An associative model of adaptive inference for learning word–referent mappings

George Kachergis, Chen Yu & Richard M. Shiffrin

Psychonomic Bulletin & Review

ISSN 1069-9384
Volume 19
Number 2

Psychon Bull Rev (2012) 19:317-324
DOI 10.3758/s13423-011-0194-6

Springer

Psychon Bull Rev (2012) 19:317–324
DOI 10.3758/s13423-011-0194-6

BRIEF REPORT

An associative model of adaptive inference for learning word–referent mappings

George Kachergis · Chen Yu · Richard M. Shiffrin

Published online: 4 January 2012

Abstract People can learn word–referent pairs over a short series of individually ambiguous situations containing multiple words and referents (Yu & Smith, 2007, Cognition 106: 1558–1568). Cross-situational statistical learning relies on the repeated co-occurrence of words with their intended referents, but simple co-occurrence counts cannot explain the findings. Mutual exclusivity (ME: an assumption of one-to-one mappings) can reduce ambiguity by leveraging prior experience to restrict the number of word–referent pairings considered but can also block learning of non-one-to-one mappings. The present study first trained learners on one-to-one mappings with varying numbers of repetitions. In late training, a new set of word–referent pairs were introduced alongside pretrained pairs; each pretrained pair consistently appeared with a new pair. Results indicate that (1) learners quickly infer new pairs in late training on the basis of their knowledge of pretrained pairs, exhibiting ME; and (2) learners also adaptively relax the ME bias and learn two-to-two mappings involving both pretrained and new words and objects. We present an associative model that accounts for both results using competing familiarity and uncertainty biases.

Keywords Statistical learning · Language acquisition · Cross-situational learning

G. Kachergis (✉) · C. Yu · R. M. Shiffrin
Department of Psychological & Brain Science,
Cognitive Science Program,
Bloomington, IN 47405, USA
e-mail: gkacherg@indiana.edu

C. Yu
e-mail: chenyu@indiana.edu

R. M. Shiffrin
e-mail: shiffrin@indiana.edu

Learning the first nouns of a new language can be challenging, since there are many possible referents in any situation. But by seeing objects in varied situations and tracking which words and objects co-occur most frequently, learners can infer words' intended referents. This ability, termed *cross-situational statistical learning*, likely plays a role in infant language acquisition (Gleitman, 1990).

In the cross-situational word-learning paradigm (Yu & Smith, 2007), adult participants are instructed that they will be learning which words go with which objects and then will view a sequence of training trials, each of which contains multiple novel objects and multiple spoken pseudowords. Although a word and its correct referent always appear together on a learning trial (i.e., there are one-to-one mappings), there is ambiguity on any single trial concerning which words are associated with which referents. Nonetheless, as pairs are repeated across trials with various other pairs, if learners are able to track which words and objects consistently co-occur, they can learn the correct mappings. Yu and Smith (2007) trained a vocabulary of 18 word–object pairs over a series of 27 learning trials, with 4 pairs/trial, and 6 repetitions/pair; in 5 min of training, adults learned about 9 of the 18 pairs.

Although learners acquire many correct word–referent mappings in this task, it is very unlikely that they track all word–object co-occurrences or attend equally to all 16 possible word–referent pairings from four words and four objects on a trial. Instead, people likely learn by selectively attending to a subset of word–referent mappings, chosen by strategies that restrict the space of possible pairings. One principle for selective attending that leads to an effective learning strategy is based on ME, a constraint holding that words are mapped one-to-one to objects (Markman, 1990; Markman & Wachtel, 1988; Merriman & Bowman, 1989). Suppose that a single trial

contains two words $\{w_1,w_2\}$ and two objects $\{o_1,o_2\}$. If a subsequent trial has words $\{w_2,w_3\}$ and objects $\{o_2,o_3\}$, a learner employing ME can infer that the new word w_3 should map to the new object o_3, if he or she remembers that w_2 previously occurred with o_2. Empirical evidence from cross-situational learning supports this: In studies with some pairs that are either explicitly pretrained or presented more frequently, learning improves—even for unpretrained and low-frequency pairs (Kachergis, Yu, & Shiffrin, 2009a; Klein, Yu, & Shiffrin, 2008).

However, if ME were the only constraint, then when a word (or object) already has a known associate, new associations to the word would not be formed. For instance, knowing w_1–o_1, learners would have difficulty associating w_2 with o_1 despite consistent co-occurrence. Thus, ME may facilitate learning of one-to-one word–referent mappings but could inhibit learning of many-to-many mappings between words and referents. Since human language learners acquire both homonyms (one-to-many word–object mappings) and synonyms (many-to-one), simple ME is unlikely to be the only mechanism at work.

