Précis of *Mechanisms for Cross-Situational Learning of Word-Referent Mappings: Empirical and Modeling Evidence*

George Kachergis

Introduction

Language acquisition is a ubiquitous, challenging problem involving fundamental cognitive abilities of attention, learning, and memory. From infants to adult travelers, language learners are faced with figuring out which words refer to which referents from situations that contain many words and referents. By remembering the words and referents (e.g., objects) that co-occur most frequently over time, learners may acquire correct noun meanings—an idea known as cross-situational learning (Gleitman, 1990). However, given that in most situations there are many possible word-object pairings but time and attention are limited, learners likely use strategies to restrict the number of meanings they consider. Building on associative learning research on attention (e.g., Kruschke, 2001), my thesis models noun acquisition using a simple learning mechanism with biases. I present and model a number of word-learning experiments that investigate both theorized learning constraints (e.g., mutual exclusivity: the assumption that each word maps to one object, and vice-versa) and factors such as word frequency, contextual diversity, and active learning, all of which occur naturally. Besides being of interest to psychologists and linguists, the word-object mapping problem is related to machine translation (language-language mappings) in computer science, and more generally to inference from contingency tables (for an overview, see Agresti, 1992). Models of word learning beg application in education, where they may be used to predict optimal training sequences and even identify individual differences in terms of underlying psychological abilities.

In adult cross-situational learning studies (e.g., Yu & Smith, 2007; Kachergis, Yu, & Shiffrin, 2009a, 2009b; Suanda & Namy, 2012), participants are asked to learn the referent of novel words by watching a series of training trials. On each trial learners see an array of unfamiliar objects (e.g., four sculptures) and hear pseudowords (e.g., *stigson, bosa*). The referent of each pseudoword is ambiguous on a given trial, because although each word refers to a single onscreen object, the intended referent is not indicated. In a typical learning scenario, participants attempt to learn 18 word-object pairings from 27 trials, with four words and four objects given per trial. In this design, each word-referent pair is presented six times over the five-minute training period. Learning a correct word-object pairing requires accumulating word-object co-occurrences in some fashion. Cross-situational learning has also been observed in infants (Smith & Yu, 2008; Vouloumanos & Werker, 2009) and toddlers (Akhtar & Montague, 1990). Research shows that even simple cross-situational learning models can learn large lexicons in a reasonable amount of time, both in simulated corpora (Blythe, Smith, & Smith, 2010) and from a corpus of parent-child interactions (Frank, Goodman, & Tenenbaum, 2009).

How might people accomplish cross-situational learning? One approach views word learning as a problem of induction with an enormous hypothesis space, and proposes a number of language-specific constraints to restrict the space (Markman, 1992). In this view, infants generate hypotheses that are consistent with this set of constraints and principles. For example, the global principle (or bias) of mutual exclusivity (ME) assumes that every object has only one name (Markman & Wachtel, 1988). At a lower level, Clark (1987) proposed the fill-the-lexical gap bias, which causes children to want to find a name for an object with no known name, a theory advanced by Merriman and Bowman (1989). When given a set of familiar and unfamiliar objects, it has been shown that 28-month-olds assume that a new label maps to an unfamiliar object (e.g., Mervis & Bertrand, 1994). Similarly, the principle of contrast states that an infant given a new word will seek to attach it to an unlabeled object (Clark, 1987). Fill-the-gap, ME, and contrast make many of the same predictions made by the more general novel name-nameless category principle (N3C), which states that novel labels map to novel objects (Golinkoff, Mervis, & Hirsh-Pasek, 1994). In order to be of aid to infant learners, such principles are thought to be either innate or developed very early in life (Markman, 1992).

