George Kachergis
Statement of Research Interests

My research seeks to explain how we learn about and represent the world. I empirically examine how people learn word-object mappings (i.e., nouns) from a series of situations comprised of multiple words and objects, as an infant or a traveler may do. Using mathematical models of memory and associative learning, I seek to discover the mechanisms underlying human word-learning behavior. I also investigate how these labels generalize to other perceptually similar objects, and whether perception itself is warped by the use of labels. I plan to research how populations of neurons might learn representations that produce the observed behaviors. In this vein, collaborators and I have proposed and tested a letter-based encoding scheme for words in a holographic model trained on text. In another project, I am exploring how input topology and network structure affects the recognition, separation, and integration of multiple inputs in spiking neural networks. Finally, I have developed paradigms that track the moment-to-moment decision state in memory, learning, categorization, and decision-making tasks. Access to decision trajectories in these game-like tasks enabled me to uncover new phenomenon such as changes of mind in recognition memory, and will provide additional constraining data for cognitive process models.

Learning Word-Referent Mappings

Much of my work has focused on how people acquire word-referent mappings (i.e., nouns) from ambiguous environments that contain many words and referents. This problem, faced by both infants and travelers when learning a new language, involves fundamental mechanisms of learning, memory, and attention. Participants in my experiments attempt to learn novel word-referent mappings from a series of situations containing several words and objects. The correct mappings are ambiguous on any given trial: only by tracking words and objects that co-occur frequently can a participant learn the pairings. However, given that there are a large number of novel words and objects (e.g., 18 pairs) to be learned over a short period of time (<5 minutes), it is unlikely that learners can precisely track how many times each word co-occurs with each object. In one study, I found that learning in two implicit study conditions (memory and signal detection cover tasks) was significantly worse than when participants were explicitly told to learn the word-object mappings—although still above-chance (Kachergis, Yu, & Shiffrin, 2010b). Thus, in developing a process model for cross-situational word learning, I have sought to use cognitively and neurally plausible mechanisms, rather than a system with perfect identification and memory.

In my word-learning experiments, I have varied a number of factors that vary in real-world environments in order to measure how people's learning and memory systems leverage these regularities. For example, one study investigated the effects of temporal contiguity (i.e., allowing some pairings to appear in consecutive trials; Kachergis, Yu, & Shiffrin, 2009), as there is often autocorrelation in our environment. I found that when half the pairs at some point appear on consecutive trials, these pairs are learned more frequently than the discontiguous pairs. However, the discontiguous pairs in this condition were also learned more often than in a condition with no contiguities. That is, participants were able to bootstrap learning of the discontiguous items by utilizing the structure of the contiguous items to reduce the complexity of the trial ordering. In a second experiment that shuffled a set of trials—thus maintaining within-trial statistics—to have different degrees of trial-to-trial overlap, I found that performance does not monotonically increase with increased temporal contiguity. I accounted for this surprising result using an associative learning model with habituation (i.e., reduced attention to pairs that overlap many times). The intuition is that as more pairs are contiguous across trials, any discontiguous pairs begin to stand out.

In another series of experiments I investigated mutual exclusivity (ME), a bias to learn 1-to-1 word-object mappings that even infants show (Markman & Wachtel, 1988). The ME bias can speed learning when pairs actually are 1-to-1, but can clearly inhibit the learning of homonyms and synonyms. In a study on adults, I varied the strength of prior knowledge and found that even with little early training, learners will use ME to quickly infer newly-introduced pairings in late training. However, learners will adaptively relax ME—when given additional late training suggesting that non-1-to-1 pairings exist (Kachergis, Yu, & Shiffrin, 2010a). I used my systematic data to motivate and construct an associative model of cross-situational word learning that incorporates both a bias for strengthening already-strong (familiar) associations and a bias for giving more attention to stimuli with no strong associates (i.e., high uncertainty or entropy). This model produces both inference-like behavior like rule- and logic-based models demonstrate, as well as trial-order effects that people show (Kachergis, Yu, & Shiffrin, under review).

I have also examined the influence of varying pair frequency and contextual diversity—how many pairs a given pair co-occurs with across training. In a study covarying frequency and contextual diversity—which are typically confounded in the real world—I found that increasing each factor helps learning. However, even low frequency pairs are learned well if they co-occur late in training with the (presumably already-known) high frequency pairs (Kachergis, Yu, & Shiffrin, 2009a). Hence, learning is bootstrapped by making inferences based on previously acquired knowledge, and early learning of high frequency pairs can later bring learning of low frequency words to surprisingly high levels. This data is accounted for using the same associative model with competing familiarity and uncertainty biases that I use for the mutual exclusivity data (Kachergis, Yu, & Shiffrin, in preparation). In my thesis, I will compare this model to extant models of classic associative learning, and evaluate its ability to show effects such as highlighting, illusory correlation, and blocking. I contend that cross-situational word learning can be usefully construed as associative learning with multiple cues (objects) and outcomes (words). In the future, I look forward to further linking disparate domains such as word-learning and associative learning by using models with simple underlying mechanisms.

In other cross-situational experiments, I asked whether learners can simultaneously acquire 1-to-1 and 1-to-many word-object mappings (i.e., instance and category labels). On each trial, three words and two objects were presented. Unbeknownst to the subject, one of the words was a category label that consistently co-occurred with four objects, while the other words each corresponded to a single object. Participants learned both ad hoc categories (i.e., with no perceptual similarity between the objects) and natural categories (in which objects shared a within-category feature; Gangwani, Kachergis, & Yu, 2010). This evidence that learners can simultaneously apprehend multiple levels of structure in their environment has yet to be modeled.

