# How to Model Implicit Knowledge? Similarity Learning Methods to Assess Perceptions of Visual Representations

Martina A. Rau
Educational Psychology
University of Wisconsin—Madison
1025 W. Johnson St
Madison, WI 53706
marau@wisc.edu

Blake Mason
Electrical and Computer Engineering
University of Wisconsin—Madison
1415 Engineering Dr
Madison, WI 53706
bmason3@wisc.edu

Robert Nowak
Electrical and Computer Engineering
University of Wisconsin—Madison
1415 Engineering Dr
Madison, WI 53706
nowak@ece.wisc.edu

## ABSTRACT

To succeed in STEM, students need to learn to use visual representations. Most prior research has focused on conceptual knowledge about visual representations that is acquired via verbally mediated forms of learning. However, students also need perceptual fluency: the ability to rapidly and effortlessly translate among representations. Perceptual fluency is acquired via non-verbal, implicit learning processes. A challenge for instructional interventions that focus on implicit learning is to model students' knowledge acquisition. Because implicit learning is non-verbal, we cannot rely on traditional methods, such as expert interviews or student think-alouds. This paper uses similarity learning, a machine learning method that can assess how people perceive similarity between visual representations. We used this approach to model how undergraduate students perceive similarity between visual representations of chemical molecules. The approach achieved good accuracy in predicting students' similarity judgments and expands expert predictions of how students might perceive visual representations of molecules.

## Keywords

Perceptual knowledge, implicit learning, visual representations, similarity learning methods, chemistry.

## 1. INTRODUCTION

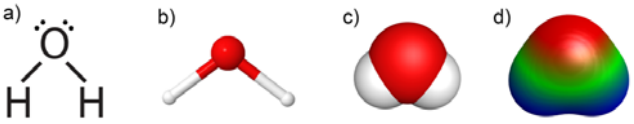


**Figure 1. Visual representations of chemical molecules: a: Lewis structure; b: ball-and-stick model; c: space-filling model; d: electrostatic potential map (EPM) of water.**

Visual representations are ubiquitous instructional tools in science, technology, engineering, and math (STEM) domains [1, 2]. For example, instructors use the visual representations shown in Figure 1 to help students learn about chemical bonding. Yet, to a novice student, these visual representations may not be helpful because the student may not know how to interpret the representations. For instance, does the red color in the ball-and-stick figure (Figure 1-b) mean the same thing as in the electrostatic potential map (EPM; Figure 1-d)? (It does not.)

Instructors often ask students to use visual representations that they have never seen before to make sense of concepts that they have not yet learned about [3, 4], an issue known as the *representation dilemma* [5]. Hence, to succeed in STEM, students need *representational competencies* that enable them to use visual representations to make sense of and solve domain-relevant problems [6, 7]. One crucial representational competency is the ability to interpret visual representations; that is, to map visual representations to the abstract concepts they depict [6, 8]. For example, students need to understand how the representations in Figure 1 show information about the molecule. For the Lewis structure (Figure 1-a), the student may map the unbonded electrons shown as dots to conceptual knowledge about how polarity in chemical molecules and infer that the water molecule has a local negative charge by the Oxygen atom.

Educational technologies are particularly suitable to support representational competencies because they can provide adaptive support while students solve domain-relevant problems [9, 10]. Such adaptive support relies on a cognitive model that infers whether the student has learned target skills based on her/his interactions with the technology. Research shows that adapting instruction to students' representational competencies can enhance those competencies [11] and learning of domain knowledge [12].

However, educational technologies for representational competencies have two critical limitations. First, they typically focus on one set of representational competencies: students' conceptual understanding of representations (e.g., the ability to explain how visual features depict concepts). This focus mimics education psychology research's focus on conceptual learning [6, 13]. Conceptual knowledge is invariably intertwined with a second type of representational competency: *perceptual knowledge* [14, 15]; the ability to rapidly and effortlessly recognize conceptual information based on visual features of the representations. This ability results from *implicit forms of learning*. For example, expert chemists simply "see" that the molecules depicted in Figure 1 have a local negative charge by the Oxygen atom, without having to make a an effortful conceptual inference.

Second, of the few educational technologies that enhance perceptual fluency, their adaptive capabilities are limited and their perceptual supports rely solely on performance measures (e.g., accuracy, response times) to adapt to students' representational competencies [15, 16]. They do not use a cognitive model of the latent skills that students acquire through perceptual learning. As a result, they cannot provide specific feedback when students make mistakes. Decades of research showing that cognitive models can dramatically increase the effectiveness of educational technologies [10, 17] suggest that we must address this limitation and create adaptive instruction for perceptual knowledge.

These limitations likely result from cognitive modeling's traditional focus on explicit, verbally accessible knowledge. To develop cognitive models, researchers analyze how students think about target skills [9, 18]. We typically ask students to verbalize their problem-solving steps [19, 20]. Yet, verbalization is not suitable for assessing perceptual learning processes, which are implicit and not verbally accessible [14, 21]. Therefore, instructional designers have to rely on "educated guesses" as to which visual features students may pay attention to. These educated

guesses are based on the novice-expert literature, which documents the fact that novices tend to rely on surface features; that is, easily perceivable visual cues such as color and shape, to judge the similarity between stimuli items. By contrast, experts rely on visual features that are conceptually relevant and hence make more refined distinctions between visual features. Thus, to create adaptive perceptual supports, we need to develop cognitive models for perceptual learning.

Our research takes a first step towards developing a cognitive model for perceptual learning by assessing students' perceptual knowledge of a common visual representation in chemistry. In particular, we investigate *research question 1:* Which visual features do students focus on when presented with visual representations? To address this question, we asked hundreds of students to judge the similarity between visual representations of molecules. We then used *similarity learning*—a machine learning method that provides a formal approach to investigating how people perceive similarity among visual stimuli. This method allowed us to estimate latent factors that account for the perceived similarity relationships between representations. Because we can map these latent factors to the visual features in the representations, this approach allows us to investigate which visual features are most salient to students' perceptions of similarity. Comparing these visual features to "educated guesses" allowed us to test *research question 2:* Do the visual features we identified as salient via metric learning correspond to visual features that students are expected to attend to based on the expert-novice literature on perceptual learning? In addition, we investigated a methodological *research question 3:* How many similarity judgments we need to assess students' perceptual knowledge?

