Research Strategy

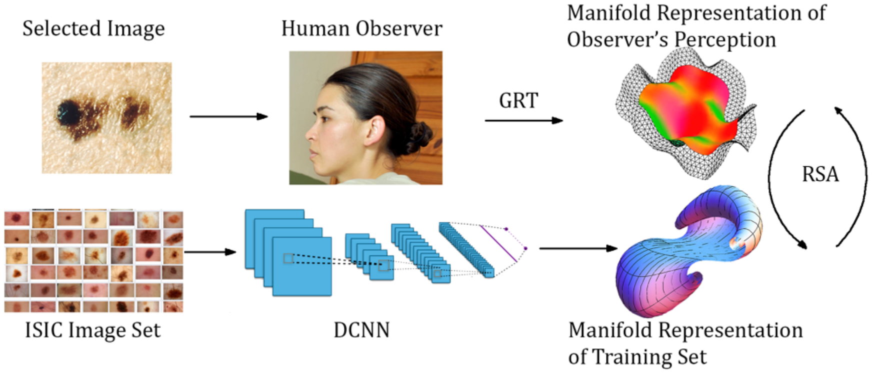
(A) Significance. The prevalence of melanoma, the deadliest form of skin cancer, is high, with nearly 1 person in 28 being diagnosed in their lifetime (1). The American Cancer Society estimates over 100,000 new melanomas will be diagnosed and over 7,000 people will die in the United States in 2021 (20). Fortunately, early detection has a large effect on survival rates (4). While public health campaigns and improved training have increased the rate of diagnosis, the number of deaths attributed to melanoma continues to rise (1-4) indicating additional efforts are needed. Proposed approaches to reducing the impact of melanoma can be broadly grouped into increased screening and the introduction of new technologies for screening.

Evidence for increased or universal screening is mixed (21), in part because many medical workers do not have sufficient training in categorizing skin lesions, and because there is no consensus for best practices for training lesion categorization (5-9). Existing guidelines (11,12) for determining whether a lesion should be assessed by a dermatologist or biopsied focus on rule-based procedures, such as the ABCDE approach, and depend in part on perceptual judgments (13). In contrast, expert dermatologists rely on holistic perceptual expertise rather than on a combination of clearly defined properties of a lesion (14,15). Together, these facts suggest the value of focusing on efficient perceptual learning (16). If training is sufficiently efficient, a much higher proportion of front-line practitioners, e.g., primary care physicians and non-physician clinicians could be trained, thus increasing early detection.

Technological advances in melanoma screening include the increased availability of imaging systems that better control light and image quality (22,23) and machine learning approaches— particularly deep convolutional neural networks (DCNN) (24), which are flexible, general mechanisms for categorizing image content —that leverage the controlled imagery. These advances have been bolstered by the International Skin Imaging Collective (18), which has made large training sets available and hosted competitions for machine vision-based lesion categorization. While this approach is promising, DCNNs have limitations in this setting. The most salient drawbacks are the black-box nature of these networks, which makes them difficult to use in collaboration with experts (25), and the fact that DCNNs can fail in unexpected ways due to unidentified limitations of the training sets (26). Even as these challenges are addressed, people—whether dermatological experts, front-line medical workers, or the general public—will continue to play a critical role in identifying potentially cancerous lesions. With the appropriate scientific advances, these technological advances can be leveraged to improve non-experts' lesion identification abilities.

(B) Innovation. Our pioneering effort is to develop an efficient perceptual training program for ultimate use by front-line medical practitioners by leveraging cognitive modeling, computer vision systems, and widely available image collections. Our proposed training efforts will be informed by geometric representations of human visual expertise at the behavioral and neurophysiological level and trained DCNNs. We will use the geometric representations to develop a novel framework for mapping between naïve and expert cognitive processes and between DCNN and human representations. The primary goal of our initial research is to collect data on expert and novice judgments of skin lesion imagery, develop models of how an expert and novices form those judgments, and establish a framework for mapping among those models and the representations of information embedded in a DCNN. To our knowledge, these approaches are novel.

