, M., Saalbach, H., Zhang, Y., Shu, H., & Okada, H. (2011). Word learning does not end at fast-mapping: Evolution of verb meanings through reorganization of an entire semantic domain. *Cognition, 118*, 48–64. https://doi.org/10.1016/j.cognition.2010.09.007.

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PLoS ONE, 6*. https://doi.org/10.1371/journal.pone.0028095.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences, 18*, 251–258. https://doi.org/10.1016/j.tics.2014.02.007.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*, 343–348. https://doi.org/10.1016/S1364-6613(03)00156-6.

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

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66*, 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition, 127*, 375–382. https://doi.org/10.1016/j.cognition.2013.02.015.

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology, 108*, 394–401. https://doi.org/10.1016/j.jecp.2010.09.011.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology, 45*, 1611–1617. https://doi.org/10.1037/a0016134.

Wojcik, E. H., & Saffran, J. R. (2013). The ontogeny of lexical networks: Toddlers encode the relationships among referents when learning novel words. *Psychological Science, 24*, 1898–1905.

Wojcik, E. H., & Saffran, J. R. (2015). Toddlers encode similarities among novel words from meaningful sentences. *Cognition, 138*, 10–20. https://doi.org/10.1016/j.cognition.2015.01.015.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x.

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. https://doi.org/10.1111/j.1467-7687.2010.00958.x.

Yurovsky, D., Fricker, D., Yu, C., & Smith, L. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review, 21*, 1–22. doi: 10.3758%2Fs13423-013-0443-y.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science, 37*, 891–921. https://doi.org/10.1111/cogs.12035.