Although we address these questions in the context of a particular domain with a particular visual representation, this paper makes two important broader contributions. First, it provides an empirical validation of the "educated guesses" that developers of perceptual learning technologies typically rely on. Second, it establishes a methodology to assess perceptual knowledge that can serve as a basis for a cognitive model of perceptual learning. These contributions build the foundation for the development of adaptive instruction for perceptual knowledge and other implicit knowledge.

# 2. EXPERIMENT

## 2.1 Visual Representations of Molecules

For our experiment, we selected visual representations of chemical molecules common in undergraduate instruction. Lewis structure representations are the most commonly used visual representations in undergraduate chemistry textbooks. We reviewed textbooks and online instructional materials and listed the frequency of all occurring molecules using their chemical names (e.g., $H_2O$) and common names (e.g., water). For our experiment, we chose the 50 most common molecules.

First, we created *educated guess features* (Figure 2, yellow) that correspond to expert assessments of which visual features students may attend to when making similarity judgments. To obtain these educated guesses, we reviewed the literature on chemistry expertise [22, 23] and on perceptual learning [14, 24], and conducted learner-centered interviews with undergraduate and PhD students in chemistry [25]. We identified 6 educated guess features: number of total letters, number of distinct letters, number of total bonds, number of single bonds, number of unbonded electrons, and molecule geometry (linear, planar, tetrahedral).

To investigate which visual features drive students' similarity judgments, we quantitatively described the visual features of the

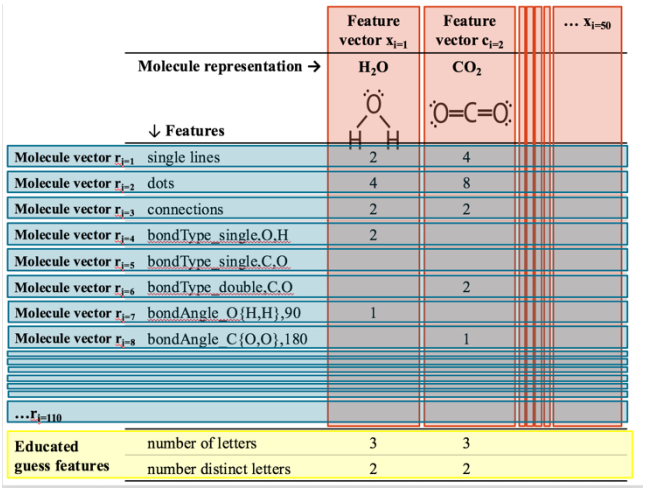


**Figure 2. Example features for $H_2O$ and $CO_2$ molecule representations with educated guess features in yellow, feature vectors in red, and molecule vectors in blue.**

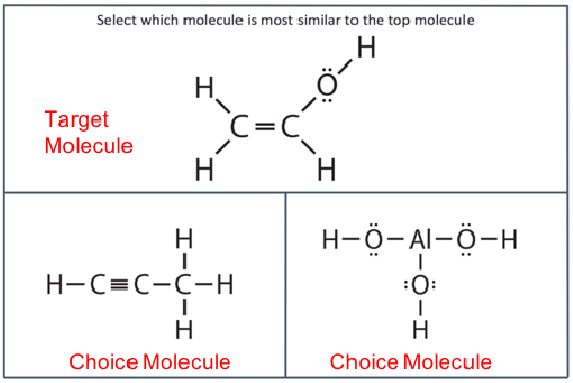


**Figure 3. Example of a similarity judgment task: given the molecule on the top, students were asked which of the two molecules at the bottom is most similar.**

molecule representations. To this end, we created *feature vectors* for each of the molecules (see Figure 2, red) that describe which visual features the representation contains (e.g., bond angles, the numbers of specific atoms, or the numbers of different atoms present). The feature vectors of our corpus of molecule representations contained a total of 110 features. The 50 feature vectors collectively form matrix $X = [x_1, x_2, x_3, \dots, x_{50}]$, where $x_i$, is the feature vector for the *i*th molecule.

We aggregated each element of the feature vectors into *molecule vector* for individual features (Figure 2, blue). Each molecule vector consisted of 50 values describing how many times the feature occurred in each representation. As molecule vectors make up the rows of our matrix of 110 features by 50 molecules shown in Figure 2, we will refer to the molecule vector for the *j*th feature as $r_j$. Thus, feature vectors provide a numeric description of the visual information present in each representation, whereas molecule vectors provide a numeric description of overall patterns of visual features in the dataset for all representations.

## 2.2 Similarity Judgment Tasks

Students completed similarity judgment tasks that were presented as triplet comparisons (see Figure 3). Given a representation of a molecule (the "target-molecule"), students were asked to choose molecules") was most similar to the given one. For each task, the student chose between one of the two choice-molecules that

he/she perceived to be more similar to the target-molecule. After each task, another triplet was generated uniformly at random from our corpus of molecule representations.

We delivered the similarity judgment tasks via NEXT; a cloud-based machine learning platform [26]. NEXT allows users to upload their own content and query participants to perform judgment tasks. It uses machine learning algorithms to automate data collection and analyze results. More information about the platform can be found at http://nextml.org. In NEXT, students first received a brief description of the study and then worked through a sequence of 50 similarity judgment tasks. Students were instructed that these tasks are not a test and that there is right or wrong answer, but that we they are simply asked about their personal perceptions of similarities among molecule representations.

