Learning of Relational Categories as a Function of Higher-order Structure

Daniel Corral (daniel.corral@colorado.edu) & Matt Jones (matt.jones@colorado.edu)
Department of Cognitive Psychology, University Colorado Boulder,
Boulder, CO 80309 USA

Abstract

Higher-order relations are important for various cognitive tasks, such as analogical transfer. The current study tested people's ability to learn new relational categories, using a learning test of pure higher-order relations. Each stimulus consisted of 4 objects varying on 3 dimensions. Each category was defined by three binary relations between pairs of objects, producing six logically different conditions. Every category was composed of the same number of relations, but differed in the manner that the relations were linked (i.e., by operating on shared objects). Various learning models were compared and the significance of their performance on the experimental task is discussed. The current findings may advance understanding of the cognitive mechanisms involved in relational learning and the manner in which people naturally represent higher-order relational structures.

Keywords: higher-order relations; schema refinement; schema elaboration; structure acquisition.

Introduction

The ability to generalize and transfer knowledge from a given problem to an analogous task has been of great interest to cognitive scientists and has led to an extensive amount of research. The large body of work on analogical transfer has converged on the idea that transfer is driven by discovering the common relational structure between two analogous scenarios (Gentner, 1983; Gick & Holyoak, 1983). Penn, Holyoak, and Povinelli (2008) posit higher-order relations are critical for most other higher cognitive processes as well, including inference, causal reasoning, and theory of mind. Nevertheless, there is little understanding of the cognitive mechanisms that subserve learning and recognition of higher-order relations.

The purpose of the current study is to explore how people learn different higher-order relations. We define a *higher-order relation* to be a system of first-order (i.e., primitive) relations operating on a common set of objects. Different higher-order relations differ in how the first-order relations are linked together by shared role-fillers. We report an experiment using a relational category-learning task, in which each subject learned a category defined by a higher-order relation, by learning to distinguish category members from non-members. The category in each experimental condition was defined by three binary relations among four objects. In the spirit of Shepard, Hovland, & Jenkins' (1961) classic study on learning feature-based categories, we conduct an exhaustive comparison of the six logically different categories of this type.

The dominant view of how people acquire relational concepts is schema refinement (e.g., Doumas, Hummel, & Sandhofer, 2008), but the present results highlight a number of conceptual problems with this approach. As an alternative, we introduce schema elaboration as a mechanism that is more psychologically plausible and better able to match human performance. We consider four variants of schema elaboration, motivated by different theoretical perspectives, and compare their ability to predict the relative learnability of different higher-order relations.

Structure-Mapping Theory

Since its initial proposal (Gentner, 1983), structure-mapping theory has provided a great deal of insight into the process of analogical learning and transfer. Structure-mapping theory posits that analogy involves aligning the relational structures of two scenarios. A relational structure is composed of multiple first-order relations that are linked together in a specific manner (i.e., the manner in which they operate on shared objects). Consider the classic solar system-atom analogy (Figure 1): Planets revolve around the sun, and planets are smaller than the sun; electrons revolve around the nucleus, and electrons are smaller than the nucleus. Although the same first-order relations are present in both scenarios (i.e., *smaller than* and *revolves around*), the analogy only works because the first-order relations share objects in the same way, such that their first roles are filled by the same object (i.e., planet and electron). In structure-mapping theory, this property is formally known as *parallel connectivity* (Gentner, 1983).

Thus, analogy can be viewed as the recognition that two scenarios are instances of the same higher-order relation, that is, first-order relations connected in the same manner. When the same first-order relations are present in two or more scenarios, but are shared differently between objects, different higher-order relations are formed. Hence, to successfully transfer between two analogous scenarios, people must learn the exact manner in which the first-order relations are connected to form a specific higher-order relation.
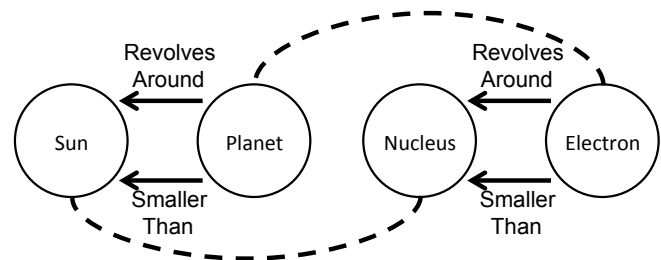

Figure 1. Diagram of solar-system–atom analogy.