Impact: Research on perceptual expertise in melanoma detection is limited. Speelman (16) found that practice discriminating skin lesions was far more effective than viewing a pamphlet. Milller (27, 28) demonstrated the value of training viewers on visual stimuli that ranged from clear exemplars of a category to ambiguous images. Additionally, Rimoin (29) found that an algorithm specialized for training perceptual expertise enhanced skin lesion categorization. Like Rimoin (29) and other applications of Kellman's approach (30,31) to adaptive learning, we will rely on both categorization accuracy and, for human observers, response speed. Our approach extends this research by integrating performance measures using cognitive processes rather than statistical models (cf. 30) and by modeling the integration of features that are treated as unrelated in models but not by observers (c.f. 29). In the current proposed work, we will build a mapping between representations of human perceptual expertise and DCNNs. This mapping will allow for a more efficient use of the available skin lesion images to train perceptual expertise. Finally, we will collect neurophysiological data to aid in identifying perceptual representations and in measuring how the representations change during learning.


Selected Image
Human Observer
GRT
Manifold Representation of Observer's Perception
RSA
ISIC Image Set
DCNN
Manifold Representation of Training Set

(C) Approach

(C1) Overview of the proposed work. The premise of our approach is that a geometric model of perceptual representation can capture 1) the experts' and novices' melanoma discrimination performance and 2) the variation in images represented by a DCNN that is trained to classify skin lesions. In this initial phase, we will rely on a single expert dermatologist to represent expert judgments and on standard psychology study participant recruitment for novices. Our target audience for training will ultimately include medical students, internists, physician assistants, nurse practitioners, etc., but we are not planning to test on those communities in this phase.

(C2) Rationale/Hypotheses. DCNNs can vary widely, and new innovations in applying the approach to melanoma detection are frequent. For the purposes of this proposal, we rely on the DCNN as described previously (32). This DCNN uses a network that was pre-trained on a wide range of imagery, with the final layer and connections from the penultimate layer defined and trained specifically for skin-lesion classification. One way of viewing the activations in the penultimate layer is as a representation of the result of transforming the input image to the best representation of that image for the classification decision. We focus on this representation of the DCNN for mapping to human expert perception.

Our cognitive modeling approach is based on general recognition theory (GRT) (17). GRT assumes that categories are represented as regions in a perceptual space separated by decision boundaries. GRT assumes that the stimulus is mapped to a noisy perceptual representation and then the observer makes a categorization decision based on the region into which the perceptual representation falls. We focus on GRT because it has been leveraged for many learning studies (33-35) and many of the core constructs in the model can be determined without parametric assumptions. GRT is also related to DCNNs in the sense that as aspects in the stimulus are varied, the changes in the probabilities of category confusions can be modeled as a probability distribution on a manifold (cf. 36). This relationship between GRT and DCNNs is key for a geometric approach for mapping among the potential training images, the DCNN representation, and representations of perception. Within the GRT framework, our goal is to model the geometry of the perceptual space of perceptual expertise in melanoma detection, including the process of learning that expertise. The critical aspects are:

1) What dimensions of the stimulus space are used for categorization?
2) What is the representation of distance in each of those dimensions?
3) What is the relationship between dimensions (e.g., the curvature)?

We plan to gather similarity rating data before and after learning to study how the perceptual space is altered by learning and how an expert's perceptual space differs from those of novices. With this initial phase of research, we will examine perceptual separability and perceptual independence for discriminating collections of images across the image set for novices and experts. These GRT constructs are directly related to how people perceived the relationship between aspects of the image, which, in turn, constrains the shape of the submanifold model of the perceptual space used for distinguishing among the manipulated stimulus dimensions. These properties of the geometric representation of the skin lesion imagery will be the foundation on which we build the mappings between different observers' representations and will ultimately enable us to identify the aspects of images the learner needs to practice.

In parallel, we will also estimate the shape of the representation of the training data implied by the activations in the penultimate layer of the trained DCNN. Our initial approach will be to rely on multidimensional scaling and similar methods which presuppose that the categorization can be represented reasonably well in a lower-dimensional subspace. We will do so for the sake of tractability. However, human decisions may be dependent on a more complex, even infinite dimensional, subspace. In the future, we hope to explore methods that do not rely on assuming a low-dimensional representation of decisions.

Once we have models of the shape of the DCNN representation, expert perceptual space, and novice representation space, we will need two more types of information. First, we will need mappings between the novice, expert, and DCNN spaces. Our initial approach will be to apply representational similarity analysis (RSA) (19). Briefly, RSA maps between two measurement approaches by comparing pairwise distances given between the two measurement approaches. For example, pairwise similarities between images judged by a human would be compared with the distances between vectors of node activations in a layer in a DCNN. These comparisons, usually quantified by rank-correlation, can then be used to compare among different aspects of each of the measurement spaces. An example application would be to compare participants' average similarity ratings to the patterns of activation in a DCNN layer when the images are grouped based on how symmetric the pictured lesion is.