To understand what role ME plays in statistical word learning, we first present an experimental study (depicted in Fig. 1). This study systematically varied the frequency of one-to-one pairs (e.g., w_1–o_1) in an early stage of cross-situational training in order to have these ME-compliant mappings learned to various degrees. In a late stage, we added both a new word (e.g., w_7) and a new object (o_7) to each early one-to-one pair. That is, a late trial always contained an early pair, pretrained to some extent, and a late pair that always appeared with that early pair (see Fig. 1). An ME-biased learner presented with late-stage trial T = $\{w_1,w_7,o_1,o_7\}$ who already knew that w_1 went with o_1 would infer that w_7 maps to o_7. A naïve learner, despite knowledge of w_1–o_1, would associate w_7 as much with o_1 as with o_7, learning a homonymous relation. Would such cross-stage associations be learned despite a large amount of early-stage training? We also ask whether sufficient late-stage training overcomes the ME bias, resulting in learning of w_7–o_1 and w_1–o_7.

Motivated by the empirical results, we introduce an associative model that explains how both fast ME-based inference and non-ME learning can arise from two simple mechanisms: competing attentional biases for familiar pairings and for stimuli with uncertain associations. In the context of the model, we show how an ME bias arises early in learning and how it gradually attenuates in response to additional evidence for non-ME pairings. The success of the model demonstrates that a general associative model with attention (controlled by both pair familiarity and stimulus uncertainty) may be a cognitively plausible learning framework for cross-situational learning of word–referent mappings.

Experiment

Each training trial contained two objects and two spoken pseudowords. Word presentation order was randomized, and there was no indication of which word referred to which

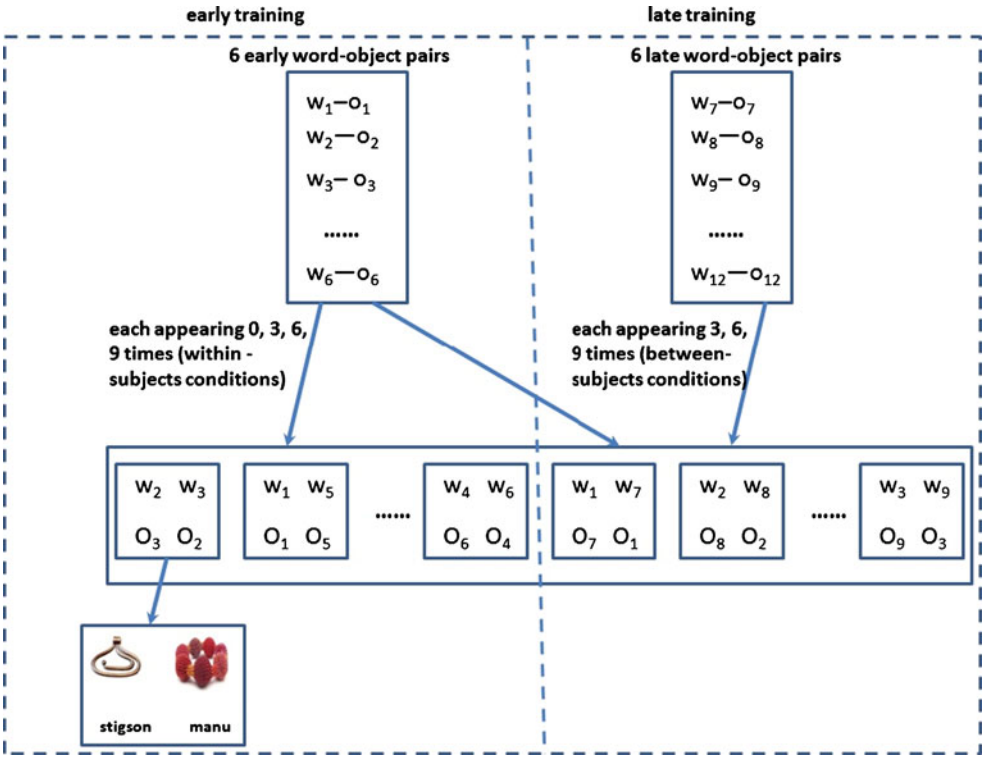

Fig. 1 Example trials from the early and late stages of training. In the early stage, the six 6 early word–object pairs co-occur randomly, allowing correct mappings to be learned cross-situationally. In the late stage, each early pair (e.g., w_1–o_1) only co-occurs only with a unique late pair (e.g., w_7–o_7). Thus, a learner employing ME will infer that the late word (w_7) maps to the late object (o_7) and not to the early object (o_1), whereas an unbiased learner may learn early-to-late (w_1–o_7) and late-to-early (w_7–o_1) associations