## 2.3 Dataset

Undergraduate students enrolled in an introductory chemistry course at a large U.S. university were invited to participate in a survey on learning with visual representations. The course had an enrolment of 781 students. Participation was voluntary. Altogether, we collected 26,180 responses from 563 (possibly non-unique) students. 61.6% of the students completed all 50 similarity judgment tasks. On average, students completed 46.5 tasks. Each similarity judgment in response to a triplet comparison task was associated with the feature vectors ($x_i$) and molecule vectors ($r_j$) of the three molecule representations, as described in 2.1.

# 3. ANALYSIS

In the following, we describe how we used similarity learning to investigate which visual features drive students' similarity judgments. We first provide a brief introduction into the metric learning method in general. Then, we describe how we applied this method to our dataset in particular.

## 3.1 Introduction to Similarity Learning

In general, the goal of similarity learning is to learn a similarity function *f* that agrees with students' similarity judgments in the following sense: if item *i* is judged to be more similar to *j* than to *k*, then f(i,j) < f(i,k). The function *f* can be thought as quantifying the perceived distance or dissimilarity between pairs. Alternatively, the function could quantify the perceptual similarity (inverse distance) between pairs, in which case f(i,j) > f(i,k).

People are better at providing ordinal (i.e., comparative) responses than at providing fine-grained quantitative judgments or ratings [27]. For example, when asked to compare the visual representations in Figure 3, people find it easier to judge whether the target molecule is more similar to the left or the right choice molecule than to judge their similarity on a rating scale. However, it is challenging to machine-learn embeddings from comparisons due to the sheer number of possible triplet comparisons that could be made; the number of distinct triplets is proportional to $n^3$. For example, in our case of n=50 molecule representations, there exist nearly 125,000 distinct triplets. Researchers have observed that while triplet comparisons are easy to answer, they can become tedious and boring after extended sessions [28]. Since we hypothesize that perceived dissimilarities can be accurately represented in d-dimensional space, it is reasonable to conjecture that if the embedding dimension is low (i.e., $d \ll n$), then there will be a high degree of redundancy among the triplet comparisons. In fact, researchers have observed that a small subset of these triplet comparisons often suffice to learn a reasonably accurate embedding, lending support to this conjecture [29-31].

## 3.2 Similarity Learning Approaches

We applied two similarity learning approaches in this paper: similarity learning by ranking [32] and non-metric multi-dimensional scaling. In both cases, we modelled the perceptual similarity between molecules *i* and *j* as

$$S_{ij} = \mathbf{x}_i^T \mathbf{A} \mathbf{x}_j$$

Here $\boldsymbol{A}$ is a symmetric matrix that parameterizes the model. The $k$,lth element of the matrix, denoted by $A_{kl}$ represents the importance of the interaction of feature *k* and feature *l* in the model. Since we assume $\boldsymbol{A}$ is symmetric, $A_{kl} = A_{lk}$ and $S_{ij} = S_{ji}$.Before introducing these approaches, let us define some notation. There are *N* triplet comparisons. For the *n*th triplet, let $i_n$ denote the target-molecule and let $j_n$ and $k_n$ denote the two choice-molecules. Let $y_n$ denote the student's judgment, specifically $y_n = +1$ if the student decided $j_n$ was more similar to $i_n$ and $y_n = -1$ otherwise. Each of the $p = 50$ diagrams also has *m* associated features (e.g., numbers of different atoms, bonds, etc.). Arrange the features for each molecule representation into an *m* x 1 molecular feature vector, and the *m* x 1 feature vectors into a *m x p* matrix, *X*. The *i*th column of *X*, denoted $x_i$, contains the m features for molecule *i*. The *j*th row of *X*, denoted $r_j$, is a molecule vector for feature *j* containing the value of feature *j* for all 50 representations.

### *3.2.1 Approach 1: Similarity Learning by Ranking*

This approach learns matrix *A* in our model of perceptual similarity directly from triplet responses via linear regression.

$$S_{ij} = \mathbf{x}_i^T \mathbf{A} \mathbf{x}_j$$

where $x_i$ and $x_j$ are *m* x 1 dimensional feature vectors of the *m* features of molecule representations *i* and *j*. The matrix *A* is *m* x *m*, and the metric learning problem is to estimate *A* that minimizes the number of disagreements between the ranking predictions for each triple (i.e., either $S_{ij} > S_{ik}$ or vice-versa) and the comparative judgments collected from the students, as proposed by [32].

The first step in this analysis was to estimate *A*. Formally, the estimation of *A* can be written as the following optimization problem. Let $\mathbb{S}_m$ be the set of all *m* x *m* symmetric matrices. Solve for *A* that minimizes:

$$\widehat{\mathbf{A}} = \arg\min_{\mathbf{A} \in \mathbb{S}_m} \sum_{n=1}^{N} \left( y_n - \mathbf{x}_{i_n}^T \mathbf{A} \mathbf{x}_{j_n} + \mathbf{x}_{i_n}^T \mathbf{A} \mathbf{x}_{k_n} \right)^2$$

where the superscript T denotes the vector transposition. The matrix *A* that minimizes the sum of squared errors weights the similarities between the diagram features so as to predict perceptual similarity judgments. In general, the solution *A* will place some weight on all *m* features. We anticipate that the visual features that are not salient do not strongly affect students' similarity judgments and therefore have lower weights in *A*.

Taking this thinking a step further, we could consider many different optimizations of the type above, where in each case we use different subsets of the features, in order to determine which are most predictive of student judgments. Indeed, some features may be totally irrelevant and worsen, rather than help, the prediction of students' similarity judgments. Unfortunately, searching over all possible subsets of features is computationally infeasible, so we instead consider the following optimization that approximates this search problem called sparse COMET [33].

$$\widehat{\mathbf{A}} = \arg\min_{\mathbf{A}\in\mathbb{S}_m} \sum_{n=1}^{N} \left(y_n - \mathbf{x}_{i_n}^T \mathbf{A} \mathbf{x}_{j_n} + \mathbf{x}_{i_n}^T \mathbf{A} \mathbf{x}_{k_n}\right)^2 + \lambda \sum_{k=1}^{m} ||\mathbf{A}(k,:)||_2^2$$

This optimization method uses a cost function that consists of two terms. The first term represents least squares data-fitting cost in the previous optimization. The second term is a Group LASSO penalty, which encourages solutions that have many columns equal to 0. If a column in *A* is all zero, then the corresponding feature is not used for prediction. The number of zero-valued columns in the solution depends on $\lambda > 0$. Note that we recover the previous optimization when $\lambda = 0$. Larger values of $\lambda$ produce sparser solutions that effectively use fewer features. Features crucial for prediction are excluded only if $\lambda$ is exceedingly large.

