Title
Learning Conceptual Hierarchies by Iterated Relational Consolidation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN
1069-7977

Authors
Foster, James
Canas, Fabian
Jones, Matt

Publication Date
2012

Peer reviewed

Learning Conceptual Hierarchies by Iterated Relational Consolidation

James M. Foster, Fabián Cañas, & Matt Jones
{james.m.foster, canas, mcj}@colorado.edu
University of Colorado, Department of Psychology & Neuroscience
Boulder, CO 80309 USA

Abstract

Learning new concepts is critical to making sense of the world. Research on analogical reasoning suggests structure mapping and schema induction can enable discovery of new relational concepts. However, existing theories of schema induction and refinement are insufficient to explain acquisition of rich, compositional hierarchies of relational concepts. This paper offers a proposal for this sort of representation construction, founded on reinforcement learning to evaluate the predictive usefulness of higher-order relations, together with a mechanism of relational consolidation by which systems of relations (schemas) can be chunked into unitary entities. A computational model of these ideas is outlined and partially tested in simulations and human experiments. Implications and moderating factors for relational consolidation are considered.

Keywords: Relational Consolidation; Analogy; Schema Induction; Predication; Refinement; Concept Learning

Introduction

Consider a second-order same-different task, in which the subject is presented with two pairs of objects and must recognize whether the pairs match in terms of whether their objects are the same or different. The pairs match if both are instances of *sameness* (e.g., A-A, B-B) or if both are instances of *difference* (e.g., A-B, C-D), and they mismatch if one is an instance of sameness and the other an instance of difference (e.g., A-A, B-C). Thompson, Oden, and Boyson (1997) tested naive chimpanzees on this task and found them unable to learn it, unless they were first trained on a first-order same-different task. In the first-order task, a single pair of objects was presented, and subjects learned to associate sameness and difference to two plastic tokens (e.g., a yellow triangle and a red circle, respectively). Thompson et al. argued this training enabled subjects to reduce the second-order same-different task to a first-order task, by mentally replacing each pair of objects with its associated token, and then determining whether those tokens matched (see also Clark, 2006).

Learning higher-order relations, such as in the second-order same-different task, is arguably critical to abstract conceptual development. In this paper, we argue that many concepts reside in relational hierarchies (relations among relations, and so on), and we investigate how such concepts might be learned. Our basic premises are that much structure in the world (or at least its mental representation) is hierarchically compositional, and that discovering (or creating) this structure is a powerful cognitive mechanism for both learning and designing complex systems.

For example, computer architecture, mathematical functions, and natural languages all exemplify multiple levels of abstraction by chunking systems of relations at one level into building blocks at the next level. In computer architecture, digital logic gates are composed to form adders, which are composed with other digital circuits to form an arithmetic logic unit (ALU), which is a building block in a computer's CPU. Software design manages complexity by continuing this hierarchy, composing primitive functions into more complex functions, and from there to objects and design patterns. The conceptual progression in mathematics proceeds similarly, composing the counting operation to define adding, which is further composed to form multiplying, and then exponentiation. In traditional views of linguistics, phonemes, morphemes, words, and sentences form another example of a relational hierarchy.[1]

These examples motivate our basic research questions. How are relational hierarchies mentally represented? How are these representations learned or constructed through experience? Once a relational concept is learned, how can one discover the higher-order relations in which it can participate?

Here we consider the proposal that much of concept learning is driven by recognizing relational structure through analogy. Research on analogical reasoning has converged on a view that episodes or scenarios are represented as patterns of role binding, in which objects are bound to roles of relations (Gentner, 1983; Hummel & Holyoak, 2003). For example, the fact that the earth revolves around the sun is represented by binding *earth* and *sun* to the first and second roles of a *revolves_around* relation. An analogy between two scenarios constitutes a determination that they share a common pattern of role binding. For example, in the analogy between the solar system and the atom (Gentner, 1983), each system has the property that the object playing role 1 of *revolves_around* is the same as the object playing role 2 of *more_massive_than*.

Analogy formation can thus be viewed as a search for a pattern of linkage among relations (i.e., in terms of how they are bound to shared objects) that holds in two different scenarios. This linkage pattern is a type of higher-order relation among the linked relations. Theories of schema induction (e.g., Doumas, Hummel, & Sandhofer, 2008) offer one way for such higher-order relations to be learned. When an analogy is formed, an abstract schema is created that captures the common structure discovered by the

[1] Relational hierarchies are not taxonomic hierarchies. In a taxonomic hierarchy, each concept or category is a union of lower-level categories. In a relational hierarchy, each *instance* of a concept is a configuration of instances of lower-order concepts.

analogy. The schema can act as a new relational concept, in that it can be analogically aligned with future instances of the higher-order relation it embodies.