object. However, since words occurred only on trials with their intended referents, "correct" pairings (those occurring most frequently across trials) were disambiguated over the series of trials. We systematically co-varied the number of times a given pair occurred in both an early stage and a late stage of learning. Half of the pairs appeared only in the late stage. As is shown in Fig. 1, when an early-stage pair w_1–o_1 appeared in the late stage, a specific late-stage pair (w_7–o_7) always co-occured with w_1–o_1. Therefore, if a learner experienced only the late stage, all four possible associations (w_1–o_1, w_1–o_7, w_7–o_1, w_7–o_7) were equiprobable. If a learner has no mutual exclusivity bias, the late stage alone (i.e., 0 early) should induce a two-to-two mapping: Each word (homonym) maps to two objects, and each object has two words (synonyms) mapping to it. But in the early stage, learners acquire one-to-one mappings (e.g., w_1–o_1), which may accentuate the learning of late-stage novel-to-novel pairs (e.g., w_7–o_7) but, meanwhile, block the learning of novel-to-familiar cross-stage mappings (w_1–o_7, w_7–o_1).

Method

Participants Ninety-six undergraduates at Indiana University participated in the three between-subjects conditions to receive course credit (33 in 3 late repetitions, 29 in 6 late, and 34 in 9 late). None had participated in other cross-situational experiments.

Stimuli Stimuli were 48 images of unusual objects and 48 spoken pseudowords (see Fig. 1). The pseudowords were computer generated, phonotactically probable in English, and pronounced by a monotone, synthetic female voice.

Design On each 8-s training trial, two objects were displayed throughout, while two pseudowords were heard successively: 2 s of silence preceded each 1-s word, and the last was followed by 2 s of silence. The 48 objects and words were randomly assigned to four sets of 12 word—object pairings, one set used for each within-subjects condition. As is shown in Fig. 1, six pairs in a set appeared only in the late (two-to-two) stage of training, and the other six in the set appeared in both the early (one-to-one) and late training stages. Within-subjects conditions varied in the amount of early training: The six early pairs appeared zero, three, six, or nine times in early training, so the blocks had 0, 9, 18, or 27 early training trials, respectively, each displaying two early pairs. Condition order was counterbalanced, and each learner participated in all four early repetition conditions for a single level of late-stage repetitions.

In each between-subjects condition, the late training stage had the same structure: Each late pair appeared three (3-late condition), six (6-late condition), or nine (9-late condition) times, for a total of 18, 36, or 54 trials. As in the early stage, each trial contained two pairs, and there was no pause or marker between the early and late stages. Each late-stage trial had one late pair and one pair from the early stage (e.g., $\{w_1,w_7,o_1,o_7\}$). To reiterate, each late pair appeared solely and consistently with a unique early-stage pair.

Procedure Learners were instructed that they would see a series of trials with two objects and two artificial words and that they should try to figure out what each word referred to for a final test. Participants were not told that there were two stages of training, and there was no perceptible break. After training, their knowledge was assessed using 11-alternative forced choice (11AFC) testing: On each test trial, a single word was played, and participants were instructed to choose the appropriate object from a display of 11 of the 12 trained referents. In order to assess knowledge about both associations, each word was tested twice (but spaced): once without its corresponding early object, and once without its late object. For example, to test w_1–o_1, the 11 objects on the testing trial excluded o_7. Similarly, the testing trial for w_1–o_7 contained all of the other 11 objects except o_1. For conditions with both early and late training stages, the data were sorted into four cases. Consider the late study trial $\{w_1,o_1, w_7,o_7\}$, in which w_1 and o_1 were studied both early and late and w_7 and o_7 were studied only late. We thus tested four associations: w_1–o_1, w_1–o_7, w_7–o_7, and w_7–o_1. The within-stage associations w_1–o_1 and w_7–o_7 were compatible with ME, whereas w_1–o_7 and w_7–o_1 were across-stage pairings that should not be learned under strict ME. Also note that for conditions with no early training, early and late pairs were indistinguishable, since both always co-occurred.