The second step in this analysis was to tune the parameter $\lambda$ and then to assess the prediction accuracy of our method. To this end, we used 10-fold cross validation. Specifically, we randomly split the complete dataset into 10 equal sized subsets. We removed 2 random subsets as hold-out data and kept the remaining data as training data. We then solved the optimization above with the training data over a range of different $\lambda$ values. For each $\lambda$, we scored prediction accuracy on one set of hold-out data to select the optimal value. Then, using our chosen $\lambda$ value, we solved the optimization again to obtain a final *A* using 9/10 of the data, and assessed the prediction accuracy on remaining 1/10 of the data.

The final step was to rank the features based on the weights in matrix. Due to the Group LASSO penalty in the loss function, many of the columns in the resulting matrix are zero. To get the aggregate weight of each relevant feature, we computed the length (norm) of each non-zero column and ranked accordingly.

### *3.2.2 Approach 2: Ordinal Embedding*

In this approach, rather than directly making predictions of similarity based on feature vectors and triplet responses, we first used students' similarity judgments to learn an embedding that spatially represents the similarity of molecule representations as distances in 2-dimensional space. We then identified molecule vectors that account for the distribution of molecule representations in the embedding space.

The first step in this analysis was to learn an embedding. We applied non-metric multidimensional scaling (NMDS) to the 26,180 triplet comparison responses collected from the experiment to learn an embedding of the 50 molecule representations in a two-dimensional space [22]. Embedding in two dimensions allows visualizing the perceived similarity computed by NMDS. The embedding reflects the consensus among students as to which molecular representations were more or less similar. We created 50 different embeddings, using multiple random initializations per embedding in order to account for the non-convexity of NMDS.

The second step was to validate the embedding. To this end, we computed a distance matrix for each embedding. To validate the distance matrices, we used the following cross-validation procedure. We selected 6000 triplet comparison responses uniformly at random to serve as a hold-out dataset. From the remaining triplets, we randomly selected training sets of different size, ranging from 1000 to 20,000 triplet comparison responses. We computed embeddings for each training set. We then used these embeddings and the associated distance matrices to predict students' similarity judgments. Next, we used the distances in the embedding as a predictor of judgments in the hold-out set; the prediction errors quantify how well the embedding reflects the judgments. We repeated this procedure for training sets of different size. We performed 50-fold cross validation to calculate average prediction error on the learned embeddings. This procedure allowed assessing how prediction performance relates to the training set size (i.e., how many triplets were used to compute an embedding).

The third step in our analysis, after validating our embedding procedure, was to compute an embedding and corresponding distance matrix from the full set of triplets. Since the distance between points in the embedding corresponds to their perceived dissimilarity, we computed a similarity matrix defined as the element-wise inverse of the distance matrix, scaled from 0 to 1.

The fourth step was to identify which features, represented by the feature vectors, drive students' similarity judgments. Because the embedding was performed in 2 dimensions, we can consider the problem of only choosing 2 feature vectors to combine and compare combinations of pairs of feature vectors to the similarity matrix. For each possible pair, we performed a least squares optimization to find the ideal uniform scaling to match an outer product of our feature vectors to the similarity matrix.

$$\widehat{A} = \arg\min_{\mathbf{A}} \sum_{i,j=1}^{p} \left(S_{ij} - \mathbf{x}_i^T \mathbf{A} \mathbf{x}_j\right)^2$$

subject to $A_{st} = 0$ for all *s,t* not equal to *k,l* or *l,k*. In other words, only let the *k,l* elements of *A* be non-zero and optimize these. This equates to fitting *S* to the molecule vectors for features *k* and *i*. Here, $S_{ij}$ represents the value of the perceptual similarity between molecules *i* and *j* from the embedding. The magnitude of resulting value of $A_{kl}$ tells us how important the interaction of features *k* and *l* is in representing the similarity. This is basically a correlation coefficient, and it only gauges the marginal value of this interaction (i.e., in isolation of all other interactions). In each case, after learning a matrix *A_* we computed the corresponding residual value between similarity matrix *S* and our combination of 2 features. After performing all possible combinations of pairs of features, we ranked pairs of features in ascending order of residual values, with the smallest residuals being the best approximation of our observed similarity matrix. To evaluate the feature rankings, we used 10-fold cross-validation by performing identical tests on 10 different similarity matrices computed from different embeddings based on equal numbers of triplets to ensure that the original embedding and the non-convexity of NMDS was not a factor in the final ranking of feature pairs.

# 4. RESULTS

## 4.1 Identifying Important Visual Features

To address *research question 1*, we used the two similarity learning approaches just described to identify which visual features account for students' similarity judgments.

### *4.1.1 Approach 1: Similarity Learning by Ranking*

Recall that the first approach entailed learning a similarity function that describes students' perceived similarity between molecule representations. This approach yielded an average 69% prediction accuracy of students' similarity judgments (assessed via 10-fold cross validation). This finding indicates that there was consensus over which representations were more or less similar, but also that there were some disagreements among students' similarity judgments.