Although we agree with theories of schema induction, we argue it is insufficient to explain human relational learning. Schemas are explicit relational structures, and thus they cannot be bound to roles of yet-higher-order relations in the way unitary objects and relations can. The Thompson et al. (1997) study suggests that newly learned relations can only fill roles of other relations if they can be represented as atomic entities. Therefore, to explain acquisition of relational hierarchies, we put forward the hypothesis that useful schemas are eventually replaced (or supplemented) with unitary representations. Thus, a concept that was represented as a system of relations (via the schema) can now be represented as an atomic entity, capable of entering into relations itself. We label this process *relational consolidation*, in a deliberate parallel to theories of episodic memory consolidation (e.g., Squire & Alvarez, 1995).

We further propose that analogy, schema induction, and relational consolidation form a cycle that, when iterated, can produce relational hierarchies of arbitrary depth (height). This form of learning leads to a dualist view of objects and relations, in which (nearly) every concept is both a relational structure among its components and an object capable of participating in relations. The conceptual systems built from this hierarchical relational chunking are potentially quite powerful and flexible.

The remainder of this paper sketches a computational model under development that formalizes these ideas. We report experimental tests and discuss implications of human learning of higher-order relational structures.

Model

We propose a computational model for learning hierarchies of relational concepts, named APEC for Analogy, Predication, Evaluation, and Consolidation. The first two stages (A, P) of the model draw on prior work on analogy and schema induction (Doumas et al., 2008; Forbus, Gentner, & Law, 1995; Larkey & Love, 2003). The last two stages make new proposals for how schemas are selected (E) and consolidated (C) into new concepts. Altogether, the model progresses through parallel processes of analogy formation, predication of relational structure by schema induction, evaluation and refinement of schemas in a reinforcement-learning setting, and consolidation of useful schemas into new atomic relations. Consolidated relations enter into new analogies, allowing the entire learning process to iterate.

The goal of the model is to identify new higher-order relations that are useful for making predictions and guiding behavior. There are an infinite number of higher-order relations that could be learned from any episode, and thus the challenge is selecting those that carry useful information (analogous to the problem of selecting configural cues in feature-based learning; e.g., Gluck & Bower, 1988). The present model addresses this problem in two ways. First, analogical mapping identifies higher-order relations that recur across multiple episodes, to determine which schemas to induce (a form of unsupervised learning). Second, schema evaluation determines how useful each higher-order relation is for predicting outcomes or reward, to determine which schemas to consolidate (a form of reinforcement learning).

The model is currently being implemented within Conway's Game of Life (Gardner, 1970), a cellular automaton exhibiting hierarchical emergent structure, to test its ability to discover that structure. The model produces interesting patterns of schema formation and evolution, which will be reported elsewhere. Here we lay out the model's basic architecture and formulation.

Analogy

APEC represents episodes as systems of role binding among entities, each of which is an instance of a known concept. Every entity is eligible to be bound to a role of one or more other entities, and all entities except primitive objects (used to seed the model) have roles that other entities can bind to. The goal of the analogy component of the model is to find correspondences between episodes that maximally preserve these role-filler relationships (i.e., parallel connectivity).

Formation of an analogy is achieved by a dynamic process of structure mapping. APEC's mapping dynamics are based on a simplified version of the Connectionist Analogy Builder (CAB; Larkey & Love, 2003). For every pair of entities (say, a_i in episode 1 and b_j in episode 2), a mapping weight (m_{ij}) is defined. Mapping weights evolve according to excitatory and inhibitory dynamics. The raw evidence, R_{ij}, for mapping weight m_{ij} is derived by summing the excitation received from all other weights:

$$R_{ij} = \sum_{kl} w_{ijkl} m_{kl}$$

The excitation weight w_{ijkl} equals 1 if m_{ij} and m_{kl} correspond to immediately adjacent and compatible mapping connections (e.g., a_k plays role r in a_i, and b_l plays role r in b_j), and it equals 0 otherwise. The raw evidence is filtered through additional inhibitory mechanisms that encourage one-to-one mappings, and the result determines the incremental change to the mapping weights. These dynamics continue until all mapping weights converge to 0 or 1.