Results and discussion

Figure 2 displays learning performance[1] for the three between-subjects levels of late stage repetitions (three, six, and nine late repetitions in top panels, left to right). The figure also shows predictions of the model (lower panels) described later. Within each of the top three panels, the results from four early-pair repetition conditions (0, 3, 6, and 9) are reported: The single dot shows performance for the condition with no early stage, and for the other three conditions with early training, learning results were separated on the basis of early-, late-, and cross-stage pairing types. For the 3-late condition, a 3 × 2 × 2 ANOVA with factors of early repetitions (three, six, or nine), pair stage (early or late), and pairing type (within-stage or across-stage) showed

[1] Two participants' data from the 3-late condition were excluded because their overall mean performance was at chance (11AFC: .091). The outcomes of statistical tests were unaffected.

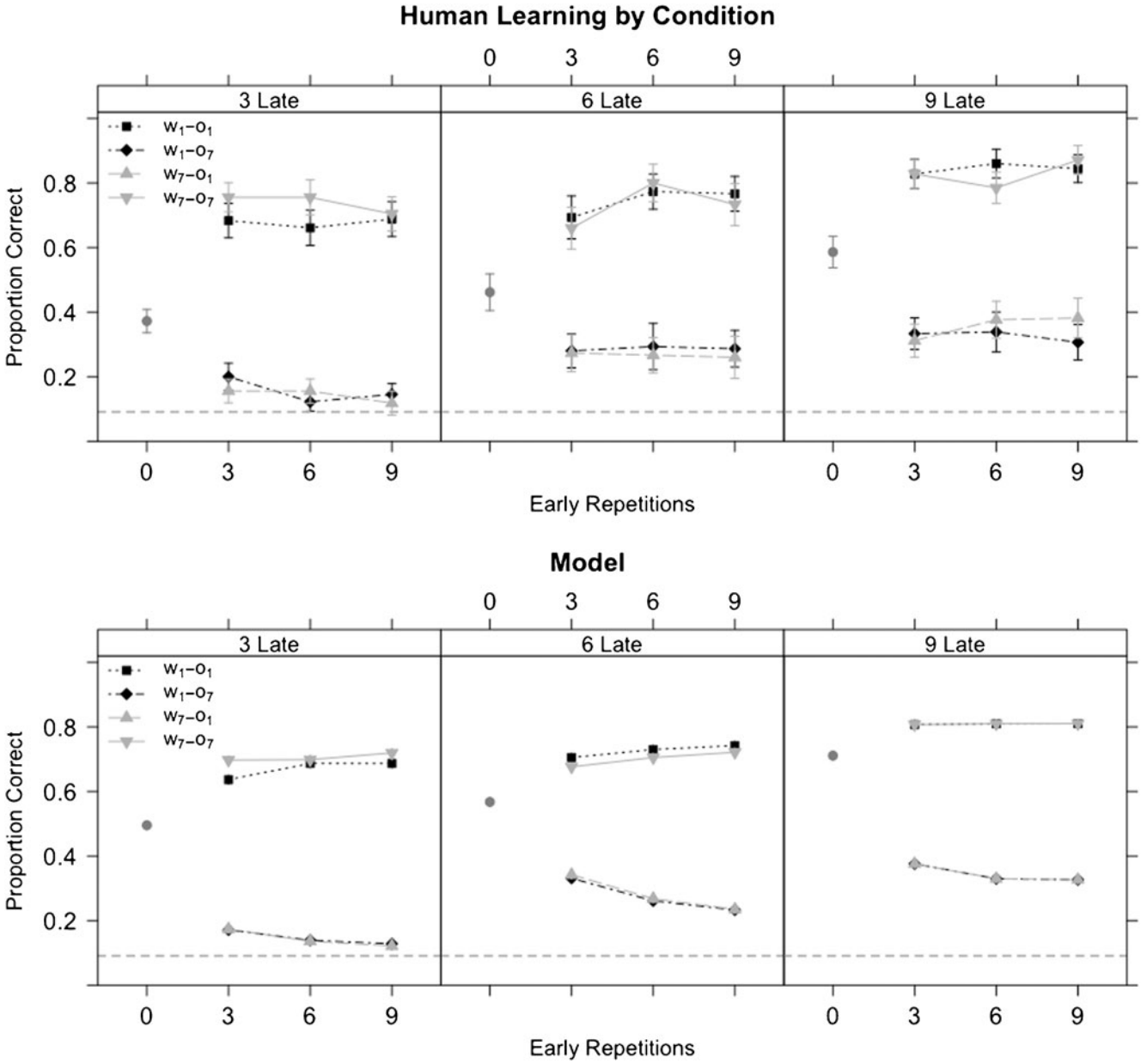

Fig. 2 The top row shows mean participant accuracy by number of late pair repetitions (panels), early pair repetitions (*x*-axis), and by association type (line and symbol type). Participants learned ME-compatible pairs (w_1–o_1 and w_7–o_7) quite well, even with only three 3 late pair repetitions (e.g., w_7–o_7). Learning of across-stage pairs (w_1–o_7 and w_7–o_1) is quite modest by comparison. As compared with to the 3- late condition, learning of across-stage pairs significantly improved in the 6- late condition, improving further in the 9- late condition. Thus, participants responded to increasing evidence for cross-stage associations by learning more such pairings. The bottom row of panels shows learning of the best-fitting model for the same training conditions. The model captures the high level of learning of the mutually -exclusive pairs——without regard for the number of early repetitions (excluding 0, for which both humans and the model show an increase), as well as the increase in learning of cross-stage associations with more late repetitions. The horizontal line shows chance (11AFC; .091). Error bars show ± +/-*SE* for humans. Model predictions are probabilistic, —not simulated, —and thus have no variability

only a significant main effect of pairing type, $F(1, 30) = 8.62$, $p = .004$. Within-stage (i.e., w_1–o_1 and w_7–o_7) pairs were learned far better than across-stage pairs ($M_{within} = .71$, $M_{across} = .15$), paired $t(30) = 12.41$, $p < .001$, $d = 3.26$, and even across-stage (e.g., w_1–o_7) learning was above chance, paired $t(30) = 2.29$, $p = .03$, $d = .60$. The indistinguishably high learning of w_1–o_1 and w_7–o_7 suggests strong, ME-based inference: Late pairs (w_7–o_7) were learned by filtering out the consistently co-occurring early pair (w_1–o_1). Surprisingly, performance did not significantly vary across three, six, and ninr early repetitions, $F(2, 30) < 1$: Three repetitions were sufficient for learning early pairs and using