To further constrain the potential process models and gain more insight into perceptual expertise, we will also measure neural activity for a subset of observers. Members of our research team and others have recently shown that by jointly measuring neurophysiological activity and behavior, we can apply GRT constructs to the neural representation of a stimulus in a task (35,37). For this work, we will use linear discriminant analysis (LDA) with EEG data, an approach that performs well in decoding analyses (38). A strength of LDA is that it makes predictions for the identification/confusion matrix for the categorization decision, which is the same level of analysis used for the GRT modeling. To obtain a model that maximizes categorization accuracy while minimizing the number of electrodes used for the decoding, we will divide the EEG time series into fixed length windows, then apply a stepwise model selection procedure. This approach will allow us to assess perceptual separability and independence in the evolving neural representation and to compare those neural based assessments to the inferences regarding perceptual separability and independence obtained from the behavioral identification/confusion matrix. We will use LDA on the data immediately prior to the onset of the lateralized readiness potential (LRP) to draw inferences regarding decisional separability in the neural representation and will compare that to inferences drawn from the behavioral identification/confusion matrix. All analyses---behavioral and EEG---will be done at the level of individual observers, with the intent of being able to document individual differences. This measure of variation across observers will be leveraged in our future work to develop individualized sequences of training images.

(C3) Study Design. Our first specific aim is to characterize the behavior and neural patterns novice and expert perceptions of images of skin lesions using mathematical cognitive models. The first two experiments and part of the third and fourth experiments will support us in achieving this aim. For Experiment 1, we will collect data from many naïve participants and one expert as they judge whether a skin lesion is malignant or benign. In Experiment 2, we will collect similar data, but with instructions to the participants to focus on specific dimensions of the images. Together, these data will be used with the GRT cognitive modeling framework to develop geometric representations of the perceptual space. To augment and refine the cognitive modeling efforts with neural characterizations, we will collect EEG on the first day of Experiments 3 and 4.

Our second specific aim is to **estimate the changes in perceptual and neural representations of skin lesions imagery over the course of extended training**. The third and fourth experiments are proposed to achieve this aim. Experiment 3 will mirror Experiment 1 in that the observers will judge whether the lesion in an image is malignant or benign; however, the observers will continue to make these judgements over the course of 35 days to learn the perceptual discriminations. Similarly, Experiment 4 will mirror Experiment 1, in that participants will be instructed on the dimensions to which they should attend, and they will participate for 35 days. The participants' patterns of errors over the course of training will be modeled using the GRT cognitive modeling framework, and like the first aim, we will use the EEG data collected at the beginning and end of training to augment our models with neural representations.

Our third specific aim is to **connect cognitive models of human perceptual representations to deep neural network representations of the information in its training imagery**. The general structure of these models will be developed in parallel with data collection as they are not dependent on the outcomes of the experiments. Additionally, the comparisons between the performance of participants using the DCNN-derived dimensions and the participants in the unstructured and ABCD based experiments can be used for basic assessments of the relationship between human perceptual representations and the DCNN representations. These assessments will include direct comparisons and estimated mappings using RSA. If we are successful in developing the representations targeted with our first aim, we will be able to use the models and data to derive richer mappings that more directly connect the cognitive models to the DCNN representations by relying on differential geometry to map between the spatial representations of both.

Experiment 1: Unstructured Classification. The goal of this experiment is to get a broad, if imprecise, a measure of observers' perceptual processes for classifying skin lesions. We will recruit approximately 85 participants to capture the representation of novice observers. Naïve observers will be recruited from undergraduate psychology research pools and online. We will screen normal color vision and acuity and query participants regarding their knowledge of skin cancer. To capture expert perceptual processes, we will focus on a thorough estimation of a single expert's categorization performance. To avoid biasing the observers, we will neither give instructions about features on which to focus, nor will we give trial-by-trial feedback.

Stimuli will be drawn randomly from either images of malignant lesions (5,714 possible images) or benign lesions (47,684 possible images) hosted by the International Skin Imaging Collaboration (ISIC). Participants will be given up to 30 seconds to categorize an image as malignant or benign. Naïve participants will be shown as many images as they can classify within a one-hour session; we expect this will be around 800 to 1000 images. The expert observer will classify 2000 images. Although the base rates of malignant and benign skin lesions are not equal and the relative frequency of viewing the types likely influences the categorization process (39), we will present each category equally frequently to focus on the perceptual discrimination.