From another view, domain-general mechanisms like those involved in associative learning are looked to before positing language-specific constraints (Smith, 2000; Kachergis, 2012). Associative learning paradigms typically present one or more perceptual cues (e.g., objects, sounds), learners make a response (e.g., a button press), and feedback is given (e.g., food, a shock). Language acquisition seems likely to be subject to the same memory, attention, and learning mechanisms that are used to explain associative learning behavior. For example, attention is a useful construct in explaining several associative learning phenomena. When one cue q_1 is paired with outcome o on each trial, the resulting q_1-o association is stronger than q_1-o when two simultaneous cues $\{q_1, q_2\}$ predict o during training; thus, q_2 is said to overshadow q_1 (Pavlov, 1927). A reasonable way to explain overshadowing is that attention is split between the two cues, and thus the associations q_1-o and q_2-o grow more slowly than when q_1 appears alone. Attention is also used to explain the blocking effect (Kamin, 1968), which can be induced using a design with two training stages. In the early stage, cue q_1 is repeatedly paired with outcome o, and in the late stage q_1 and q_2 appear jointly preceding o. The association between q_2 and o is found to be much weaker than when only the late stage occurs. Thus q_2 has been blocked by q_1's earlier association with o–with much the same effect as a mutual exclusivity bias preventing learners from mapping a second label (q_2) to a known object (o). Learning models, updating knowledge trial-to-trial, account for blocking using selective attention to q_1: since q_1 already predicts o, there is no need to strengthen q_2-o (e.g., Rescorla & Wagner, 1972; Pearce & Hall, 1980). Kachergis (2012) shows that highlighting, another associative learning effect explained with attention (Kruschke, 2003; Kruschke, 2009), is also observed in a word-learning context, and can be accounted for using an associative word-learning model with attentional biases. This work is explained in detail in Chapter 2 of my thesis.

Vlach (2012) links word-learning to memory, showing that fast-mapped words show the same pattern of forgetting as other memorized items. Moreover, recent work has found that children show a 1-to-1 bias in domains other than language: voices to faces (Moher, Feigenson, & Halberda, 2010) and actions to objects (Childers & Tomasello, 2003), suggesting that ME–whatever its explanation–may not be a language-specific constraint. Finally, both infants (Carey & Bartlett, 1978) and dogs (Pilley & Reid, 2011) have been shown to fast map: given a new word, they will choose a new object over an object with a known label, retaining the mapping weeks later (see Bloom, 2000). Thus, fast mapping, a valuable tool for word learning which is thought to rely upon language-specific principles such as the above, may instead be based on more general abilities. In this thesis, I will explore various real-world factors that influence word-learning, and show that domain-general mechanisms are sufficient to explain the bulk of the findings without necessitating language-specific constraints.

Overview of Thesis Studies

Using a cross-situational word learning experiment, Chapter 2 investigates the double-edged blade of mutual exclusivity, which can be leveraged to quickly infer novel word-object pairs if old pairs are present, but which would, strictly applied, block the learning of non-1-to-1 mappings such as synonyms and homonyms. The results show how the ME bias changes as a function of the strength of prior knowledge, and that it relaxes in the face of greater evidence that mappings are not 1-to-1. I introduce an associative model that accounts for the empirical results using competing familiarity and uncertainty biases–biases shown even by infants (for an overview, see Hunter & Ames, 1988). The model learns trial-by-trial, word-by-word, distributing attention (i.e., associative weight) among the objects that are present. The amount of attention given to each word-object association on a trial is determined by the current strength of that association, as well as the relative entropy of a word/object's associates: both strong associations and stimuli with uncertain associates (e.g., novel stimuli) demand more attention. Baseline models that operate solely on the familiarity bias or solely on the uncertainty bias are provided to understand how the biases compete to produce the inference-like behavior that is consistent with mutual exclusivity. This experiment was previously published in Kachergis, Yu, and Shiffrin (2010), and later modeled in Kachergis, Yu, and Shiffrin (2012b).

The proposed model assumes that learners do not attend equally to all possible word-object pairings (i.e., store all co-occurrences). The model will naturally show a mutual exclusivity bias via competing selective attention based on two factors: strengthening associations between words and objects that have co-occurred previously, as well as attending to stimuli that have no strong associates (e.g., novel stimuli). These competing familiarity and uncertainty biases allow the model to exhibit fast mapping, since a novel word-novel object combination will demand more attention, and a novel word will only become weakly associated with an already-known referent (Kachergis et al., 2012b). For example, suppose word w_1 and object o_1 have appeared together and are thus somewhat associated, while w_7 and o_7 are novel. Given a trial with both pairs: $\{w_1,o_1,w_7,o_7\}$, w_1-o_1 demands more attention than w_7-o_1, w_1-o_7, or w_7-o_7, since w_1-o_1 is stronger than baseline. However, attention is also pulled individually to w_7 and to o_7, since both of these novel stimuli have no strong associates. Uncertainty is measured by the entropy of each stimulus' association strengths. Because of the high joint uncertainty of w_7 and o_7, more attention is given to the association w_7-o_7. Thus, attention is mostly divided between w_1-o_1 and w_7-o_7, although the other pairings will be strengthened a bit.