them for inference in the late training, and further repetitions did not significantly improve learning. With no early stage, as might be expected, performance falls between the other performance levels (M = .32, well above chance), paired t (30) = 8.83, $p < .001$, d = 2.32.

A mixed ANOVA (three, six, or nine late repetitions [between subjects] × 3, 6, or 9 early repetitions × early or late pair stage × across- or within-stage pairing type) showed a main effect of pairing type, $F(1, 101) = 161.76$, $p < .001$, and an interaction between pairing type and the number of late repetitions, $F(2, 101) = 10.89$, $p = .001$. As in the 3-late condition, filtering with the use of ME produced excellent learning of the late pairs. However, the increased late training did not much improve the relative amount of learning of the ME-compliant within-stage pairs ($M_{3\text{-Late}}$ = .71, $M_{6\text{-Late}}$ = .74, $M_{9\text{-Late}}$ = .84). Nonetheless, learning of the across-stage pairs (pairings that are inconsistent with strong ME) increases with additional late training ($M_{3\text{-Late}}$ = .15, $M_{6\text{-Late}}$ = .28, $M_{9\text{-Late}}$ = .34).

These results suggest that learners initially utilize an ME bias, leveraging it to infer that new words refer to new objects. However, in the face of additional co-occurrences of early pairs with late pairs, they come to realize that words refer to multiple objects (and that objects have more than one word) and begin to learn cross-stage associations. We now develop a computational model to provide a mechanistic account of the underlying learning processes producing these results. Although both rule-based and associative approaches have been used successfully to model language acquisition (for an overview, see Broeder & Murre, 2002), models using rule-based ME (e.g., Siskind, 1996) would have difficulty showing the gradual relaxation of ME demonstrated in our experiment. Thus, we propose an associative model that captures the results via an interaction of two previously implicated mechanisms: familiarity and novelty biases.

Model

The model assumes that learners do not attend equally to all possible word–object pairings (i.e., store all co-occurrences). Rather, attention to and storage of the pairings on a trial is preferentially directed to those that have previously co-occurred. However, this bias for familiar pairings competes with a bias to attend to stimuli that have no strong associates (e.g., novel stimuli). For example, on the first trial in the late stage $\{w_1,o_1,w_7,o_7\}$, w_1–o_1 demands more attention than does w_7–o_7, w_1–o_7, or w_7–o_1, since w_1 and o_1 have been associated. However, attention is also pulled individually to w_7 and to o_7, since these stimuli have no strong associates yet and need to be learned. That is, they have high uncertainty, quantified by the entropy of their association strengths, and they thereby attract attention. Thus, at first, attention is mostly divided between w_1–o_1 and w_7–o_7, although the remaining two possible pairings (e.g., w_7–o_1 and w_7–o_1) may also receive a small amount of attention.