The analysis model will assume the same fixed perceptual representation across naïve individuals allowing for different attentional weights across dimensions. The two fixed spaces we will consider are the four-dimensional space represented by the first four dimensions of the ABCDE, algorithm (asymmetry, border, color, and diameter) and a representation based on the activation pattern of the penultimate layer in the trained DCNN. We anticipate being able to measure upper and lower performance limits with these images. We will also assess the variation in how stimulus dimensions are treated among novice observers and the distinctive properties of the expert's approach.

The largest risk with this experiment will be whether we can acquire data from enough subjects. For the novice subjects, the potential to recruit from three different universities mitigates this risk. The expert is not easily replaced; however, we have connections with dermatology programs at The University of Oklahoma and The University of Texas at San Antonio for potential backups.

Experiment 2. Structured Classification. The goal of this experiment is to assess individual perception and categorization performance when observers are explicitly asked to attend to specific dimensions of the stimulus. We will examine performance with both the ABCD dimensions and dimensions derived from the DCNN classifier. To allow for more direct comparison between the ABCD approach and the DCNN approach, we will reduce the number of dimensions from the penultimate layer of the network to four using one of the following three methods: 1) Principal components analysis (PCA) on the activation space; 2) selecting the four nodes with the highest magnitude weights for the final decision node; or 3) training the network with a four-node layer as the penultimate layer. The approach to dimensionality reduction will be determined through analysis of the DCNN activations and limited pilot data.

The number of participants, recruitment approach, and stimulus set will be the same as Experiment 1, but we will only recruit participants that did not participate in the first experiment. Instead of asking participants to classify images as malignant or benign, we will ask them to classify the image based on one or two of the four dimensions, ABCD or DCNN. The dimension will be varied within subjects and the basis for classification (ABCD or DCNN) will be manipulated between subjects. Each subject will complete half of the trials using a single dimension and half of the trials using a pair of dimensions. For the two- dimensional trials, participants will be asked to indicate whether the image looks malignant on both dimensions, following standard complete identification GRT procedures. Each half of the trials will be distributed evenly among the dimension and pairs of dimensions evenly, and the order will be counterbalanced.

For the single-dimensional analyses, we will use hierarchical Bayesian estimates of the function mapping between the stimulus scale and the probability of classifying the image as malignant (40). To do this, we need a metric for the stimulus dimensions. For the DCNN dimensions, we will use the activation in the neural network at the corresponding node or derived dimension based on PCA. For the ABCD dimensions, we will use either a neural net explicitly trained to measure ABCD dimensions (41) or assess the dimensions using the Bradley-Terry-Luce model from the group level data (42,43). For the two-dimensional analyses, we will estimate Bayesian hierarchical GRT models assuming either linear decision bounds between categories or optimal decision bounds, which we will determine based on model fit to the data.

These data will be valuable for three main purposes: 1) estimate how sensitive novices are to variation in important stimulus dimensions; 2) indicate novices' abilities with neural net derived dimensions and ABCD dimensions; 3) provide a baseline of sensitivity prior to training. The largest risks for this experiment are that subjects will not be able to understand the task when it relies on DCNN derived dimensions and that we are not able to obtain a good metric for the ABCD dimensions in the images. To mitigate these risks, we have proposed multiple potential solutions, three alternatives for DCNN based dimensions (activations of a select subset of nodes in the penultimate layer, PCA based representation of the activation in the penultimate layer, or a trained network with only four nodes in the penultimate layer) and two alternatives for measuring the ABCD levels within an image (derived from a trained neural network or using a model fit to the group level data). If naïve observers are not able to use DCNN dimensions with any of the three approaches, results are still valuable and indicators that the DCNN dimensions would not be useful to novices, and hence not something that could replace ABCD. If neither of the options for the ABCD dimensions work, a costly but safe option would be to have observers explicitly rank images on each of the four dimensions.

Experiment 3: Unstructured Learning. The goal of the next two experiments is to examine the change in observer's perceptual representation of skin lesions as a function of experience and to measure the influence of targeted training on learning. To characterize the learning process, we will follow a similar structure to the first set of experiments, examining both structured and unstructured stimulus environments. Likewise, we will measure perceptual change through the lens of geometric representation of the perceptual decision-making processes. In addition to the basic behavioral data, we will collect EEG data during training to gain information about how the internal representations change over practice and within trials.