Schema Refinement

Formation of an analogy has been proposed to lead to induction of a schema, an abstract representation that captures the relational structure common to both analogues (Gick & Holyoak, 1983; Hummel & Holyoak, 2003; Kuehne, Forbus, Gentner, & Quinn, 2000). Subsequent analogy between a schema and a new episode can result in a new schema (replacing or supplementing the original schema) that contains only the structure that is common to the original schema and the new episode. This process is referred to as *schema refinement*, and it has been proposed to operate by a mechanism of *intersection discovery* (Doumas et al., 2008). An analogy between two episodes may lead to a "dirty" schema that includes idiosyncratic properties common to both episodes but not universal to other instances of the abstract concept being acquired (Doumas et al., 2008; Hummel & Holyoak, 2003). Through comparison to successive instances of the abstract concept (as they are encountered), the schema can be refined to contain only information that belongs to the concept.

As a model of relational learning, schema refinement has several shortcomings. First, because refinement models only allow for a schema to decrease in size, the model cannot add new information. Consequently, upon its first encounter with a member of a relational category, the model must retain all information contained in the exemplar, as it may be necessary in defining the category. Contrary to the results presented below, this assumption leads to a prediction of no false alarms during learning of a relational category. A false alarm can only occur when a subject's current schema is missing relational constraints for what constitutes category membership. Under an idealized model of pure schema refinement, there is no way for the model to delete relations that are present in all category members.

A related assumption of schema refinement is that the model can start off with and maintain highly complex schemas. Given the processing constraints of working memory (Baddeley, 2003), such an assumption seems psychologically implausible. Instead, subjects should be expected to quickly forget a large amount of the information that was initially processed. When the number of objects and predicates contained within a higher-order relation exceeds the processing capacity of working memory, schema refinement may not accurately reflect how the concept is acquired., Thus, we propose that a more complete model of relational learning must incorporate forgetting, and, consequently, the ability to add information to the schema rather than only simplifying it.

Schema Elaboration

When a learning model contains processing constraints similar to those of working memory, the ability to elaborate upon a schema (i.e., to add new information) may be better suited than schema refinement alone for the acquisition of higher-order relations. We propose that in cases where a schema is insufficiently complex (i.e., is missing appropriate relational constraints), people are capable of updating their schema by adding new relations. We refer to this process as *schema elaboration*.

Because the pure schema refinement model is unable to add new information, the model has no room for error if true relational constraints are mistakenly discarded. This makes schema refinement an unrealistically rigid learning model. The schema elaboration model described in detail below is more flexible, as it is able to reincorporate information that it has mistakenly discarded or forgotten.

Experiment

The current study investigated people's ability to learn arbitrary new higher-order relations. The study used a standard category-learning paradigm, with an A/not-A design in which subjects were asked to decide whether each stimulus did or did not belong to the category. The category to be learned was manipulated between subjects. The category structures all contained the same number and types of first-order relations but differed in how those relations were connected (i.e., in the higher-order relation they formed). We aimed to test models of relational concept acquisition by assessing how this manipulation of higher-order structure affects learning.

Figure 2 shows an example stimulus. Each stimulus comprised four objects, known in the literature as *Shepard circles*, arranged in a square configuration. The objects varied along three separable dimensions: brightness, size, and radius tilt. Each dimension had four levels, assigned without replacement to the four objects on each trial.

Each dimension defines a comparative binary relation among the objects (i.e., brighter, larger, steeper). The category to be learned by each subject was defined by three such relations, one on each dimension. Thus, a stimulus was a member of the category if it satisfied all three of these relations (e.g., upper-right object must be larger than upper-left object, lower-left object must be brighter than lower-right object, and lower-right object must have its radius more tilted than upper-left object). The category structures varied in how the relations were connected to each other, in terms of the objects they were defined on. For example, any two relations could operate on the same pair of objects, on disjoint objects, or on one shared object with one unique object for each relation. This design leads to six topologically unique category structures, shown schematically in Figure 3. The manner in which these topological structures were instantiated (i.e., the roles of the four spatial locations) was counterbalanced across subjects within each condition.
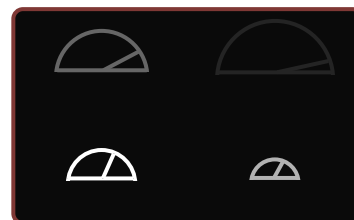

Figure 2. Example stimulus from main task.