Formally, given n words and n objects to be learned over a series of trials, let M be an n word × n object association matrix that is incrementally built during training. Cell $M_{w,o}$ will be the strength of association between word w and object o. Strengths are subject to forgetting (i.e., general decay) but are augmented by viewing of the particular stimuli. Before the first trial, M is empty. On each training trial t, a subset S of m word–object pairings appears. If there are any new words and objects are seen, new rows and columns are first added. The initial values for these new rows and columns are k, a small constant (here, 0.01).

Association strengths are allowed to decay, and on each new trial, a fixed amount of associative weight, χ, is distributed among the associations between words and objects and added to the strengths. The rule used to distribute χ (i.e., attention) balances a preference for attending to unknown stimuli with a preference for strengthening already-strong associations. When a word and referent are repeated, extra attention (i.e., χ) is given to this pair—a bias for prior knowledge. Pairs of stimuli with no or weak associates also attract attention, whereas pairings between uncertain objects and known words, or vice versa, do not attract much attention. To capture stimulus uncertainty, we allocate strength using entropy (H), a measure of uncertainty that is 0 when the outcome of a variable is certain (e.g., a word appears with one object and has never appeared with any other object) and maximal ($\log_2 n$) when all of the n possible object (or word) associations are equally likely (e.g., when a stimulus has not been observed before, or if a stimulus were to appear with every other stimulus equally). In the model, on each trial, the entropy of each word (and object) is calculated from the normalized row (column) vector of associations for that word (object), $p(M_{w,\cdot})$, as follows:

$$H\left(M_{w,\cdot}\right) = -\sum_{i=1}^{n} p\left(M_{w,i}\right) \cdot \log\left(p\left(M_{w,i}\right)\right)$$

The update rule for adjusting and allocating strengths for the stimuli presented on a trial is

$$M_{w,o} = \alpha M_{w,o} + \frac{\chi \cdot e^{\lambda \cdot (H(w)+H(o))} \cdot M_{w,o}}{\sum_{w \in S} \sum_{o \in S} e^{\lambda \cdot (H(w)+H(o))} \cdot M_{w,o}}$$

In this equation, α is a parameter governing forgetting, χ is the weight being distributed, and λ is a scaling parameter governing differential weighting of uncertainty [novelty; H (.)] and prior knowledge (familiarity; $M_{w,o}$). As λ increases, the weight of uncertainty (i.e., the exponentiated entropy term, which includes both the word's and the object's association entropies) increases, relative to familiarity. The

denominator normalizes the numerator so that exactly χ associative weight is distributed among the potential associations on the trial. For stimuli not on a trial, only forgetting operates. After training and prior to test, a small amount of noise ($c = .01$ here) is added to M. At test, learners choose the associated referent for the word from the m alternatives in proportion to their strengths to the word. We fit the model separately to each between-subjects condition (13 means/condition), in which the number of late repetitions varied. In the bottom row of Fig. 1, we show the best model fits for these three conditions. In the 3-late condition, (parameters: 3-late, $\chi = 0.31$, $\lambda = 2.34$, $\alpha = 0.91$; 6-late, $\chi = 0.20$, $\lambda = 0.88$, $\alpha = 0.96$; 9-late, $\chi = 3.01$, $\lambda = 1.39$, $\alpha = 0.64$). The total sum of squared error between these fits and subject means is 0.1. Although there are some differences between data and predictions, the model captures the major findings: Early pairs (w_1–o_1) were learned quite well, and late pairs (w_7–o_7) were quickly learned when introduced, because of the uncertainty bias. Thus, the model shows advantageous ME behavior as a result of a bias to associate uncertain words with uncertain objects. The competing bias to strengthen previous associations keeps uncertain stimuli from quickly becoming associated to stimuli that already have associates.

Remarkably, the level of learning of both ME-compatible pairing types did not much depend on the number of early or late pair repetitions. The model also shows the increase in learning of cross-stage (e.g., w_7–o_1) pairings with increasing late repetitions and the increased performance in the 0-early-stage conditions, although, here, the model overlearned in all three late stages. These are conditions in which two word–object pairs always co-occur, but never with other stimuli. Thus, the model has little trouble choosing the object at test that is one of only two that the word has been associated with. A more elaborate decision mechanism may alleviate this, but our focus is learning.