Formally, given n words and n objects to be learned over a series of trials, let M be an n word $\times$ 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 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 (e.g., 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, strength is allocated 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:

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

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

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

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 ($H(\cdot)$; roughly novelty) and prior knowledge

($M_{w,o}$; familiarity). As λ increases, the weight of uncertainty (i.e., the exponentiated entropy term, which includes both the word and 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 the model chooses an associated referent for each word from the m alternatives in proportion to their strengths to the word.

The new associative model word-learning is contrasted with the classic Rescorla and Wagner (1972) model of associative learning, showing that the prediction-based error correction mechanism of the latter does not match my empirical data. I discuss how word learning can be thought of as classic associative learning with multiple cues (i.e., objects) and outcomes (words). Grounding the study of language acquisition in associative learning mechanisms by adapting and applying these models to word-learning bridges two areas of research that are often thought of separately. I also compare the proposed model to other recent models of cross-situational word learning, showing that this model produces both inference-like behavior typically exhibited by rule- and logic-based models (e.g., Siskind, 1996), as well as trial-order effects that humans demonstrate, but which some models cannot show (e.g., Frank et al., 2009). Finally, Chapter 2 includes an experiment demonstrating that an attentional associative learning effect also occurs in a cross-situational word-learning context, further suggesting that domain-general mechanisms may be sufficient to explain a variety of language learning behaviors. This experiment was published and modeled in Kachergis (2012). Despite its success at explaining away some language-specific constraints, a domain-general associative account is not satisfactory to all language acquisition researchers.

Another proposed process for word learning is hypothesis testing, which assumes that learners eliminate incorrect word-referent mappings from the hypothesis space using a combination of observations and logical constraints. This approach is used in the formal analysis of language acquisition (e.g., Gold, 1967; Pinker, 1979), and can be seen to stem from logic-based approaches to human concept learning (Bruner, Goodnow, & Austin, 1956) and a long line of inferential methods in the philosophy of science. Many developmental theories of language acquisition are built upon a rationale of hypothesis testing (e.g., Carey, 1978; Clark, 1987). One intuition that seems common among these approaches is that the world and perhaps the language environment are far too complex (cf. Quine, 1960) for learners to be able to store, track, and update a multitude of associations between words and referents (e.g., Medina, Snedeker, Trueswell, & Gleitman, 2011). In Chapter 3, I address the question of whether a simple hypothesis-based model built upon the assumptions of Medina et al. (2011) can account for individual word-learning trajectories as well as graded associations. I measured word-learning trajectories over four blocks of the same cross-situational training. Individuals' trajectories from this experiment were used to investigate the flexibility of two models: the associative model proposed in Chapter 2, and a verbal model proposed by Medina et al. (2011) that posits people store only a single hypothesized object for each word, and do not remember other possible associates. I found that the model derived from the assumptions in Medina et al. (2011) is not capable of creating the range of learning trajectories shown by individuals, whereas the associative model is. This work was published in Kachergis, Yu, and Shiffrin (2012a).

Chapter 4 examines and models the effects of frequency and contextual diversity in cross-situational word learning. Word frequency is known to vary in the real world (Zipf, 1949), and children typically acquire higher frequency words earlier (Huttenlocher, Haight, Bryk, Seltzeer, & Lyons, 1991). I investigate the effects of allowing some word-object pairs to appear more often than others. Results show that frequent word-referent pairs are often–but not always–learned better, and also boost learning of low frequency pairs. This superior learning for vocabularies with heterogeneous frequency may result from learning frequent pairs first, and leveraging this knowledge in later trials to learn low frequency pairs. However, contextual diversity – the number of other pairs a given pair appears with – is naturally confounded with frequency, and presents an alternative explanation (see also Hills, Maouene, Riordan, & Smith, 2010). I manipulate three critical factors in cross-situational learning: frequency, contextual diversity, and within-trial ambiguity (i.e., the number of pairs per trial). While greater frequency and contextual diversity are separately correlated with performance, they also interact. Specifically, when infrequent pairs are given greater contextual diversity, especially mixed with high frequency pairs, learning increases for these pairs, with little or no detriment to the high frequency pairs. The highest learning performance observed was in conditions with varied frequency and high contextual diversity, in which learners likely used the high-frequency pairs, learned early in training, to quickly boostrap the low frequency pairs late in training. The familiarity- and uncertainty-biased associative model naturally produces such bootstrapping via an attention shift to the infrequent (uncertain) stimuli once the frequent (familiar) pairs are well-known. Once again, inference-like behavior is accomplished using simple biases that use only the current state of knowledge. I also show that the associative model outperforms the recent incremental probabilistic model (Fazly, Alishahi, & Stevenson, 2010b; Fazly, Ahmadi-Fakhr, Alishahi, & Stevenson, 2010a, 2008). Moreover, the associative model grants additional insight into the timecourse of learning, where we see how early learning of high-frequency pairs bootstraps late learning of low-frequency pairs. The experiments in Chapter 4 were published in Kachergis et al. (2009a), and a manuscript with the modeling portion is in preparation.