Experiment 3 will obtain a baseline estimate of the perceptual learning process that is unbiased by specific strategies. We will recruit nine participants with the goal of obtaining complete training data from six. All participants will possess normal vision and have no prior experience classifying skin lesions. Stimuli will be drawn at random from the same ISIC image set used in Experiment 1, and observers will be asked to classify the images as malignant or benign. A key difference from Experiment 1 is that participants will be given immediate feedback as to whether they correctly classified the image. Training will last for 35 one-hour long sessions, each with at least one night and at most four days between each session.

EEG will be measured on the first and 35th day of training using a 128-channel EGI system (Magstim/EGI, Eugene OR). EEG will be recorded continuously with hardware filters set from 0.1 to 100 Hz, a sampling rate of 1 K Hz, and an online vertex reference. Impedances will be kept below 50 KΩ for the entire session. Continuous EEG will be epoched with reference to the behavioral response, running from -500 ms to 1500 ms. Data will be preprocessed and analyzed using EEGLab (44): Data will be first inspected visually, and bad channels will be deleted, then artifacts will be rejected, and any additional bad channels will be deleted on the basis of probability estimated by EEGlab.

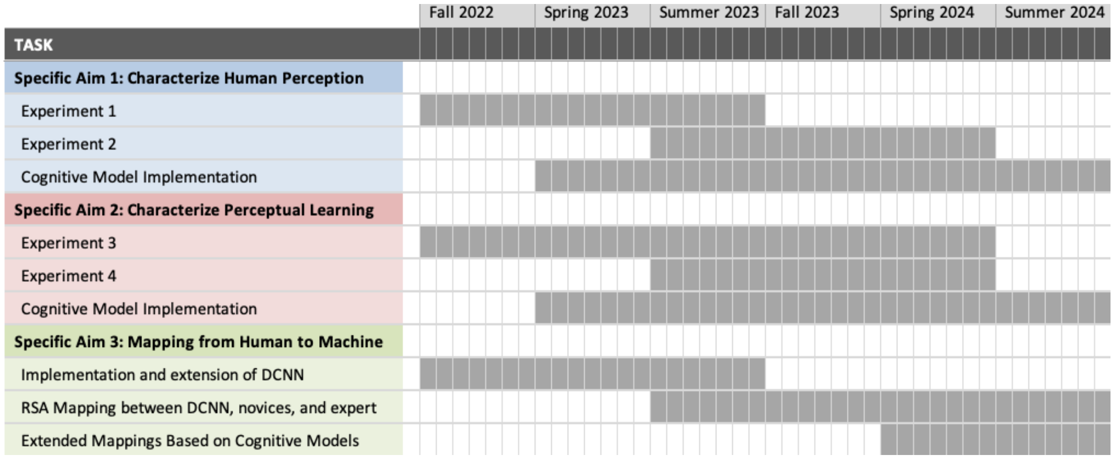
We will fit the same models as in Experiment 1 to the performance data during training. Our goal is to use independent model fits for each day of training so as minimize the bias introduced into potential learning structures. If the data prove insufficient for individual level model fits, we will assume an autoregressive structure on the parameters across training sessions and jointly fit the full training set for an individual and/or fit hierarchical models of the parameter variation across participants and time. We will also track more straightforward metrics of learning, including changes in average accuracy and response times across days, and in particular look for asymptotes in those metrics that may indicate performance plateaus. EEG data will be analyzed using the LDA approach described above.

We expect a gradual, but relatively smooth, change in categorization performance across sessions. Given that we will have naïve participants before training, we expect the model parameters to be very similar to the parameters estimated in Experiment 1 before training. Through the course of training, we expect the parameters to shift toward representations more in line with either the expert, the DCNN, or both.

Experiment 4. Structured Learning. The goal of this study is to examine whether a focus on specific perceptual dimensions in training is more efficient for learning to categorize skin lesions. We will focus on two sources of perceptual dimensions: those derived from the cognitive model analysis of expert image classification and those derived from the DCNN. By querying participants on two dimensions at a time, we can estimate the dependencies in the perception-decision process through the GRT model.