I have also studied active cross-situational word learning, which moves beyond the passive tasks used above and allows learners to choose which items they would like to experience on the next trial. Thus, learners control when to repeat pairs, when to stop experiencing pairs they feel they know, and when to attempt to learn more pairs. This gives us a glimpse of their preferred strategies and rate of learning, which in turn can tell us what information and mechanisms they have at their disposal. Even when constrained to the same degree of ambiguity (i.e., pairs per trial) and maximum number of training trials, active learners turn out to be much better than passive learners. Individual strategies vary somewhat, but most successful learners repeat slightly more than one pair on each consecutive trial. This indicates that learners are using temporal contiguity—as manipulated above—to deduce correct mappings. Modeling of individual differences is underway, and will be included in my thesis.

Categorical Perception and Verbal Labels

Cognitive scientists have long pondered whether the way language carves up the world affects how we perceive the world. Does attaching a label ('rose') to a category of objects make those objects less perceptually similar to other groups of objects that have different labels? If dandelions were called roses, would we find them harder to discriminate visually? I have addressed these questions in a series of experiments using a 2-dimensional space of morphed faces split into two categories, and with two clusters of faces in each category. We found that visual discrimination improves for between-category judgments as participants learn the categories, and improves for within-category but between-cluster judgments, and does not improve for within-cluster judgments (Hendrickson, Kachergis, Gureckis, & Goldstone, 2010). However, we were unable to disrupt the between-category categorical perception effect using a verbal interference task. This suggests that categorical perception is not merely an online influence of verbal processing, but may involve lower-level changes to the perceptual system.

Hendrickson, Fausey, Goldstone and I are continuing to investigate the effects of verbal labels in a category-learning study in which participants learn to label arbitrary sets of blob-shapes, and are then given a transfer task with either new labels, new blobs, or the same blobs in new sets. Surprisingly, learning was hardest in the transfer conditions that put blobs into new sets, implying that had found a way to group these blobs according to a latent variable (e.g., the label), even though blobs were only ever seen individually. In a follow-up, the blob categories either initially align with some perceptual feature or do not, and in the transfer phase the blobs are assigned to new sets and a different feature becomes diagnostic. Participants who first learned that categories align with one feature have no trouble switching to a different diagnostic feature, but cannot switch to item-based categorization; perhaps they are searching for a diagnostic feature that is not there? We are currently modeling the results with an associative model that learns to selectively attend to different features.

Holographic Models of Semantics

Dictionaries define words in terms of other words; models of semantic memory seek to define syntactic and semantic relationships between words based on the words they appear with in large bodies of text. Although these text corpora typically lack any perceptual grounding, comparing models' representations to human experimental data can tell us whether we are proposing reasonable representations, based on semantic information alone. Holographic reduced representations use neurally-plausible operators (circular convolution and vector addition) over distributed representations to store information according to a scheme defined by the researcher. Thus, instead of learning an arbitrary representation as in other connectionist models, various schemes can be explicitly proposed, tested, and compared. I have worked on extending BEAGLE, a model of semantic memory and word order (Jones & Mewhort, 2007), to incorporate orthographic information about words. That is, instead of having a random representation for each word, colleagues and I created a principled scheme to construct letter-based representations for each word so that similar words (e.g., *cat* and *catch*) will have greater analytical similarity. Our encoding scheme corresponds well with human data in several different tasks (Kachergis, Cox, & Jones, 2011; Cox, Kachergis, Recchia, & Jones, 2011), and opens the door for further such representational extensions (e.g., phonology).

Computational Properties of Spiking Neural Networks

I have worked with Richard Veale on determining what network properties of spiking neural networks are important for processing input spike trains. We use the framework of liquid state machines (e.g., Maass & Markram, 2004), in which a large small-world random network of leaky integrate and fire neurons are used as a kernel into which inputs are injected, and then recognized using trained perceptrons. Generic problems that such cortical microcircuits should be able to solve include recognizing particular inputs, recognizing conjunctions of particular inputs, and determining the source of a particular input. We ask how performance on these tasks varies as a function of the location and separation of our different input spike trains, and also as a function of the connectivity of the liquid. Two papers are in preparation.

References

Cox, G. E., Kachergis, G., Recchia, G., & Jones, M. N. (2011). Towards a Scalable Holographic Word-form Representation. *Behavior Research Methods*.

Gangwani, T., Kachergis, G., & 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*. Austin, TX: Cognitive Science Society.

Hendrickson, A. T., Kachergis, G., Gureckis, T. M., & 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*. Austin, TX: Cognitive Science Society.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1-37.

Kachergis, G., Yu, C., & Shiffrin, R. M. (under review). An associative model of adaptive inference in cross-situational word learning.

Kachergis, G., Cox, G. E., & Jones, M. N. (2011). OrBEAGLE: Integrating Orthography into a Holographic Model of the Lexicon. *In Lecture Notes in Computer Science* (Springer). International Conference on Artificial Neural Networks. Espoo, Finland.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2010a). 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. 1210-1215). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2010b). 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. 1189-1194). Austin, TX: Cognitive Science Society.

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 the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009b). Temporal contiguity in cross-situational statistical 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.

Maass, W., and H. Markram. (2004). On the computational power of recurrent circuits of spiking neurons, *Journal of Computer and System Sciences*, *69*(4), 593-616.