Following the MAC/FAC model of analogical retrieval (Forbus et al, 1995), APEC uses a measure of structural match to determine the quality of an analogy. An initial score is assigned to every matched pair of nodes to enforce a size preference. A preference for deep analogies (systematicity) is implemented via a trickle-down method, whereby initial match scores are passed down to increment the scores of matching components. The match scores are summed to get a global measure of structural match quality.

Predication

If the analogy achieves a minimum match quality, a schema is induced that represents the structural commonalities of the analogues and encodes the shared pattern of role binding embodied by the analogy. Specifically, an abstract entity is created for every mapping weight in the analogy, and these entities are role-bound to each other if the corresponding entities in the analogues are so bound. Once created, schemas are treated identically to episodes (they are just more abstract). This simple mechanism is drawn from prior work on schema induction (Gick & Holyoak, 1983; Doumas et al., 2008; Kuehne et al., 2000). A schema can be thought of as codifying the higher-order relation embodied by the analogy, hence turning it into an explicit predicate. Aligning the schema with any new episode enables a test of whether that episode instantiates the higher-order relation.

Evaluation

When a schema is retrieved and compared to a new episode, it is refined, by abstracting the common structure between schema and episode (Doumas et al., 2008). This process is a form of intersection discovery, where the intersection of the set of relations in a schema and episode are encoded as a new schema. In this way, schemas shrink in size because the variability between episodes is abstracted over, leaving only the structure that is consistent across episodes. However, there may also be a need for schema elaboration, where schemas can grow in size (Corral & Jones, in press). We are currently exploring implementing schema elaboration in the model.

In parallel with schema refinement, schemas are evaluated as candidates for consolidation as new relational concepts. New concepts are useful because they can facilitate generalization. Learning about one instance of a concept can be applied to other instances. Jones & Cañas (2010) show how representations can be learned by improving generalization in a reinforcement-learning framework. The basic idea is that reward prediction error (TD error) can be used to determine when generalization from some past stimulus to the current stimulus was or was not helpful. In the present context, if a learner encounters an episode that is alignable with some stored schema, then analogical inference from that schema enables generalization from past instances of that schema. If this inference leads to improved prediction or behavior, then the schema is strengthened, and if not it is weakened. This process tunes generalization to depend more on higher-order relations that are predictive and less on those that are not.

Consolidation

Relational consolidation is the process of a schema becoming chunked into a unitary concept that can be recognized automatically, retrieved from memory in parallel, and represented as an element of yet-higher-order relations. As summarized in Table 1, consolidation is hypothesized to confer properties to a concept that are not true of (unconsolidated) schemas, because consolidated concepts are recognizable perceptually, without explicit (working-memory dependent) structure mapping.

Table 1. Predicted consequences of consolidation

Not Consolidated	Consolidated
More affected by WM demands	Less affected by WM demands
Quicker at analogical inference, because structure mapping is active	Easier to learn higher-order structure, because instances can be represented by tokens
Serial retrieval	Parallel retrieval

It is important to note that consolidation is not a change in the declarative knowledge embodied by a concept. Rather, it is a proceduralization of the concept that enables future changes in knowledge – similar to the interaction between declarative and procedural knowledge in production systems (Anderson & Lebiere, 1998).

The DORA model of relational predication (Doumas et al., 2008) has an operation similar to relational consolidation, in which it recruits a new proposition node to bind lower-order relations. This new node can be bound to roles of yet-higher-order relations, but the relation is still explicitly structured. In contrast, consolidation might be viewed within the DORA framework as enabling the new proposition node at the top of the relational structure to evolve into a new semantic node at the bottom. We conjecture this difference has important implications for recognition and retrieval of instances of the concept.

Relational consolidation is best explained from the perspective of the MAC/FAC model of analogical retrieval (Forbus et al, 1995). MAC/FAC embodies the assumption that verifying the lower-level elements of an episode (i.e., objects and relations) is fast and automatic, whereas verifying relational structure is slower and requires working-memory resources. In the first stage of analogical retrieval (Many Are Called), the target episode is converted to a flat feature vector that is used to probe all episodic memories in parallel. Importantly, the dimensions in the MAC feature vector are predefined, based on the concepts the learner currently knows. Stored episodes that share content (objects and relations) with the target are retrieved, without regard for how those objects and relations are connected by role binding. In the second stage (Few Are Chosen), the episodes retrieved by the MAC stage are filtered by structural alignment to the target. Only those episodes that are alignable with the target survive this stage.