Recruitment, target sample size, and image set will be identical to Experiment 3. In contrast to Experiment 3, we will ask learners to focus on two image dimensions at a time. Learners will be given trial-by-trial feedback. Our goal is to train participants on subsets of the perceptual dimensions derived from the model of the expert viewer's perception. Given that we expect the relationships between perceptual dimensions to depend on the degree to which an image represents a dimension (e.g., how symmetric the depicted lesion is), we may need to train on a subset of the dimensions. On the first and 35th days of training, when EEG is measured, participants will complete the unstructured classification version of the task from Experiment 3.

The analyses will mirror those proposed for Experiment 3. We expect the relationships across perceptual dimensions recovered by the GRT modeling to more quickly converge to the structures measured from the expert. We also expect faster learning of the image sets and higher levels of asymptotic performance. The largest risk to this portion of the study is if the analyses of Experiment 1 fail to result in coherent dimensions for structured training. In this case, our fallback plan is to train on dimensions separately, ideally with the DCNN derived dimensions, but if those turn out to be too difficult to convey to the learner, we will use the ABCD dimensions.



TASK	Fall 2022	Spring 2023	Summer 2023	Fall 2023	Spring 2024	Summer 2024
Specific Aim 1: Characterize Human Perception						
Experiment 1	■	■	■			
Experiment 2			■	■	■	
Cognitive Model Implementation		■	■	■	■	■
Specific Aim 2: Characterize Perceptual Learning						
Experiment 3	■	■	■	■	■	
Experiment 4			■	■	■	
Cognitive Model Implementation		■	■	■	■	■
Specific Aim 3: Mapping from Human to Machine						
Implementation and extension of DCNN	■	■	■			
RSA Mapping between DCNN, novices, and expert			■	■	■	■
Extended Mappings Based on Cognitive Models					■	■

Summary. There will be three main products of this research. In accordance with our first specific aim, we will have a collection of data containing information about the patterns of discrimination performance for both novice viewers and an expert. Even with simple models, these data could be used to predict what types of lesions lay people might view as worrisome that are likely not melanoma, as well as lesions that people tend to see as non-threatening that are more likely to be melanoma. These data will be made available to the research community through the Open Science Foundation near the end of the research period. Additionally, at the conclusion of this phase of the project, we will have a model of how information in skin lesion imagery is used by experts, novices, and machine vision systems. These models will be valuable to understanding melanoma perception by each of those three, predicting performance for each of the three, and for understanding the differences among the three. In satisfaction of our second specific aim, we will have a dataset and cognitive models focused on how people's perception of skin lesion imagery changes over the course of 35 days of training. Finally, for the third specific aim, we will obtain estimates of the mappings between each pair of novices, our expert viewer, and a DCNN. These mapping will be valuable for understanding the learning process for novices, determining the images that best represent the deficits in a novice viewers knowledge, and for understanding what information in skin lesion images is used by both experts and machine vision and which sources are used exclusively by one or the other.

While these advances are valuable on their own, our goals are to use the products of the currently proposed research as the foundation of a larger project. Our primary goal is to develop an optimized perceptual training regime for identifying potential melanomas. In the follow-on work, we will leverage the mappings between human experts, machine vision systems, and human novices to identify the critical information that a novice needs to learn and display the appropriate example images for learning that information. By relying on spatial models of perception, we will be able to train based on abstract features of lesions that may not be readily named (e.g., in contrast to "symmetry"), but nevertheless are relied upon by experts. Using the technical innovations of this currently proposed research, we will perform iteratively larger studies on the effectiveness of this type of training both in controlled lab-based settings and in clinical settings. Even incremental progress in training people to identify lesions will result in many lives saved and significant reductions in costs related to treating advanced stages of melanoma.

In addition to the applied value of this work, our contribution to connecting machine vision to human vision is potentially transformative across a range of sciences. Significant basic research funding, particularly from defense research agencies, has gone into explainability in AI systems, and more generally into improving the joint performance of human-AI systems. Our framework for mapping between human and machine vision enables communication from the AI system to the human system, allowing the AI system to determine what image best exemplifies the visual information that needs to be communicated. The advantages of our approach are mirrored in the potential value to computer vision as well. A solid representation human visual expertise allows for more focused training of machine vision algorithms as well as serves as a source of information to complement standard training based on large image sets. In future work, we plan to build on the constrained spatial models that we will develop in this project toward spatial representations that allow for more complex relationships among dimensions of the input, including leveraging modern approaches to multidimensional scaling.