Method

137 undergraduates were randomly assigned to six (between subjects) conditions, differing in the category structure to be learned. Subjects were given a cover story in which the stimuli were optical key cards for a building; their task was to learn which key cards would open a door.

To familiarize subjects with each of the first-order relations, they were given three training tasks prior to the main task, one for each first-order relation (i.e., brighter, larger, and steeper). The training tasks were the same as the main task, except that each stimulus contained only two objects instead of four, and each category was defined by only one relation (e.g., right object must be darker than left object). Each training task ended once the subject answered eight consecutive correct responses. The order of training tasks was counterbalanced.

All four tasks followed the same procedure. On each trial, a stimulus was sampled randomly, subject to equal probability of choosing a stimulus in or out of the category. The subject responded by pressing Y or N (indicating the key card does or does not open the door), and then the correct answer was displayed. The instructions for each task indicated the categories were different (i.e., each task was about a different door of the building) and included a random, positive example (i.e., a key card that opens the current door). The full experiment (i.e., training and main tasks combined) was programed to end after 55 minutes.

Figure 3. Diagram of each category condition. Lines connecting objects indicate first-order relations.

Models

Before presenting the results, we describe a series of models that were compared to the data. These models were designed to test the need for augmenting theories of schema refinement with mechanisms of forgetting and schema elaboration. Only the main task was modeled.

Control Models

Three control models—pure refinement, refinement with forgetting, and refinement with forgetting and elaboration— were formulated to provide a baseline for the more sophisticated elaboration models discussed below.

All of the models operate by maintaining a schema from trial to trial that contains some set of relations among the objects within the stimuli. Each stimulus is classified as in the category if it satisfies all relations currently in the schema. The schema is initialized as a complete representation (i.e., all 18 binary relations) of the example stimulus provided in the instructions for the main task.

The pure refinement (PR) model learns only following a miss, meaning a trial on which the stimulus belongs in the category but is mistakenly classified as a nonmember. This can occur when the schema includes relational constraints that are not part of the true category rule. Feedback after a miss causes the schema to be updated (refined) by intersection discovery, discarding all relational constraints the stimulus violates. All other relations in the schema are retained. This learning process will continue until all incorrect relational constraints are removed, at which point the schema will necessarily coincide with the true category rule.

The refinement-with-forgetting (RF) model incorporates processing constraints meant to mimic those of working memory. Soft capacity limitations result in the model losing (i.e., forgetting) relational constraints prior to each trial. Each relation has an independent probability p of being forgotten, which depends on the total number of relations currently in the schema (r):

$$p = 1 - \frac{L}{r}\left(1 - e^{-r/L}\right), \tag{1}$$

where L is a processing-capacity parameter. This formulation has the property that the expected number of retained relations equals $L * (1 - \exp(-r/L))$, that is, exponential approach to some limiting capacity L.

The random elaboration model (RE) includes refinement, forgetting, and elaboration. The interplay between refinement and elaboration leads the model to add and remove constraints one at a time until the schema converges on the true category structure. The forgetting mechanism in the model allows for false alarms, as true relational constraints can be lost, making the schema under-constrained. Indeed, a subject may commit a false alarm if his or her working schema lacks a relational constraint that is part of the category rule.

According to the elaboration assumption, false alarms lead to the appendage of a new relation, which the stimulus satisfies but was not part of the initial hypothesis. This mechanism allows the schema to increase in size and complexity. Unlike with misses, after receiving feedback of a false alarm the subject does not know which relational constraints must be added (i.e., which relation in the stimulus constitutes a violation of the rule). Therefore, the model identifies all relations in the stimulus that are absent from the schema, and treats each as a candidate to be added. For simplicity, we assume exactly one relation is added to the schema following any false alarm. In the RE model, this choice is made at random among the candidates.