To identify which visual features account for students' similarity judgments, we estimated the weights for each feature in the ma-

**Table 1. Top 10 features from the ranking of features with strong weights obtained by Approach 1.**

| Feature | Avg weight |
|---|---|
| Distinct letters | 4.50% |
| Single bonds between Oxygen and Hydrogen | 3.45% |
| 180-degree angle in Hydrogen-Carbon-Fluorine | 3.16% |
| Double bonds between Oxygen and Nitrogen | 3.03% |
| Number of Nitrogen atoms | 2.99% |
| Double bonds between Carbon and Oxygen | 2.78% |
| 120-degree angle in Hydrogen-Carbon-Hydrogen | 2.73% |
| Number of Oxygen atoms | 2.64% |
| 180-degree angle in Carbon-Carbon-Oxygen | 2.62% |
| Single bonds between Carbon and Oxygen | 2.37% |

chine-learned matrix *A*. The stronger a feature's weight in *A*, the more this feature affected students' similarity judgments. Hence, the feature's weight corresponds to its saliency in students' perception of molecule representations.

Table 1 shows the 10 most important features, as determined by a ranking of features according to their aggregate weight computed from matrix *A*. These results show that the most highly ranked feature is the number of distinct letters, which corresponds to an aggregate educated guess feature. Specific visual features that are relevant to organic molecules were also ranked highly (e.g., the number of single bonds between Oxygen and Hydrogen atoms, the number of bonds between Carbon and Oxygen, the number of Nitrogen and Oxygen atoms). These specific visual features were present in many of the molecules in our dataset. Several visual features also included geometric aspects, specifically bond angles. These features indicate the presence of chemical functional groups that are relevant to predicting molecule's reactive behaviors.

### *4.1.2 Approach 2: Ordinal Embedding*

Recall that approach s learns an embedding that represents the similarity of molecule representations as distances in a d-dimensional space, from which we then extracted the most important features. First, we established how many dimensions we need to consider (i.e., which d to choose in representing similarity of molecule representations in a d-dimensional space). Using the process of 50-fold cross validation described above, we calculated unit through 20 dimensional embeddings of perceptual similarity. We used 20,000 triplets in this computation to ensure that the number of triplets did not affect the prediction accuracy as the dimension became large. Figure 4 shows that there is no drop in prediction accuracy when embedding in low dimensions versus high, suggesting that perceptual similarity can be accurately represented in a low dimensional subspace, and that there is a high degree of redundancy in the data. This result shows that students' responses agreed on the relative similarity about 70% of the time.

Next, we generated a 2-dimensional embedding that describes students' perceived similarity between the molecule representations. Figure 5 shows this embedding, illustrating that molecules naturally form clusters based on their perceptual similarity. These clusters correspond to specific chemical properties shared among the molecules, such the presence of a particular type of bond or a functional group. We color-coded and labeled some of these clusters to illustrate these characteristics of students' perceptions. This illustration lends face validity to our embedding approach.

From this embedding, we extracted an ordered list of the feature pairs that best capture students' similarity judgments, shown in Table 2. The feature pairs in this table were ranked based on how well they approximate the similarity matrix computed from the

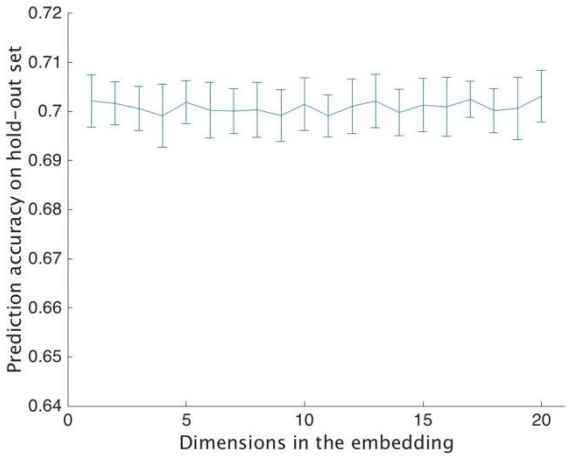


**Figure 4. Prediction accuracy on hold-out set by number of dimensions in embedding.**

**Table 2. Top 10 feature pairs from Approach 2. Each row corresponds to a pair of feature vectors ranked in accordance with how accurately they described the observed similarity structure from the embedding.**

| Rank | Feature pairs | | |
|---|---|---|---|
| 1 | Distinct letters | & | Distinct letters |
| 2 | Total letters | & | Distinct letters |
| 3 | Distinct letters | & | Single bonds |
| 4 | Total bonds | & | Distinct letters |
| 5 | Distinct letters | & | Carbons |
| 6 | Hydrogens | & | Distinct letters |
| 7 | Total letters | & | Total letters |
| 8 | Total letters | & | Single bonds |
| 9 | Total letters | & | Unbonded electrons |
| 10 | Distinct letters | & | Single Carbon-Hydrogen bonds |

embedding in Figure 5. The same feature may appear twice in a pair to account for the possibility that a weighted combination of a feature with itself better reflects the observed similarity structure than does a pair of features. In sum, these results show that the most highly ranked features are general visual features, which correspond to the aggregate educated guess features (e.g., number of letters, number of lines). Specific visual features that are relevant to hydrocarbon molecules were also ranked highly (e.g., the number of Carbon and Hydrogen atoms). These specific features were present in many of the molecules in our dataset.

### *4.1.3 Comparing the Similarity Learning Approaches*

While both methods agreed upon the top ranked feature, the similarity learning by ranking approach ranked structural features of the representations that were relevant to hydrocarbons and organic molecules more highly. As the ranking from this method follow predictive power, this ranking indicates that students' judgments of similarity can best be predicted, and therefore explained, through a combination of the number of different letters and the structural features involving Carbon, Hydrogen, and Oxygen.