Chapter 5 relaxes the assumption that learners are passive observers, simply absorbing whatever stimuli are provided. Learners in the world are not completely passive, but can affect how their environment is structured by moving their heads, eyes, and nearby objects. These actions can indicate attention to a language teacher, who may then be more likely to name the attended objects. Thus, I allow learners to actively choose which objects they would like to see named on the next trial. Learners control when to repeat pairs, when to stop experiencing pairs that they are confident they already know, and when to attempt to learn more pairs. This gives us a glimpse of their preferred strategies and rate of learning, and can be quite diagnostic for model selection (e.g., Kruschke, 2008). For example, since the model I propose assumes learners have access to prior knowledge as well as uncertainty, it predicts that active learners will perform better than passive learners. Finding that active learning produces superior performance, I investigate the types and range of strategies used by learners, and try to infer system constraints (e.g., working memory) as well as individual differences. I propose and test a working memory extension to the familiarity- and uncertainty-biased associative model that is used throughout this thesis.

I found that all learners preferred to repeat at least one pair from trial-to-trial, on average. Repeating one pair allows the learner to infer that the repeated object goes with the repeated word using working memory, and also reduces the number of associations to be considered among unrepeated stimuli. However, one group of learners preferred to repeat more than one pair, on average, indicative of a more complex set of strategies. Overall, this group showed higher learning than the cluster of learners repeating only a single pair. Extending the associative model with working memory mechanism, I found individual differences in attention to repeated/unrepeated items that may indicate a variety of strategies deserving of further study.

Chapter 6 summarizes the work presented here, situates it in the broader context of language acquisition, and discusses its limitations and possible future extensions. Overall, in this thesis I argued that a wide range of human cross-situational learning behaviors can be captured by a domain-general associative model, without recourse to language-specific principles or constraints.

Conclusion

This thesis proposed a simple associative model for learning the intended word-object mappings from a series of scenes, each of which contain multiple words and objects. The proposed model uses domain-general biases for uncertainty and familiarity, which are two fundamental psychological aspects of stimuli that have been previously implicated in directing infant and adult attention. I have linked word-learning to associative learning, showing that an attentional associative learning effect occurs in a word-learning context. I have shown that the proposed associative model is better than a single-hypothesis model at accounting for individual learning trajectories, and can naturally account for the effects and interactions of frequency and contextual diversity better than other current models. Finally, using the proposed model I investigated and evaluated what strategies people use in active cross-situational learning, finding that immediately repeating more than one pair is beneficial. In summary, I have proposed a model of cross-situational learning that accounts for a wide variety of human word-learning behaviors, including mutual exclusivity and semantic bootstrapping and other results that stem from varying factors known to vary in the natural language environment. This simple, intuitive model is both robust and extensible, and the explanations and predictions it offers may prove useful in language acquisition research as well as education.

The proposed model adaptively allocates attention trial-by-trial to pairings based on both entropy (i.e., uncertainty) 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 which will again drive attention in subsequent learning. Other proposed models of word learning, including the Frank et al. (2009) Bayesian model and the (Yu, 2008) machine translation model are batch learners: they are unaffected by trial order, which has been shown to affect cross-situational word learning (e.g., Kachergis et al., 2009b).

Thus, the contribution of the proposed 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. Note that these biases are also present in infants, who show a familiarity preference after brief exposure to a stimulus, but a preference for novel stimuli after longer exposure (for an overview, see Hunter & Ames, 1988). A recent infant study provides evidence that novelty plays an important role in early word learning (Mather & Plunkett, 2012). 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 for non 1-to-1 mappings accumulates. The model therefore displays biases claimed to be important mechanisms for language acquisition (e.g., Golinkoff, Hirsh-Pasek, Bailey, & Wegner, 1992) by formalizing the competition between attending to familiar associations and attending to stimuli with uncertain (i.e., high entropy) associates. Thus, I 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., Smith (2000)). The success of the model suggests a learning system that does not learn independent associations between individual words and referents, but one that rather learns a system of associations. In such a system, a single word-referent pairing is correlated with all the other pairings that share the same word and all the other pairings that share the same referent, which are in turn correlated with more word-referent pairs–an entire system of them. I contend that the improvement in statistical word learning is in part due to the recruitment of accumulated latent lexical knowledge, used to learn subsequently appearing pairs. The associative model, which learns associations between all co-occurring words and objects incrementally, leverages prior knowledge and uncertainty to accomplish this, and also accounts for data from associative learning experiments.