From the perspective of the MAC/FAC framework, relational consolidation enables a higher-order relation to be chunked and treated as a dimension of the feature vector used for memory probing. Prior to consolidation, retrieval of instances of a higher-order relation require something like the FAC stage, in which subjects explicitly map between those instances and the schema. Following consolidation, retrieval can rely solely on the MAC stage, thus operating much more rapidly and without requiring working memory. We also propose a similar difference for perceptual recognition of instances of the concept in the

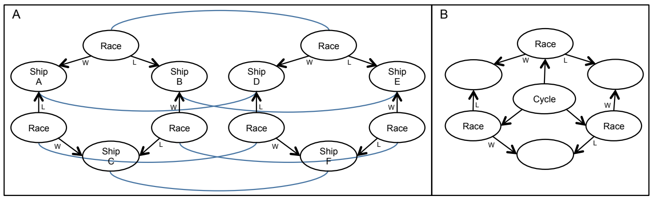

Figure 1: The two ways a predicated relation can be represented or recognized. (A) Before consolidation, episodes must be structurally aligned to a schema. (B) After consolidation, an instance of the concept is explicitly represented and bound to the lower-order relations. The labels on the nodes refer to Experiments 1 & 2.

environment, through the creation of a perceptual detector for each consolidated concept. Figure 1 illustrates the difference between recognizing an instance of a higher-order relation that has versus has not been consolidated.

Experiments 1 & 2

The goal of the present experiments was to test learning of categories defined purely by higher-order relations. That is, the set of objects and relations present in instances of each category were identical; only the way the relations were linked into a higher-order structure differed. If people can learn this type of category distinction, it would support our basic proposal for how higher-order relations are defined in constructing relational hierarchies.

Each trial showed animations of three spaceships racing each other in pairs. The categories were defined by the two logically possible structures these races could form: a *cycle* (e.g., A beats B, B beats C, C beats A) and an *ordering* (e.g., B beats A, B beats C, C beats A). The races are thus first-order relations between spaceships, and cycle and ordering are the possible higher-order relations (see Figure 1 for an example of the cycle structure).

According to APEC, three types of learning potentially contribute to this task. First, analogical alignment between episodes (trials) leads to induction of schemas capturing their common structure or properties. Some of these schemas will capture the true category structure, whereas others will be based on irrelevant information (e.g., ship color or spatial position). Second, feedback following each trial is used to strengthen or weaken schemas that contributed to each response, so that eventually the correct higher-order relations should come to dominate performance. Third, with sufficient learning, the correct higher-order relations may become consolidated. The results reported below primarily bear on the first of these mechanisms.

Methods

110 and 62 undergraduates participated in Experiments 1 and 2, respectively. Subjects were told they would observe alien spaceship races, in sessions of three races each. The aliens were said to have two names for possible outcomes of a session, and the subject's task was to learn their meaning. On each trial, three spaceships (differing only in color) raced in pairs in a sequence of three races. The subject classified the session as "Dekal" or "Koplu" by typing D or K. The correct answer was then displayed. The experiment lasted until the subject met a learning criterion of 8 out of 10 trials correct, or until 25 minutes elapsed (50-70 trials).

Each experiment included two orthogonal manipulations designed to bias attention between objects (spaceships), relations (races), and higher-order relations. In Experiment 1, the trials were described as either "tournaments" or "sessions". In addition, the main task was preceded by a series of footraces among 5 cartoon characters, after which subjects were asked either which character had done best overall or which of two characters had won a specific race. The tournament label and overall-winner question were predicted to make rankings salient, thus shifting attention to higher-order relations.

In Experiment 2, half the subjects were given a first-person perspective by adding a gold star to mark "your ship" on each trial. In addition, the three colors used for the spaceships were either constant or variable from trial to trial. The marked ship and constant colors were predicted to make individual objects more salient, thus shifting attention away from higher-order relations.

Results

There were no differences across conditions in either experiment in meeting the learning criterion or number of trials to criterion. All subsequent analyses are based on collapsing the groups of both experiments.

The results demonstrate that subjects could learn the difference in categories. 113 subjects (65.7%) met the learning criterion. Figure 2A shows the distribution of trials to criterion for these subjects (M = 30.9). Ten subjects learned without making a single error; they were excluded from the remaining analyses, which are based on errors.