## 4.2 Comparison with "Educated Guesses"

To address *research question 2* (do the visual features we identified as salient via metric learning correspond to visual features that students are expected to attend to?), we compared the results from the similarity learning approaches to the educated guess features that we had determined based on the expert-novice litera-

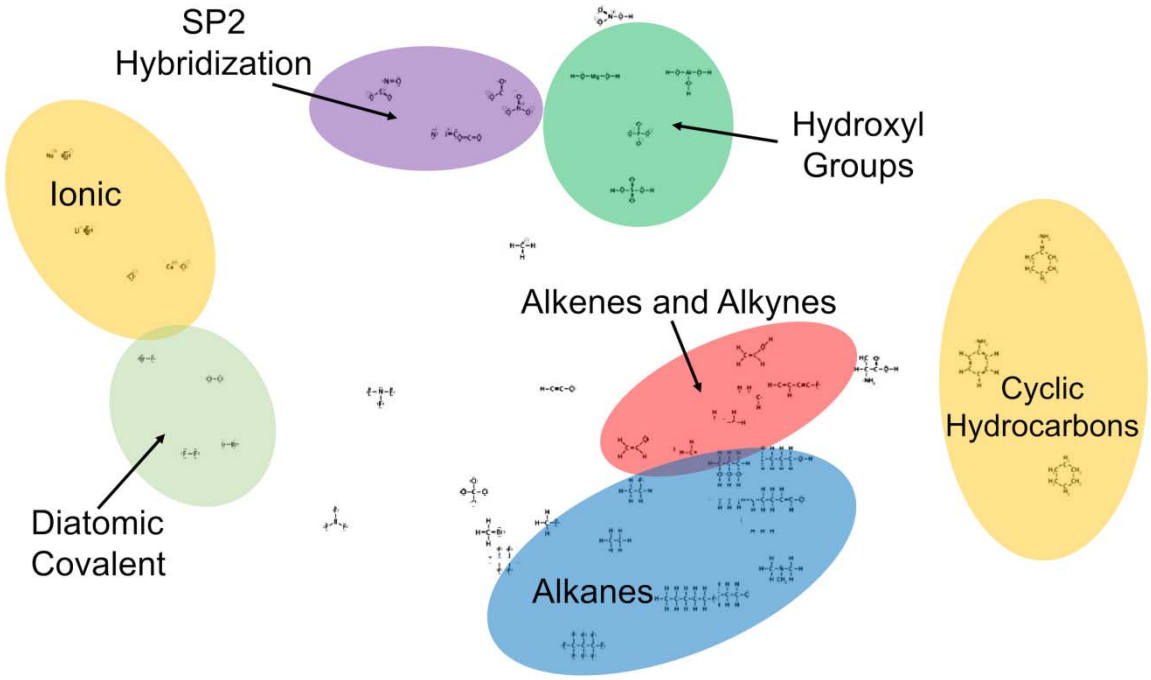


**Figure 5. 2-dimensional similarity embedding from Approach 2. Distances between molecule representations correspond to students' perceptions of dissimilarity (i.e., molecule representations that are depicted close to one another are perceived to be similar).**

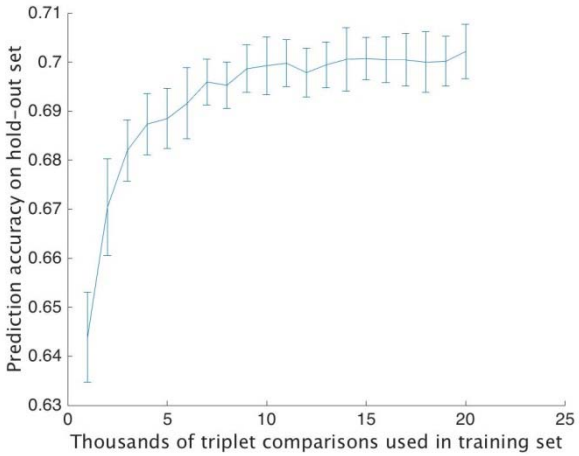


**Figure 6. Prediction accuracy on hold-out set by number of triplet comparison judgments used in the training set.**

ture on perceptual learning. Overall, the results from both metric learning approaches agree with the educated guesses: aggregate features that describe general visual features were ranked to be most important by both metric learning approaches. The similarity learning by ranking approach also yielded a number of visual features that are specific to the types of molecules in our corpus; in particular, visual representations that are highly relevant for comparing organic molecules.

## 4.3 Number of Similarity Judgments Needed

We addressed our methodological *research question 3* (how many similarity judgments we need to assess students' perceptual knowledge) with the ordinal embedding approach. Specifically, we tested how many triplet comparisons are required to compute a representative embedding of the underlying similarity. Figure 6 shows that gains in prediction accuracy of the embedding were no longer statistically significant beyond 7000 triplet comparisons.

## 4.4 Differences Between the Two Approaches

The two methods are different and potentially complementary. There is no definitively correct way to fit the common model $S_{ij} = \boldsymbol{x}_i^T \boldsymbol{A} \boldsymbol{x}_j$ to data. The main differences in the final rankings they produce stems from how we are learning matrix $A$ and the restrictions we put on its structure. In approach 1 we are directly working with triplet responses which are perhaps noisy due to disagreements in students' individual judgments of perceptual similarity, but we are placing fewer restrictions on the learned matrix, allowing for more feature interaction. In approach 2, NMDS is useful for capturing perceived similarity in aggregate, but we enforce much stronger restrictions on the structure of $A$, namely that only two features may interact at once, giving a clearer picture of the importance of a pair of features.

If we had to recommend one approach, we prefer the regression approach (approach 1) because it optimizes prediction error, which is an objective measure of model quality. The embedding approach (approach 2) has its own potential virtues: The low-dimensional embedding provides an implicit form of regularization that may be helpful especially if the amount of response data is small. Also, the embedding provides a visual representation of perceptual similarities which is helpful for model interpretation.

# 5. DISCUSSION

We applied similarity learning approaches to assess which visual features students focus on when presented with visual representations. We compared two approaches, one that allows us to assess the predictive power of the identified features, and one that allows representing the perceived similarity in a d-dimensional space. Both approaches yield similar results as to which visual features

are salient to students. Hence, both approaches address research question 1: Which visual features do students focus on when presented with visual representations? We found that students' similarity judgments of Lewis structures appear to be driven by general visual features such as the number of total and distinct letters, as well as by visual features specific to the types of molecules in our dataset (e.g., number of Hydrogen / Carbon atoms).