By linking word learning to associative learning, as suggested by Smith (2000), we may find that the plurality of overlapping language-specific constraints (e.g., ME, N3C, the principle of contrast, and the fill-the-gap bias) are unnecessary to explain many language learning behaviors. Instead, I predict that a more parsimonious explanation will emerge, built upon a foundation of domain-general mechanisms formalized in computational models. Language-specific principles and constraints may yet exist, but we should first see how far more universal mechanisms take us.

The contributions of this thesis are interdisciplinary, for I address theoretical debates in linguistics and psychology using models that are of interest to computer science and that have application in education. The diverse techniques used in this thesis–from psychology experiments to computational modeling–are used to address theoretical issues relevant to many cognitive scientists. I believe my thesis is therefore a good candidate for the Robert J. Glushko Dissertation Prize. Thank you for your consideration.

References

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, *7*(1), 131–153.

Akhtar, N., & Montague, L. (1990). Early lexical acquisition: the role of cross–situational learning. *First Language*, *19*, 34–358.

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Blythe, R. A., Smith, K., & Smith, A. D. M. (2010, January). Learning Times for Large Lexicons Through Cross-Situational Learning. *Cognitive Science*, *34*(4), 620–642.

Bruner, J. A., Goodnow, J. S., & Austin, G. J. (1956). *A study of thinking*. New York, NY: Wiley.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Report on Child Language Development*, *15*, 17–29.

Childers, J. B., & Tomasello, M. (2003). Children extend both words and non-verbal actions to novel exemplars. *Developmental Science*, *6*(2), 185–190.

Clark, E. V. (1987). The principle of contrast: a constraint on language acquisition. In B. MacWhinney (Ed.), (1-33). Hillsdale, NJ: Erlbaum.

Fazly, A, Alishahi, A, & Stevenson, S. (2008). A probabilistic incremental model of word learning in the presence of referential uncertainty. In *Proceedings of the 30th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Fazly, A, Ahmadi-Fakhr, F, Alishahi, A, & Stevenson, S. (2010a). Cross-situational learning of low frequency words: The role of context familiarity and age of exposure. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 2362–2367). Austin, TX: Cognitive Science Society.

Fazly, A., Alishahi, A., & Stevenson, S. (2010b, May). A Probabilistic Computational Model of Cross-Situational Word Learning. *Cognitive Science*, *34*(6), 1017–1063.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009, May). Using Speakers' Referential Intentions to Model Early Cross-Situational Word Learning. *Psychological Science*, *20*(5), 578–585.

Gleitman, L. (1990). The structural sources of word meaning. *Language Acquisition*, *1*, 3–55.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *16*, 447–474.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994, February). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language*, *21*(1), 125–155.

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*(1), 99–108.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010, October). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273.

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).

Huttenlocher, J., Haight, W., Bryk, A., Seltzeer, M., & Lyons, T. (1991). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Developmental Psychology*, *27*(2), 236–248.

Kachergis, G. (2012). Learning nouns with domain-general associative learning mechanisms. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 533–538). Austin, TX: Cognitive Science Society.

Kachergis, G, Yu, C, & Shiffrin, R. M. (2010). Adaptive constraints and inference in cross-situational word learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 2464–2469). Austin, TX: Cognitive Science Society.

Kachergis, G, Yu, C, & Shiffrin, R. M. (2012a). Cross-situational word learning is better modeled by associations than hypotheses. In *Ieee conference on development and learning-epirob 2012*.

Kachergis, G, Yu, C, & Shiffrin, R. M. (2009a). Frequency and contextual diversity effects in cross-situational word learning. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual meeting of the cognitive science society* (pp. 2220–2225). Austin, TX: Cognitive Science Society.

Kachergis, G, Yu, C, & Shiffrin, R. (2009b). Temporal contiguity in cross-situational statistical learning. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual meeting of the cognitive science society* (pp. 1704–1709). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012b). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic Bulletin and Review*, *19*(2), 317–324.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior, 1967: aversive stimulation* (pp. 9–31). Coral Gables, FL: University of Miami Press.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, 171–175.

Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning & Behavior*, *20*, 121–157.

Kruschke, J. K. (2009). *Highlighting: A canonical experiment*. Psychology of Learning and Motivation. Elsevier.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.

