BRETT D. ROADS | RESEARCH STATEMENT

Summary
My goal is to boost human learning and performance by developing and applying formal models of cognition. I am interested in producing software that enables individuals to learn and perform tasks efficiently and effortlessly. My approach draws on methods from machine learning and theories from cognitive science in order to construct robust psychological models that characterize the computational challenges faced by an individual attempting to complete a task. My research lies at the interface of human learning, machine learning, and computer-assisted decision making.

My research has predominantly focused on helping individuals categorize visual images. I have approached this objective from two perspectives: *decision support* and *efficient training*. Decision support enables expert-like levels of performance—without training—by exploiting ordinary but powerful human visual capabilities. Efficient training promotes the discovery of the visual features necessary to correctly categorize the images. Both approaches leverage a latent space representation of human-perceived similarity, which we refer to as a *psychological embedding*.

Psychological Embeddings
My research relies on theories of visual categorization that assume that visual images can be represented as points in a multidimensional psychological space. Embedding algorithms can be used to infer latent representations from human similarity judgments. While there are an infinite number of potential visual features, an embedding algorithm can be used to identify the subset of salient features that accurately model human-perceived similarity. Existing techniques for inferring embeddings from similarity judgments are either not constrained by existing cognitive theory or only weakly so. Using a generalized similarity kernel that encompasses a range of psychological theories, we developed an embedding algorithm that jointly infers embeddings and the underlying similarity kernel. Initial results suggest that a similarity kernel based on an exponential distribution and student-t distribution produce similar predictions of held-out data. However, we are still conducting experiments to better understand the different predictions of the two classes of similarity functions. In addition, the embedding procedure is capable of inferring different sets of attentional weights for different populations (e.g., novices vs. experts). The embedding procedure includes an active-selection procedure based on information gain that recommends what stimulus set should be presented next in order maximally reduce uncertainty in the resulting embedding.

Roads, B. D., Mozer, M. C. (in preparation). Obtaining psychological embeddings through joint kernel and metric learning.

Decision Support
In many situations, it isn't practical to train an individual to be an expert in binary classification. In those situations, it is useful to have a paradigm where novices can be guided to make expert-like decisions. Using a decision support paradigm, we help novices classify an unknown image by comparing it to a set of reference images with known labels. While novices may not know what visual features are diagnostic, they still have a powerful perceptual apparatus capable of judging similarity. By constructing a cognitive model of similarity judgments, reference images can be intelligently selected such that when novices

select based on similarity, they implicitly categorize an unknown image. Importantly, the selection procedure leverages within-class structure, as well as between-class structure, in order to determine the optimal reference images. Currently, we are exploring how neural networks can be used to better model contextual and sequential effects inherent in human decision making and choice.

Roads, B. D., & Mozer, M. C. (2017). Improving human-machine cooperative classification via cognitive theories of similarity. *Cognitive Science: A Multidisciplinary Journal 41* (5), 1394–1411.

Efficient Training
Cognitive science has numerous computational models that explain various aspects of visual category learning. Existing work typically provides qualitative theory-based recommendations on how to enhance visual task performance. However, relatively little work has been done to quantify these recommendations and to identify *optimized* training policies. My research has approached the issue of optimized training polices from three different perspectives. The first two approaches exploit structure in the psychological embedding by (a) predicting difficulty of learning and (b) guiding learners' attention to relevant feature dimensions. The third approach allows learners to request clues, adjusting the difficulty of the task and providing a richer set of behavioral data.

Difficulty-based Scheduling
Previous visual category training studies suggest that an easy-to-hard schedule benefits learning outcomes, i.e., begin with easy exemplars and then advance to more challenging exemplars. However, few studies have explored scheduling based on difficulty with complex, real-world images. Furthermore, conventional empirical methods for determining the learning difficulty associated with each image are costly. A cost-effective strategy is to infer an embedding from similarity judgments and use a simple category learning model to estimate ease of learning. We derive a simple parameter-free category learning model from the psychological literature and show that it accurately predicts exemplar difficulty. This approach makes it feasible to search over the space of scheduling policies indexed by difficulty in arbitrary image domains. For the domain of skin lesions, we compare easy-to-hard and hard-to-easy training policies, demonstrating that the easy-to-hard scheduling procedure is more efficient, requiring 27% fewer trials to train novices to correctly identify melanoma skin lesions.

Roads, B. D., Xu, B., Robinson, J. K., Tanaka, J. W. (submitted). The Easy-to-Hard Training Advantage with Real-World Medical Images.

Roads, B. D., & Mozer, M. C. (in preparation). Predicting the Difficulty of Human Category Learning Using Exemplar-Based Neural Networks.

Attention-Guiding Sequencing
Visual category learning is supported by guiding learners' attention to diagnostic visual information: features shared within a category and features that differ between categories However, the proper procedure for implementing these guidelines depends on the category structure within the embedding space. For example, research suggests that if categories are dissimilar from one another, trials should be *blocked*. In a blocked schedule, successive training trials tend to show exemplars from the same category. In contrast, *interleaved* schedules tend to be more efficient when the categories are highly similar. We argue that

blocked and interleaved schedules are special cases of a more general class of *attention-sensitive* schedules. As a proof-of-concept, we demonstrate that a mixture of the two scheduling procedures yields more efficient learning for a particular type of category structure. We are currently extending this work to optimize over the sequence of images in order to exploit attentional learning to diagnostic features.

Roads, B. D., & Mozer, M. C. (in preparation). Using human-surrogate models to optimize training sequences during visual category training.

Enriched Training Environments

The typical category learning paradigm involves a simple user interface that presents a query image and asks learners to classify either via multiple choice or free response. This paradigm has two potential drawbacks. First, each trial yields little information regarding the learner's knowledge state, making it challenging to adaptively prioritize content. Second, the monotonous task does not engage learners as active, information-seeking participants in the learning process. We compare the typical bare-bones training environment to an enriched environment. The enriched environment allows learners to request clues, thereby better adjusting the difficulty of the trial to the learners' state of understanding. For example, by eliminating some potential responses by providing a list of category labels or by providing visual examples of the category. Surprisingly, results indicate that an enriched training environment is not more efficient than a simple environment when training time is held constant. However, fewer enriched training trials are necessary to reach the same level of performance. Using recurrent neural networks, we tested which paradigm provides the best source of information for inferring a learner's knowledge state. Surprisingly, we found that the additional information obtained from the learner in the enriched environment did not benefit prediction; however, neural networks do predict better than traditional psychological category learning models.

Roads, B. D., Mozer, M. C. (in preparation). Using enriched training environments for visual category training.