The qualitative fit to all 39 conditions is quite good. For comparison, the online Appendix shows the fit of two baseline models, one with only the familiarity mechanism, and the other with only the uncertainty mechanism. For the present data, the baseline uncertainty model fits about as well as the full model, but we focus on the full model in this article because the uncertainty model fails to account for frequency effects we have reported elsewhere (Kachergis, Shiffrin, & Yu, 2009a); the full model is able to handle those results. The Appendix also shows that the dual-mechanism model's fit across conditions is robust to perturbations in the parameters.

An analysis of the dual-process model reveals interesting properties that underlie the predictions. Figure 3 shows trial-by-trial changes in the fitted model's performance (Fig. 4 in the online Appendix gives details of the actual association strengths). In early training, the model quickly associates the early words and referents (e.g., w_1–o_1). When the late stage begins, attention switches almost entirely to the late pairs because the early words and objects are less uncertain (the new stimuli have maximal entropy). Thus, our model exhibits a bias for finding names for unlabeled objects, much like human infants and adults (Golinkoff, Hirsh-Pasek, Bailey, & Wegner, 1992). The early associations then drop in strength while the late associations grow, eventually becoming equal; the amount of time to reach equality is larger for more early

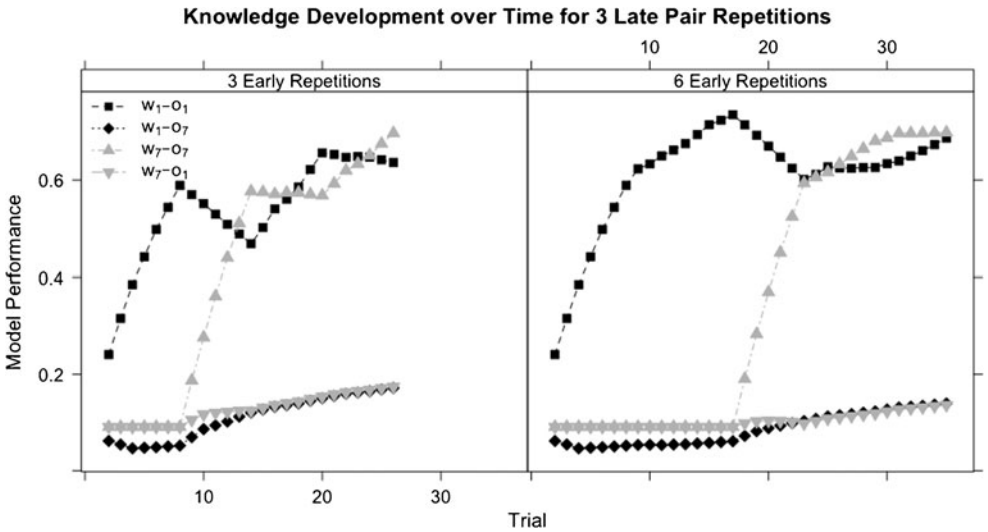

Fig. 3 The model's testing performance for different types of pairings over the course of training with 3 or 6 early pair repetitions (left and right panels) and with 3 late pair repetitions. In the early stage (the first 9 trials for the left panel, 18 for the right), early pairs (w_1–o_1) are quickly learned cross-situationally. When the late pairs (w_7–o_7) first appear, most attention is devoted to learning them, since their uncertainty is high. Early pairs suffer from neglect until late pairs are familiar enough for attention to balance out. Cross-stage pairs (e.g., w_1–o_7) are given slight attention throughout, initially due to the high uncertainty of the new late stimulus (o_7), and later because they have acquired some strength of their own

training, but after equality is reached, the two pair types act similarly. As the entropy of the late stimuli decreases, learning of cross-stage (ME-violating) pairings gradually increases. A late stimulus (w_7) will not become associated with an early stimulus (o_1) as strongly as o_7–w_7 because the early stimulus has less uncertainty.

General discussion

Learning word–referent mappings from a series of individually ambiguous trials is helped by using a form of ME bias. Such a bias significantly reduces the number of pairings a learner must consider on each trial by allowing learners to apply their current knowledge to infer the mappings of novel words and referents. Our results show behavior consistent with such a bias: Even with only three repetitions of the early pairs and three repetitions of the two-to-two late-stage training, late pairs (o_7–w_7) were learned to the same high degree as their corresponding early pairs (o_1–w_1). Surprisingly, learning of the ME-consistent mappings did not significantly improve with more early- or late-stage pair repetitions.