Figure 2B shows a backward learning curve, aligned on each subject's last error (for subjects who solved the task but made at least one error). Performance prior to the last error is only slightly above chance, indicating learning was nearly all-or-none. This is consistent with our hypothesis that learning is triggered by inducing the correct schema.

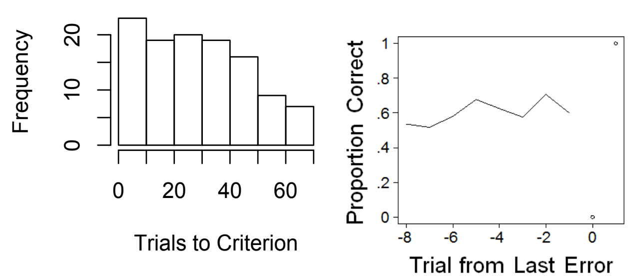

Figure 2. A: Distribution of trials to criterion, for subjects solving the task. B: Backward learning curve.

If learning depends on analogical mapping, then it should be most evident following consecutive stimuli in the same category (assuming subjects most often compare stimuli from consecutive trials). Assuming the schema was induced following the trial of the last error (t_{last}), this predicts the stimuli on trials $t_{last} - 1$ and t_{last} should tend to be of the same category (see Figure 3). This prediction holds for 59 of the 103 subjects (p = .084, one-tailed binomial test). If we relax the all-or-none assumption and examine performance of all subjects on all trials t, we find a highly reliable advantage when trials $t - 2$ and $t - 1$ are of the same category (mean difference 4.3%, paired $t_{169} = 3.42$, $p <$ .001).[2] This suggests comparison of recent stimuli from the same category had a significant effect on performance.

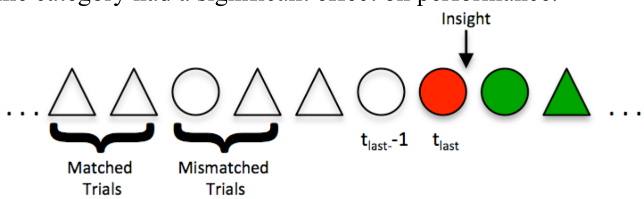

Figure 3. Example sequence of trial types and the most likely moment of schema induction. Triangles indicate ordering trials and circles indicate cycle trials.

Discussion

Analogy and metaphor are pervasive in cognition (Hofstadter, 2001; Lakoff, 1980) and play a critical role in abstract reasoning. The past three decades of research have led to a strong consensus that analogy hinges on recognition of common relational structure between two or more situations (e.g., Gentner, 1983; Hummel & Holyoak, 2003). This suggests that acquisition of new higher-order relations plays a critical role in human conceptual development.

We propose that many (if not most) abstract concepts exist in relational hierarchies, in which entities are at once relational structures among their components and elements of higher-order relations. The structure-mapping process of analogy (Gentner, 1983) can be viewed as a search for relations among relations, in the sense of how relations are connected to one another by operating on the same objects. Successful analogies—those that lead to useful inferences or predictions—might thus be treated as candidates for new relational concepts.

Our approach builds on models of analogical retrieval (Forbus et al., 1995), structure mapping (Falkenhainer, et al., 1989; Hummel & Holyoak, 1996; Larkey & Love, 2003), predication, and refinement (Doumas et al., 2008). Importantly, our model goes beyond previous models of schema induction (Doumas et al., 2008) by positing relational consolidation as a means for learning new relational concepts. A further contribution of our approach is embedding predication within a reinforcement-learning framework in order to modify analogical similarity, and thus generalization, by representation change (Jones & Cañas, 2010; see also Tomlinson & Love, 2006).

Taken together, these ideas lead to a model, APEC, which iterates the Analogy, Predication, Evaluation, and Consolidation stages to build relational hierarchies in a long-term conceptual learning system. Our eventual aim for the model is a system that can autonomously discover useful structure in its environment by construction of these relational hierarchies.

Although the present experimental results indicate a role for analogical comparison and schema induction, we do not have strong evidence here for consolidation. Indeed, we believe it more likely that the categories were learned only as schemas. The experiments do provide a test of one fundamental assumption of the model: that people can learn higher-order relations defined solely by the configuration of shared role binding among lower-order relations.

Future experiments could build the tournaments into third-order structures that require consolidation of the tournament type in order to solve the task. Another future experiment is to test transfer of learned relational structure to an alternate domain with different lower-order relations. The lack of effect of our manipulations suggest it is an open question what factors influence the kind of learning tested in these experiments.