Our results also address research question 2: Do the visual features we identified as salient via similarity learning correspond to visual features that students are expected to attend to based on the expert-novice literature on perceptual learning? We found that the identified general visual features align with educated guesses based on the literatures on expertise and perceptual learning, which validates the common "educated guess" approach that instructional designers have to rely on in the absence of assessments of perceptual knowledge. Our results also suggest that, in addition to these general features, students learn to pay attention to key visual features that are highly domain-specific; such as features that indicate the presence of functional groups that are predictive of chemical behaviors. Furthermore, our results show that a few key features predict students' perceptions of similarity between visual representations with accuracy of about 70%.

Finally, we addressed our methodical research question 3: How many similarity judgments we need to assess students' perceptual knowledge? Our results show that about 7,000 responses to triplet comparison tasks are sufficient in assessing a population's perceptual knowledge. Using a survey with 50 triplet comparison tasks (as in our experiment), that means an N of 140 participants will yield valid assessments of perceptual knowledge.

# 6. LIMITATIONS

Although both similarity learning approaches had rigorous theoretical backing, we made a few assumptions about our triplet comparison data that had inherent limitations of note. In both of these methods, we are not modelling individual students, but rather the population as a whole. Consequently, we assume that the triplets and therefore the judgments of similarity are independent of one another. This assumption allows us to learn the rankings of features and feature pairs for the students' collectively, but it does not provide a ranking for an individual. Further, because judging similarity representations is a subjective task, students' judgments may in certain cases conflict with one another. Even with an extremely large number of similarity judgments, complete consensus is unlikely, and therefore, perfect prediction of student judgments is similarly difficult to achieve. Hence, future research needs to investigate how to expand the present approach to modeling individual perceptual knowledge.

Another limitation pertains to the ordinal embedding procedure. For visualization purposes, we embedded the molecules into a 2-dimensional space. Higher dimensional embedding may more accurately capture perceptual dissimilarities. Future research should explore this question.

# 7. FUTURE DIRECTIONS

We will expand our research to other types of visual representations typically used in chemistry instruction (see Figure 1). Further, we will gather data from expert chemists and compare them to data from novices and advanced learners. Based on this comparison, we will identify a "perceptual knowledge gap" between students and experts. Specifically, we will identify visual features that experts attend to but students do not.

Further, we will expand similarity learning so that it can assess an individual student's perceptual knowledge in real time. The current approach is limited in that it requires a large number of similarity judgments to assess students' perceptual knowledge, which is only feasible if we are interested in assessing perceptual knowledge of a population of interest (e.g., novices, advanced students, experts), and because we assume independence among similarity judgments. To address this limitation, we will combine our similarity learning approach with cognitive modeling methods (e.g., Bayesian knowledge tracing). For example, a similarity judgment survey may provide a prior for in a cognitive model, and students' performance on perceptual learning tasks may inform the choice of representations for a small number similarity judgment tasks interspersed in the learning activity.

This expansion will provide the basis for the design of adaptive instruction for perceptual knowledge that can provide appropriate sequences of perceptual learning tasks that draw students' attention to visual features they yet have to learn. Further, knowing which visual features students have not yet learned can serve as a basis for the design of visual feedback that highlights visual features when students make mistakes on perceptual learning tasks.

In sum, we will use the similarity learning approach described in this paper both to design instruction for perceptual learning and to assess perceptual knowledge as a learning outcome.

# 8. CONCLUSIONS

This paper described a new approach to assess students' perceptual knowledge. We used this approach to validate the "educated guesses" approach. In addition, we offer more formal pathways for instructional designers to create perceptual learning assessments. Because developing adaptive instruction for perceptual knowledge relies on such assessments, this paper makes an important contribution to cognitive modeling research.

This paper also makes important contributions to machine learning. We provide a new mathematical approach to quantify the accuracy of perceptual embeddings learned from similarity judgments. Specifically, we derived bounds on the accuracy of embeddings learned from small numbers of comparative judgments by adapting recently developed large-sample analysis methods [34]. This approach provided new algorithms for generating embeddings that are provably accurate. We investigated new methods for embedding based on spectral methods inspired by spectral ranking algorithms [35]. Our experiment yielded an empirical validation with perceptual data from undergraduates, as well as new machine learning methods to assess how visual features predict or encode perceptual similarity judgments. Specifically, we explored the application of group Lasso algorithms for automatically selecting the most perceptually salient features [36]. Our experiment empirically evaluated the group Lasso approach.

In sum, our work provides a crucial stepping stone towards adaptive instruction for perceptual knowledge. Perceptual knowledge is by definition implicit and does not lend itself to the kinds of techniques used in traditional cognitive modeling approaches (e.g., think-alouds, interviews). We presented and evaluated two similarity learning approaches that can determine which visual features students attend to when perceiving visual representations.

# 9. ACKNOWLEDGMENTS

We thank Professor John Moore for his help in recruiting participants for this study, and the LUCID group for their suggestions.

# 10. REFERENCES

[1] Ainsworth, S.: 'The educational value of multiple-representations when learning complex scientific concepts',

in Gilbert, J. et al. (Eds.): 'Visualization' (Springer, 2008), pp. 191-208

[2] NRC: 'Learning to think spatially' (National Academies Press, 2006)

[3] Wertsch, J., & Kazak, S.: 'Saying more than you know in instructional settings', in Koschmann, T. (Ed.): 'Theories of Learning and Studies of Instructional Practice' (Springer, 2011), pp. 153-166

[4] Airey, J., & Linder, C.: 'A disciplinary discourse perspective on university science learning', J. of Research in Science Teaching, 2009, 46, pp. 27-49

[5] Dreher, A., & Kuntze, S.: 'Teachers facing the dilemma of multiple representations being aid and obstacle for learning', Journal für Mathematik-Didaktik, 2014, pp. 1-22