The PR model is an ideal observer for the present task, and hence performance was expected to be high for all conditions. In the RF model, once a true relational constraint is forgotten, it has no way of being reincorporated into the schema; hence all relations will eventually be lost and the

model should asymptote chance performance. In the RE model, elaboration and forgetting can combine to produce intermediate levels of performance (depending on the capacity parameter L). However, it was expected that none of the three control models would predict any learning differences among the different category conditions. Indeed, all relations are treated independently (except for a global effect of schema size on the forgetting probability), and hence the manner in which relations are linked through operating on shared objects should have no effect on model performance.

The Search for Relational Constraints

The RE model assumes schema elaboration involves random selection of a candidate relation that is in the stimulus but not in the schema. Alternatively, the selection could be preferential, sensitive to higher-order structure. Different assumptions about preferences guiding the relations that are added lead to different models of preferential elaboration, each making unique predictions about the relative learnability of the category structures in the current study. Here we consider four possibilities, motivated by different theoretical perspectives in the literature. Importantly, each model is inspired by the corresponding theoretical perspective but is not meant to be a formal implementation of that theory.

Each of the four models presented below works in the following manner. Following a false alarm, each candidate relation for addition to the schema is assigned a score that determines its probability of being selected. The probability each candidate is selected is given by

$$\frac{e^{\varphi s}}{\sum_{s'} e^{\varphi s'}} \tag{2}$$

where s is the score for the candidate, s' ranges over all candidates, and φ is a parameter determining the degree of stochasticity in the decision process. The models differ in how the scores are determined by higher-order structure.

Conceptual Coherence

Murphy and Medin (1985) proposed that people's lay theories about the world make categories conceptually coherent. One reason for this may be that a theory provides a conceptual filter through which relational information can be processed and organized around. Furthermore, research has shown that features that are central to a concept more strongly influence the concept's conceptual coherence (Sloman, Love, & Ahn, 1998). Taken together, these ideas suggest that a category composed of a central object that participates in all three relations will be more conceptually coherent than other category structures, as the category representation can be organized around the central object, providing a critical conceptual reference point. Thus, performance should be higher for conditions 2, 4, and 6 than for other category structures. To formalize this principle, the score (s) for each candidate relation was defined as the sum,

over the two objects that relation operates on, of how many relations already in the schema each object participates in. This assumption leads the conceptual coherence (CC) model to prefer relations built on already more central objects, thus favoring categories with a centralized structure.

Economy of Objects

Due to the processing constraints of working memory (Baddeley, 2003), it may be easier to discover an analogy that requires mapping fewer objects between scenarios. Therefore, when elaborating a schema, subjects may be inclined to select relations that minimize the total number of objects involved. This hypothesis predicts that learning will be superior for category structures involving a smaller number of objects, such as condition 6 and to a lesser extent 3 and 4. This principle was formalized in the economy of objects (EO) model, by defining the score for each candidate relation as the number of its objects (0, 1, or 2) that participate in other relations already in the schema.

Plurality of Objects

In contrast to the EO model, cognitive load theory (van Merriënboer & Sweller, 2005) suggests that objects that do not participate in any of the category's relations will act as extraneous distractors. Therefore, subjects' attention may be drawn to such objects, making them more likely to add relations on new objects when elaborating the schema. Because people often struggle to recognize surface features as irrelevant information (Cooper & Sweller, 1987), irrelevant objects may place an unnecessary amount of strain on working memory, while concurrently obscuring the category's higher-order structure. Consequently, category structures that contain the greatest number of irrelevant objects (condition 6, followed by 3 & 4) would be most difficult to learn. This principle was formalized in the plurality of objects (PO) model, by defining the score for each candidate relation as the number of its objects (0, 1, or 2) that do not participate in any relations currently in the schema. This scoring rule is opposite that used in the EO model, and the two models are equivalent under a substitution $\varphi \to -\varphi$.