Evidence for relational consolidation could also come from process dissociation between the two modes of representation outlined in Table 1. Other evidence may come from neurological studies. We speculate that relational consolidation is implemented neurally by a process of hippocampal-to-cortical feedforward training, in line with models of episodic memory consolidation (Gluck & Myers, 1993; McClelland, McNoughton, & O'Reilly, 1995). The hippocampus is well suited for storing schemas, which are inherently structured, given the conjunctive and localist nature of hippocampal representations.

Relational consolidation is similar to the career-of-metaphor hypothesis, by which a metaphor is originally an analogical mapping between the base and target domain, but it can become conventionalized so that the target is recognized immediately as an instance of the base category (Gentner et al., 2001). This transition from novel to conventional metaphor resembles the transition from non-consolidated to consolidated relations, in that both involve a transition from recognition via structure mapping to more automatic, perceptual recognition. The major difference is that in the career of metaphor, the base concept was already consolidated, and the conceptual change is a form of sense extension of that base concept. The conventionalization process does not create new concepts; it just extends their meaning. Therefore it does not function to build up relational hierarchies. Nevertheless, the two ideas seem intimitely linked, in that the career-of-metaphor mechanism might play an important role in extending and refining the meaning of concepts after they have been consolidated.

Language is almost certainly an important factor in relational consolidation. Thompson et al.'s (1997) finding

[2] Two additional subjects were excluded because they experienced no alternation trials before meeting the learning criterion.

of chimps learning higher-order relations depended on initial training with material tokens. In humans, words can act as linguistic tokens (Clark, 2006) and have been shown to aid category learning (Lupyan, Rakison, & McClelland 2007). Son, Doumas, and Goldstone (2010) offer two possible roles of language in relational learning: "(1) words invite learners to compare, highlight, and represent relations (the Generic Tokens [GT] hypothesis), and/or (2) words carry semantic cues to common structure (the Cues to Specific Meaning [CSM] hypothesis)" (p. 55). The lack of significant effect of the CSM manipulation in our Experiment 1 could be explained by the cue word appearing only at the start of the experiment, or by the relatively subtle difference in semantics evoked by the cues "tournament" vs. "session". A stronger test of CSM would be to cue subjects with category labels whose meanings structurally match or mismatch the category schema.

Finally, we do not claim that relational consolidation is the only mechanism for acquiring new relational concepts. Research on basic-level objects (Rosch et al., 1976) suggests there are truly primitive object concepts that are not originally constructed as relational systems. Clearly a lot of discovery comes from analyzing the substructure of objects, and that process should be included in any complete model. For example, categories can be induced for objects that fill the same roles (Jones & Love 2007). Although we have been working on concept learning through mechanisms of synthesis, a future goal is to explore how combinations of analytic and synthetic mechanisms of relational learning might be more powerful than both alone.

Acknowledgments

This work was supported by AFOSR Grant FA-9550-10-1-0177 to MJ.

References

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, *10*(8), 370–374.

Corral, D., & Jones, M. (in press). Learning of relational categories as a function of higher-order structure. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.

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), 1–43.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, *41*(1), 1–63.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*(2), 141–205.

Gardner, M. (1970). Mathematical Games - The fantastic combinations of John Conway's new solitaire game "life". *Scientific American, 223*, 120-123.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science*, *7*(2), 155–170.

Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K.J. Holyoak, & B.K. Kokinov (eds.), *The analogical mind: Perspectives from cognitive science*, 199–253.

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

Gluck, M. A. & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227-247.

Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*, *3*(4), 491–516.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*(2), 220.

Jones, M., & Canas, F. (2010). Integrating reinforcement learning with models of representation learning. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology*, *55*(3), 196–231.

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* (pp. 770–775).

Larkey, L. B., & Love, B. C. (2003). CAB: Connectionist analogy builder. *Cognitive Science*, *27*(5), 781–794.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not Just for Talking. *Psychological Science*, *18*(12), 1077.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382–439.

Son, J. Y., Doumas, L. A. A., & Goldstone, R. L. (2010). When Do Words Promote Analogical Transfer? *The Journal of Problem Solving*, *3*(1), 4.

Squire, L. R. & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology, 5*, 169-177.

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (Pan troglodytes) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*(1), 31-43.

Tomlinson, M., & Love, B. C. (2006). Learning abstract relations through analogy to concrete exemplars. *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 2269–2274).