Markman, E. M. (1992). Constraints on word learning: speculations about their nature, origins and domain specificity. In M. R. Gunnar & M. P. Maratsos (Eds.), *Modularity and constraints in language and cognition: the minnesota symposium on child psychology* (pp. 59–101). Hillsdale, NJ: Erlbaum.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*(7), 1157–1177.

Medina, T., Snedeker, J, Trueswell, J., & Gleitman, L. (2011, May). How words can and cannot be learned by observation. *PNAS*, 1–6.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. In *Monographs of the society for research in child development* (Vol. 54, *3*, pp. 1–129).

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name - nameless category (n3c) principle. *Child Development*, *65*, 1646–1662.

Moher, M., Feigenson, L., & Halberda, J. (2010). A one-to-one bias and fast mapping support preschoolers learning about faces and voices. *Cognitive Science*, 1–33.

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.

Pearce, J., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectivenessof conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552.

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*, 184–195.

Pinker, S. (1979). Formal models of language learning. *Cognition*, *1*, 217–283.

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

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. New York, NY: Appleton Century Crofts.

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

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

Smith, L. B. (2000). How to learn words: an associative crane. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Breaking the word learning barrier* (pp. 51–80). Oxford: Oxford University Press.

Suanda, S. H., & Namy, L. L. (2012, April). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science: A Multidisciplinary Journal*, *36*(3), 545–559.

Vlach, H. A. (2012, February). Fast mapping across time: memory processes support children's retention of learned words, 1–8.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–1617.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development*, *4*(1), 32–62.

Yu, C, & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420.

Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.

George Kachergis

Leiden University
Department of Psychology
Leiden, the Netherlands

Email: george.kachergis@gmail.com
Homepage: http://www.kachergis.com

Education

- Postdoctoral Researcher, Leiden University, Advisor: Bernhard Hommel, RoboHow Project, *January 1, 2013–present.*
- Ph.D. (December 15, 2012), Indiana University, Department of Psychological and Brain Sciences and Cognitive Science Program. Advisory Committee: Richard M. Shiffrin, Michael N. Jones, John Kruschke, Robert Goldstone, Chen Yu. GPA 3.98/4.
- B.A. Cognitive Studies and Computer Science, Carleton College, 2007. *Magna cum laude* with distinction in Cognitive Studies.
- Probabilistic Models of Cognition Summer School, Institute for Pure and Applied Mathematics 2011.
- Budapest Semester in Cognitive Science, Eötvös Loránd Science University, Fall 2005.

Publications

Journal Articles

- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (in prep). *A Bootstrapping Model of Frequency and Contextual Diversity Effects in Word Learning.*
- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (in press). *Actively Learning Object Names Across Ambiguous Situations.* Topics in Cognitive Science.
- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (in prep). *Temporal Contiguity Improves Cross-Situational Word Learning.*
- Trueblood, J. S., **Kachergis, G.**, and Kruschke, J. K. (in prep). *A Cue Imputation Bayesian Model of Information Aggregation.*
- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (2012). *An Associative Model of Adaptive Inference for Learning Word-Referent Mappings.* Psychonomic Bulletin & Review, *19*(2), 317-324.
- Cox, G., **Kachergis, G.**, Recchia, G., and Jones, M. N. (2011). *Towards a Scalable Holographic Word-form Representation.* Behavior Research Methods, *43*(3), 602-615.

Refereed Conference Proceedings

- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (2012). Cross-situational Word Learning is Better Modeled by Associations than Hypotheses. *IEEE Conference on Development and Learning-EpiRob 2012.* **Best Experiment Combined with Computational Model**

- **Kachergis, G.** (2012). *Learning Words with Domain-General Associative Learning Mechanisms.* In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 533-538). Austin, TX: Cognitive Science Society.

- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (2012). Actively Learning Nouns Across Ambiguous Situations. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 527-532). Austin, TX: Cognitive Science Society. **Marr Prize Winner**

- Hendrickson, A. T. **Kachergis, G.**, Fausey, C., and Goldstone, R. L. (2012). Re-learning labeled categories reveals structured representations. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1668-1673). Austin, TX: Cognitive Science Society.

- Cox, G. E., **Kachergis, G.**, and Shiffrin, R. M. (2012). Gaussian Process Regression for Trajectory Analysis. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1440-1445). Austin, TX: Cognitive Science Society.

- **Kachergis, G.**, Recchia, G., and Shiffrin, R. M. (2011). Adaptive Magnitude and Valence Biases in a Dynamic Memory Task. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 819-824). Austin, TX: Cognitive Science Society.