Relational Chaining

Lastly, learning may be better for category structures composed of relations that are chained together (e.g., condition 1), as people may be intuitively inclined to link known relational structure to new objects. Such a preference might arise from a causal learning perspective, in which subjects seek to discover causal chains among the objects. For example, upon learning that object A must be bigger than object B, people may be inclined to test whether object B must be brighter than object C. Thus, structures composed of relations that can be more readily chained together may be easier to acquire. This principle was formalized in the relational chaining (RC) model, by defining the score for each candidate as

$$s = \frac{1}{c_{\min} + 2} + \frac{\frac{1}{c_{\min}+1} - \frac{1}{c_{\min}+2}}{c_{\max} - c_{\min} + 1}, \qquad (3)$$

where $c_{\min}$ and $c_{\max}$ are the counts of relations currently in the schema in which the candidate's two objects participate. This rule was designed to implement a lexicographic preference for small values of $c_{\min}$ followed by small values of $c_{\max} - c_{\min}$. As a special case, the score for relations with $c_{\min} = c_{\max} = 0$ was set to zero, to implement a preference not to add isolated relations. Thus, the ideal candidate is one that extends an existing chain: $c_{\min} = 0$, $c_{\max} = 1$.

Summary of Models

The preceding subsections fully specify the models tested. The categorization response is determined by whether the stimulus satisfies all relations currently in the schema. Refinement follows misses, by intersecting the schema with the stimulus. Forgetting precedes each trial, following Equation 1. Elaboration follows a false alarm, adding a single relation from the stimulus, chosen by Equation 2 (or randomly, in the RE model). The models differ in whether they include forgetting and elaboration, and in the preferences guiding elaboration.

Experiment and Model Results

Subjects varied in how much time they took to learn the training tasks, and hence in how much time they had for the main task. To reduce statistical noise and ensure all subjects had enough time to learn their condition's category structure, a selection criterion was used to exclude subjects who spent over 35 minutes on the training tasks (leaving less than 20 minutes for the main task). Because this criterion is based on events prior to the experimental manipulation, it introduces no bias in estimating differences among conditions. The selection left 104 subjects in the analysis. Of these subjects, the fewest number of trials completed on the main task was 245.

An ANOVA comparing average proportion correct on the first 245 trials across conditions revealed a non-significant trend, $F(5, 98) = 1.83$, $p = .114$, $MSE = .004$. Because of the complexity of the main task, 245 trials may be insufficient for learning. Therefore, we repeated the analysis while excluding the 6 additional subjects who had completed the fewest trials. The remaining 98 subjects all completed over 350 trials. An ANOVA on these subjects' performance on the first 350 trials indicates a significant main effect of category structure, $F(5, 92) = 2.76$, $p = .023$, $MSE = .012$. Figure 4 shows the mean performances by condition, compared to model predictions. Importantly, the ordering among conditions remained unchanged from the initial analysis, and means were nearly unchanged. Finally, the 6 excluded subjects were re-included, with their proportions correct defined based on the number of trials actually completed. The analysis again revealed a significant effect of condition, $F(5, 98) = 2.75$, $p = .023$, $MSE = .012$. Again, the ordering of performance between

conditions was unchanged, and condition means were nearly identical to the previous analyses.

Figure 4 shows the behavioral results and the simulated predictions of all models. Model parameters (as applicable) were the same for all models and were chosen by hand, with $L = 9$ and $\varphi = 10$.

Evaluation of the models in sequence provides support for each of our theoretical proposals (see Table 1). As predicted, the PR model far outperformed the subjects, suggesting the need for some sort of forgetting mechanism in addition to schema refinement. However, the RF model performed nearly at chance, because it eventually forgot all relations, suggesting a further need for some sort of schema elaboration mechanism. The RE model can match subjects' intermediate performance level, but it fails to predict differences among conditions. The last four models predict condition differences because they are sensitive to higher-order structure. However, the differences are all weaker than in the empirical data. The predicted condition differences are greater when L is increased to produce levels of performance (Figure 5, with $L = 40$), but still none of the models reproduces the correct ordering among conditions. Therefore, further work is required to understand exactly how higher-order structure affects relational learning.
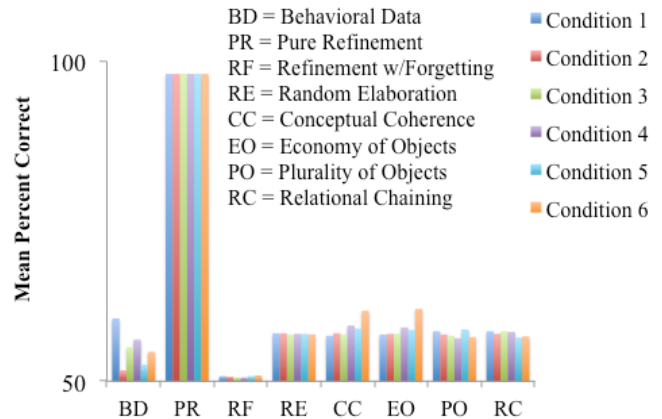