If pairings are not one-to-one, or if word–referent mappings may change over time, strict ME would be maladaptive. Between subjects, we increased the number of late pair repetitions and found increased learning of the across-stage mappings (e.g., w_7–o_1), a result that could be viewed as ME-violating. This suggests that participants do not use strict ME but use an adaptive approach that allows learning of one-to-many and many-to-one mappings, given sufficient evidence. What are the underlying cognitive mechanisms generating this flexible bias? There are several ways to think about this. For example, if the storage or retrieval of the early learning was equally imperfect for three, six, and nine early repetitions, ME might have been used when retrieval succeeded and learning distributed to both referents when retrieval failed. Given that we test each word twice, once without each of its two possible referents, it may be that what we call learning of a secondary association is, in fact, an episodic memory of that word often co-occurring with that object. Although we use a simple associative model rather than an episodic memory model, we recognize that knowledge develops in part from episodic memory; thus, we do not find this explanation entirely independent of our own: Memory and learning are two faces of the same coin. However, in future experiments, it would be useful to give a test option "None of these objects," rather than forcing participants to choose one of the studied objects.

Our model adaptively allocates attention trial-by-trial to pairings on the basis of both entropy and prior knowledge. Built upon a simple associative mechanism, this process model captures the dynamic feedback loop between attention and learning: Internal learning states drive attention to certain pairs, and attention on these pairs, in turn, strengthens associations between those pairs (leaving unattended pairs relatively weak), updating internal learning states that will again drive attention in subsequent learning. Other proposed models of word learning, including the Frank, Goodman, and Tenenbaum (2009) Bayesian model and the Yu (2008) machine translation model, are batch learners and are unaffected by trial order, which has been shown to affect cross-situational word learning (Kachergis, Yu, & Shiffrin, 2009b).

Thus, the contribution of the present model is to incorporate two attention mechanisms—biases for prior knowledge and uncertainty—and show how they jointly control statistical learners' attention in real-time learning. We note that these biases are also present in infants, who show a familiarity preference after brief exposure to a stimulus but a novelty preference after longer exposure (Hunter & Ames, 1988). These factors cause our model to show a strong, early ME bias—consistent with children's ability to fast-map (Markman & Wachtel, 1988)—but allow this bias to gradually relax as additional evidence accumulates. The model therefore displays biases claimed to be important mechanisms for language acquisition (e.g., Golinkoff et al., 1992) by formalizing the competition between attending to familiar associations and attending to stimuli with uncertain (i.e., high-entropy) associates. Thus, we have demonstrated that an associative process model with attention can successfully explain how early adaptive biases may arise from simple mechanisms and still yield general learning in the long run, as do human learners. This approach attributes developmental changes in word learning to general cognitive mechanisms (a view shared by others—e.g., Yu & Smith, 2011). The model works reasonably well for the present data, and future research will extend it to other language tasks and associative learning effects.

Acknowledgments This article is an extended and updated version of a paper that appeared in the *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*. This research was supported by National Institute of Health Grant R01HD056029 and National Science Foundation Grant BCS 0544995. Special thanks to Tarun Gangwani for data collection, to Gregory E. Cox for useful discussions, and to Hedderik van Rijn and two anonymous reviewers for helpful comments.

References

Broeder, P., & Murre, J. M. J. (Eds.). (2002). *Models of language acquisition: Inductive and deductive approaches*. Oxford: Oxford University Press.

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

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 1–55.

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

Hunter, M., & Ames, E. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. Libsitt (Eds.), *Advances in infancy research* (Vol. 5, pp. 69–95). Stamford, CT: Ablex.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009a). Frequency and contextual diversity effects in cross-situational word learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of 31st Annual Meeting of the Cognitive Science Society* (pp. 755–760). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009b). Temporal contiguity in cross-situational word learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of 31st Annual Meeting of the Cognitive Science Society* (pp. 756–761). Austin, TX: Cognitive Science Society.

Klein, K. A., Yu, C., & Shiffrin, R. M. (2008). Prior knowledge bootstraps cross-situational learning. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1930–1935). Austin, TX: Cognitive Science Society.

Markman, E. M. (1990). Constraints children place on word learning. *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.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*(3/4), 1–129.

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

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development, 4*, 32–62.

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.

Supplemental materials

An Appendix with additional modeling details, as well as code for baseline models, figures, and fitting procedures, may be downloaded from http://pbr.psychonomic-journals.org/content/supplemental.