- **Kachergis, G.**, Cox, G. E., and Jones, M. N. (2011). OrBEAGLE: Integrating Orthography into a Holographic Model of the Lexicon. *21st Annual International Conference on Artificial Neural Networks*, Espoo, Finland. *June 2011.*

- Trueblood, J. S., **Kachergis, G.**, and Kruschke, J. K. (2011). A Cue Imputation Bayesian Model of Information Aggregation. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1298-1303). Austin, TX: Cognitive Science Society.

- Gangwani, T., **Kachergis, G.**, and Yu, C. (2010). Simultaneous Cross-situational Learning of Category and Object Names. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1595-1600). Austin, TX: Cognitive Science Society.

- Hendrickson, A. T., **Kachergis, G.**, Gureckis, T. M., and Goldstone, R. L. (2010). Is categorical perception really verbally mediated perception? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings* *of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1216-1221). Austin, TX: Cognitive Science Society.

- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (2010). Cross-Situational Statistical Learning: Implicit or Intentional? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2362-2367). Austin, TX: Cognitive Science Society.
- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (2010). Adaptive Constraints and Inference in Cross-Situational Word Learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2464-2469). Austin, TX: Cognitive Science Society.
- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (2009). Frequency and Contextual Diversity Effects in Cross-Situational Word Learning. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2220-2225). Austin, TX: Cognitive Science Society.
- **Kachergis, G.**, Yu, C., and Shiffrin, R. M. (2009). Temporal Contiguity in Cross-Situational Statistical Learning. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1704-1709). Austin, TX: Cognitive Science Society.
- Barbella, D., **Kachergis, G.**, Liben-Nowell, D., Sallstrom, A., and Sowell, B. Depth of Field and Cautious-Greedy Routing in Social Networks. *18th International Symposium on Algorithms and Computation*, Sendai, Japan. *17 December 2007.*

Selected Conference Presentations

- Brown, S., **Kachergis, G.**, Donkin, C., Heathcote, A., Rae, B. *How do people make fast decisions?*. 45th Annual Meeting of the Society for Mathematical Psychology, Columbus, OH. *July 21, 2012.*

  Kachergis, G. (2012). *Learning Words with Domain-General Associative Learning Mechanisms.* 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan. *August 2012.*

  Kachergis, G., Yu, C., and Shiffrin, R. M. (2012). *Actively Learning Nouns Across Ambiguous Situations.* 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan. *August 2012.*
- **Kachergis, G.**, Cox, G. E., Shiffrin, R. M. *Dynamic Effects of Perceptual and Categorical Similarity on Recognition Memory.* Poster presented at the Context and Episodic Memory Symposium, Bloomington, IN. *May 10, 2012.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *Modeling Cross-situational Word Learning: Hypotheses or Associations?.* Midwest Cognitive Science Conference, Bloomington, IN. *May 7, 2012.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *Active Cross-situational Statistical Word Learning.* Poster presented at the 33rd Annual Meeting of the Cognitive Science Society, Boston, MA. *July 2011.*

- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *An Associative Model of Inference in Statistical Word Learning.* 44th Annual Meeting of the Society for Mathematical Psychology, Boston, MA. *July 2011.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *Modeling the Acquisition of the Mental Lexicon.* 43rd Annual Meeting of the Society for Mathematical Psychology, Portland, OR. *August 2010.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *Modeling Frequency and Context Effects in Statistical Word Learning.* 42nd Annual Meeting of the Society for Mathematical Psychology, Amsterdam, the Netherlands. *August 2009.*
- Gangwani, T., **Kachergis, G.**, Yu, C. *Simultaneous Noun and Category Learning via Cross-Situational Statistics.* 31st Annual Meeting of the Cognitive Science Society, Amsterdam, the Netherlands. *July 2009.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *The Semantics of Eye Movements in Cross-Situational Statistical Word Learning.* 41st Annual Meeting of the Society for Mathematical Psychology, Washington, D.C. *27 July 2008.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *Temporal Continuity in Cross-Situational Statistical Learning.* 30th Annual Meeting of the Cognitive Science Society, Washington, D.C. *25 July 2008.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *The Automaticity of Statistical Word Learning.* 30th Annual Meeting of the Cognitive Science Society, Washington, D.C. *25 July 2008.*
- **Kachergis, G.**, Shiffrin, R. M. *The Effects of Repeated Sequential Context on Recognition.* 7th Annual Summer Interdisciplinary Conference, Madonna di Campiglio, Italy. *10 July 2008.*
- **Kachergis, G.**, Yu, C., Shiffrin, R. M. *The Automaticity of Cross-Situational Statistical Learning.* Redhawk Mental Life, Oxford, OH. *8 March 2008.*
- Ohnesorge, C., **Kachergis, G.** *Testing the Whorfian Hypothesis: Lateralized Presentation and Color Recognition.* 48th Annual Meeting of the Psychonomic Society, Long Beach, CA. *17 November 2007.*
- **Kachergis, G.**, Ohnesorge, C. *Whorf hypothesis not supported at a perceptual level.* MidBrains, Macalester College, St. Paul, MN. *28 April 2007.*
- Ohnesorge, C., Fanta, A., **Kachergis, G.** *Color Recognition and Laterality: Does Language Affect Color Processing?* 6th Annual Summer Interdisciplinary Conference, Kalymnos, Greece. *28 June 2007.*
- **Kachergis, G.**, Olson, J. F. *Lexical Variety in Human Processing of Syntactically Ambiguous Sentences.* Minnesota Undergraduate Psychology Conference, University of St. Thomas, St. Paul, MN. *22 April 2006.*