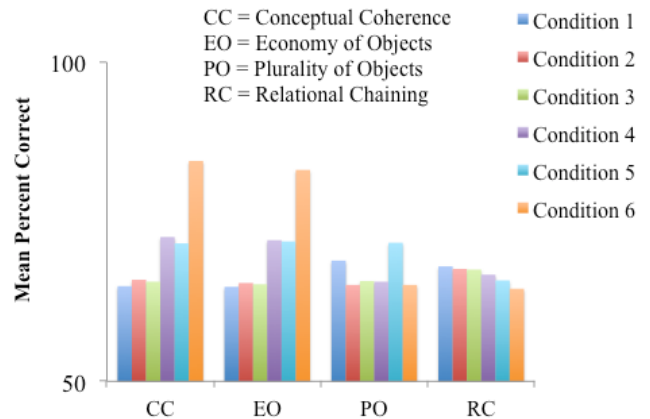*Figure 4*. Mean performance for subjects and models.

Figure 5. Predictions of structure-sensitive models at higher levels of performance, to accentuate condition differences.

Table 1. Strengths (+) and weaknesses (–) of the simulated models.

	Model						
	PR	RF	RE	CC	EO	PO	RC
Performs within the range of subject data	–	–	+	+	+	+	+
Predicts differences across conditions	–	–	–	+	+	+	+
Predicted differences match subjects	–	–	–	–	–	–	–

Discussion

Although previous research has not directly addressed how readily people learn different types of higher-order relations, the acquisition of such concepts is integral to the development of expert representations (Chi, Feltovich, & Glazer, 1981). The current behavioral data suggest that acquisition of higher-order relations is indeed affected by the manner in which the elementary relations within a relational structure are connected. Subjects' performance was best for conditions in which relations could be chained together and where single objects participated in multiple relations (i.e., Conditions 1, 4, and 3).

Although schema refinement has been the dominant model of relational learning (e.g., Doumas et al., 2008), the PR model incorrectly predicts no learning differences across the different category conditions. Further, the model's performance differed dramatically from that of subjects. As expected, when processing constraints were introduced, the RF model failed to retain any of the relational constraints in the category, performing at chance in all conditions. Taken together, these results suggest that schema refinement alone is an insufficient explanation of human relational learning.

The predictions from the elaboration models allow us to address several important issues. That the elaboration models make predictions within the range of the subjects' performance supports the proposal that people employ elaboration mechanisms (in addition to schema refinement) when acquiring higher-order concepts. Additionally, the differences that were found across conditions in the behavioral data provide support for the idea that people indeed have preferences for seeking out certain types of higher-order relations, as formalized in the four structure-sensitive elaboration models.

However, the condition differences predicted by these models were weaker then those exhibited by subjects, and none of the models reproduces the correct ordering across conditions. Thus, it remains an open question as to the specific mechanisms that drive people's search for higher-order relations.

Understanding what drives the differences among the present experimental conditions may provide important theoretical insight into the mechanisms of relational learning, as well as the manner in which people acquire more abstract, higher-order concepts. Such results may also have practical applicability for areas where the recognition of higher-order structures is important for deep learning, such as education, problem solving, and decision-making.

Acknowledgements

References

Baddeley, A. D. (03). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience, 4*, 829-839.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*, 347-362.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1-43.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155-170.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220-264.

Kuehne, S., Forbus, K., Gentner, D., & Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, 770-775.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Penn, D. C., Holyoak, K J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109-178.

Shepard, R. N, Hovland, C. I., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*, Whole No. 517.

Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189-228.

van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147-177.