[6] Ainsworth, S.: 'DeFT: A conceptual framework for considering learning with multiple representations.', Learning and Instruction, 2006, 16, pp. 183-198

[7] Gilbert, J.: 'Visualization: A metacognitive skill in science and science education', in Gilbert, J.K. (Ed.): 'Visualization in science education' (Springer, 2005), pp. 9-27

[8] Schnotz, W.: 'An integrated model of text and picture comprehension', in Mayer, R.E. (Ed.): 'The Cambridge Handbook of Multimedia Learning' (Cambridge University Press, 2005), pp. 49-69

[9] Koedinger, K., & Corbett, A.: 'Cognitive Tutors: Technology bringing learning sciences to the classroom', in Sawyer, R. (Ed.): 'Cambridge Handbook of the Learning Sciences ' (Cambridge University Press, 2006), pp. 61-77

[10] VanLehn, K.: 'The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems', Educational Psychologist, 2011, 46, (4), pp. 197-221

[11] Tuckey, H., Selvaratnam, M., & Bradley, J.: 'Identification and rectification of student difficulties concerning three-dimensional structures, rotation, and reflection', J. of Chemical Education, 1991, 68, pp. 460-464

[12] Davidowitz, B., & Chittleborough, G.: 'Linking the macroscopic and sub-microscopic levels: Diagrams', in Gilbert, J., and Treagust, D. (Eds.): 'Multiple representations in chemical education' (Springer, 2009), pp. 169-191

[13] Seufert, T.: 'Supporting coherence formation in learning from multiple representations', Learning and Instruction, 2003, 13, pp. 227-237

[14] Kellman, P.J., & Massey, C.M.: 'Perceptual Learning, cognition, and expertise', The Psychology of Learning and Motivation, 2013, 558, pp. 117-165

[15] Massey, C., Kellman, P., Roth, Z., & Burke, T.: 'Perceptual learning and adaptive learning technology', in Stein, N., and Raudenbush, S. (Eds.): 'Developmental cognitive science goes to school' (Routledge, 2011), pp. 235-249

[16] Kellman, P., Massey, C., & Son, J.: 'Perceptual learning modules in mathematics.': 'Topics in Cognitive Science' (2009), pp. 285-305

[17] Anderson, J., Boyle, C., Corbett, A., Lewis, M.: 'Cognitive modeling and intelligent tutoring' (MIT Press, 1990)

[18] Rau, M., Aleven, V., Rummel, N., & Rohrbach, S.: 'Why interactive learning environments can have it all: Resolving design conflicts between conflicting goals': 'Proceedings of SIGCHI 2013' (ACM, 2013), pp. 109-118

[19] Clark, R., Feldon, D., Van Merrienboër, J., Yates, K., & Early, S.: 'Cognitive task analysis', Spector, J. et al. (Eds.): 'Handbook of research on educational communications and technology' (Lawrence Erlbaum, 2007), pp. 577-593

[20] Schraagen, J., Chipman, S., & Shalin, V.: 'Cognitive Task Analysis' (Erlbaum Associates, 2000)

[21] Koedinger, K., Corbett, A., & Perfetti, C.: 'The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning', Cognitive Science, 2012, 36, pp. 757–798

[22] Rappoport, L., & Ashkenazi, G.: 'Connecting levels of representation: Emergent versus submergent perspective', Int. J. of Science Education, 2008, 30, (12), pp. 1585-1603

[23] Talanquer, V.: 'On cognitive constraints and learning progressions: The case of "structure of matter"', Int. J. of Science Education, 2009, 31, (15), pp. 2123-2136

[24] Goldstone, R., Landy, D., & Son, J.: 'The education of perception', Topics in Cognitive Science, 2010, 2, pp. 265-284

[25] Rau, M.: 'Multi-methods approach for domain-specific grounding: An ITS for connection making in chemistry', Under review at IEEE TLT.

[26] Jamieson, K., Jain, L., Fernandez, C., Glattard, N., & Nowak, R.: 'NEXT: A system for real-world development, evaluation and application of active learning': 'Advances in Neural Information Processing Systems' (2015), pp. 2638-2646

[27] Stewart, N., Brown, G., & Chater, N.: 'Absolute identification by relative judgment', Psychological Review, 2005, 112, pp. 881–911

[28] Bijmolt, T., & Wedel, M.: 'Effects of alternative methods of collecting similarity data for multidimen-sional scaling', Int. J. of Research in Marketing, 1995, 12, pp. 363–371

[29] Agarwal, S., Wills, J., Cayton, L., Lanckriet, G., Kriegman, D., & Belongie, S.: 'Generalized non-metric multidimensional scaling': 'Proceedings of the 12th Int. Conference on Artificial Intelligence and Statistics' (2007)

[30] Johnson, R.: 'Pairwise nonmetric multidimensional scaling', Psychometrika, 1973, 38, (1), pp. 11–18

[31] Tamuz, O., Liu, C., Belongie, S., Shamir, O., & Kalai, A.: 'Adaptively learning the crowd kernel': 'Proceedings of the 28th Int. Conference on Machine Learning' (2011)

[32] Chechik, G., Sharma, V., Shalit, U., & Bengio, S.: 'Large scale online learning of image similarity through ranking', Journal of Machine Learning research, 2010, pp. 1109–1135

[33] Atzmon, Y., Shalit, U., & Chechik, G.: 'Learning sparse metrics, one feature at a time', Journal of Machine Learning Research, 2015, 1, pp. 1-48

[34] Arias-Castro, E.: 'Some theory for ordinal embedding', arXiv preprint arXiv:1501.02861, 2015

[35] Negahban, S., Oh, S., & Shah, D.: 'Iterative ranking from pair-wise comparisons': 'Advances in Neural Information Processing Systems' (2012), pp. 2474-2482

[36] Yuan, M., & Lin, Y.: 'Model selection and estimation in regression with grouped variable', J. of the Royal Statistical Society: Series B, 2006, 68, (1), pp. 49–67