Teaching, Supervision, and Departmental Service

- *Inverting the Classroom*, Pedagogy Seminar Talk, Indiana University *Spring, 2012*
- Teaching Assistant for Human Learning and Cognition, *Spring, 2012*
- Guest lecture presented in Introduction to Psychology, Instr. Hilary Kalagher, *Spring, 2009*
- Coordinator for Indiana University Cognitive Lunch talk series, *Fall 2010 to Fall 2011*
- Instructor for P211 Experimental Methods in Psychology, *Spring, 2010*
- Grader for Artificial Intelligence, Prof. David Musicant, Carleton College *Winter, 2007*
- Prefect for Data Structures, Prof. David Liben-Nowell, Carleton College *Winter, 2006*
- Supervised student research assistants: J. Booher, E. Farmer, T. Gangwani, Z. Horwitz, K. Mullen, E. Lee, P. LaFree, A. Salisbury, B. Jenkins

Awards

- **Best Paper: Experiment Combined with Computational Model, International Conference on Development and Learning-EpiRob**, *2012*
- **Marr Prize for Best Student Paper, Cognitive Science Society**, *2012*
- **National Science Foundation East Asia & Pacific Summer Institute / Japan Society for the Promotion of Science** (OISE-1209475; JAIST), *2012*
- **Robert J. Glushko and Pamela Samuelson Foundation Student Travel Grant**, Cognitive Science Society, *2012*
- **Indiana University Cognitive Science Supplemental Research Fellowship**, *2011*
- **William K. Estes Summer Research Award**, Indiana University Psychology Department, *2011*
- **European Neural Network Society Student Travel Award**, *2011*
- **Society for Mathematical Psychology Student Travel Award**, *2011*
- **Society for Mathematical Psychology Student Travel Award**, *2010*
- **Society for Mathematical Psychology Student Travel Award**, *2009*
- **Indiana University Cognitive Science Supplemental Research Fellowship**, *2009*

Invited Talks

- *Domain-General Mechanisms for Learning Word-Referent Mappings*, Leiden University, *October 10, 2012*
- *Domain-General Mechanisms for Learning Word-Referent Mappings*, Japan Advanced Institute for Science and Technology, *June 27, 2012*
- *How Children Learn Words*, Carleton College, *April 25, 2012*
- *Domain-General Mechanisms for Learning Word-Referent Mappings*, Syracuse University, *April 15, 2012*
- *Dynamic Effects of Perceptual and Categorical Similarity on Recognition Memory*, Syracuse University, *April 15, 2012*
- *Domain-General Mechanisms for Learning Word-Referent Mappings: Empirical and Modeling Evidence*, Cognitive Lunch, Indiana University, *April 4, 2012*
- *Actively Learning Nouns Across Ambiguous Situations Using Associative Mechanisms*, Stanford University, *February 21, 2012*
- *Modeling the Acquisition of the Mental Lexicon*, Cognitive Lunch, Indiana University, *April 28, 2010*

Membership and Professional Service

- **Ad hoc reviewer**: Cognition, Cognitive Science, Cognitive Science Conference, Journal of Mathematical Psychology, Language Learning, Psychological Review
- **Member** of Cognitive Science Society, *2005–Present*
- **Member** of The Society for Mathematical Psychology, *2008–Present*
- **Member** of Sigma Xi Scientific Research Society, *2007–Present*
- **Member** of Association for Computing Machinery, *2004–2007*

Last updated: January 15, 2013