Analogical Reinforcement Learning

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

Abstract

Research in analogical reasoning suggests that higher-order cognitive functions such as abstract reasoning, far transfer, and creativity are founded on recognizing structural similarities among relational systems. Here we integrate theories of analogy with the computational framework of reinforcement learning (RL). We propose a computational synergy between analogy and RL, in which analogical comparison provides the RL learning algorithm with a measure of relational similarity, and RL provides feedback signals that can drive analogical learning. Initial simulation results support the power of this approach.

Keywords: Analogy; Reinforcement Learning; Schema Induction; Similarity; Generalization

Introduction

The goal of the present work is to develop a computational understanding of how people learn abstract concepts. Previous research in analogical reasoning suggests that higher-order cognitive functions such as abstract reasoning, far transfer, and creativity are founded on recognizing structural similarities among relational systems (Doumas et al., 2008; Gentner, 1983; Hummel & Holyoak, 2003). However, we argue a critical element is missing from these theories, in that their operation is essentially unsupervised, merely seeking patterns that recur in the environment, rather than focusing on the ones that are predictive of reward or other important outcomes.

Here we integrate theories of analogy with the computational framework of reinforcement learning (RL). RL offers a family of learning algorithms that have been highly successful in machine learning applications (e.g., Bagnell & Schneider, 2001; Tesauro, 1995) and that have neurophysiological support in the brain (e.g., Schultz et al., 1997). A shortcoming of RL is that it only learns efficiently in complex tasks if it starts with a representation (i.e., a means for encoding stimuli or states of the environment) that somehow captures the critical structure inherent in the task. We formalize this notion below in terms of similarity-based generalization (Shepard, 1987) and kernel methods from statistical machine learning (Shawe-Taylor & Cristianini, 2004). In other words, RL requires a sophisticated sense of similarity to succeed in realistically complex tasks. Psychologically, the question of how such a similarity function is learned can be cast as a question of learning sophisticated, abstract representations.

This paper proposes a computational model of analogical RL, in which analogical comparison provides the RL learning algorithm with a measure of relational similarity, and RL provides feedback signals that can drive analogical learning. Relational similarity enables RL to generalize knowledge from past to current situations more efficiently, leading to faster learning. Conversely, the prediction-error signals from RL can be used to guide induction of new higher-order relational concepts. Thus we propose there exists a computationally powerful synergy between analogy and RL. The simulation experiment reported here supports this claim. Because of the strong empirical evidence for each of these mechanisms taken separately, we conjecture that the brain exploits this synergy as well.

Analogy

Research in human conceptual knowledge representation has shown that concepts are represented not just as distributions of features (cf. Nosofsky, 1986; Rosch & Mervis, 1975) but as relational structures. This relational knowledge includes both internal structure, such as the fact that a robin's wings allow it to fly (Sloman et al., 1998), as well as external structure, such as the fact that a dog likes to chase cats (Jones & Love, 2007). Theories of analogical reasoning represent relational knowledge of this type in a predicate calculus that binds objects to the roles of relations, for example CHASE(DOG,CAT). According to these theories, an analogy between two complex episodes (each a network of relations and objects) amounts to recognition that they share a common relational structure (Gentner, 1983; Hummel & Holyoak, 2003).

At a more mechanistic level, the dominant theory of analogy is *structural alignment* (Gentner, 1983). This process involves building a mapping between two episodes, mapping objects to objects and relations to relations. The best mapping is one that maps objects to similar objects, maps relations to similar relations, and most importantly, satisfies *parallel connectivity*. Parallel connectivity means that, whenever two relations are mapped to each other, the objects filling their respective role-fillers are also mapped together. An example is shown in Figure 1. Parallel connectivity is satisfied here because, for each mapped pair of ATTACK relations (red arrows), the objects filling the ATTACKER role are mapped together (knight is mapped to queen), and the objects filling the ATTACKED role are also mapped together (rook to rook and king to king). Thus structural alignment constitutes a (potentially partial or imperfect) isomorphism between two episodes, which respects the relational structure that they have in common. Importantly, if the search for a mapping gives little emphasis to object-level similarity (as opposed to relation-level similarity and parallel connectivity), then structural alignment can find abstract commonalities between episodes having little or no surface similarity (i.e., in terms of perceptual features).

We propose structural alignment is critical to learning of

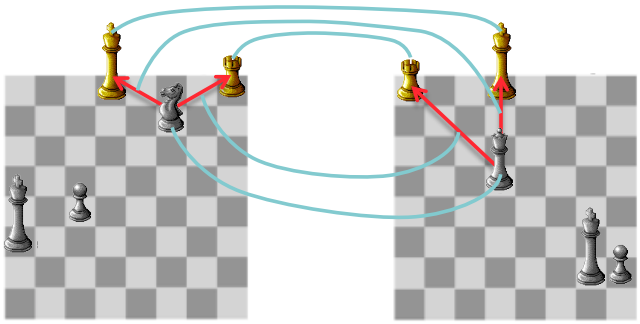

Figure 1: An example of structural alignment between two chess positions. Both positions contain instances of the abstract concept of a FORK: black's piece is simultaneously attacking both of white's pieces. These attacking relations are represented by the red arrows. Cyan lines indicate the mapping between the two episodes. The mapping satisfies parallel connectivity because it respects the bindings between relations and their role-fillers.

abstract concepts for three reasons. First, perceived similarity of relational stimuli depends on structural alignability (Markman & Gentner, 1993). Second, structural alignment is important for analogical transfer, which is the ability to apply knowledge from one situation to another, superficially different situation (Gick & Holyoak, 1980). For example, a winning move in one chess position can be used to discover a winning move in a different (but aligned) position, by translating that action through the analogical mapping. Third, a successful analogy can lead to *schema induction*, which involves extraction of the shared relational structure identified by the analogy (Doumas et al., 2008; Gentner, 1983; Hummel & Holyoak, 2003). In the example of Figure 1, this schema would be a system of relational knowledge on abstract (token) objects, including ATTACK(PIECE1,PIECE2), ATTACK(PIECE1,PIECE3), and potentially other shared information such as NOT_ATTACKED(PIECE1) and KING(PIECE2).

These three observations suggest that analogy plays an important role in learning and use of abstract relational concepts. The first two observations suggest that analogical transfer can be cast as a form of similarity-based generalization, as we elaborate in the next two sections. In brief, structural alignment offers a sophisticated form of similarity that can be used to generalize knowledge between situations that are superficially very different. The third observation suggests that analogy can discover new relational concepts (e.g., the concept of a chess fork, from Figure 1), which can in turn lead to perception of even more abstract similarities among future experiences.

One potential shortcoming of the basic theory of analogy reviewed here is that is it essentially unsupervised. In this framework, the quality of an analogy depends only on how well the two systems can be structurally aligned, and not on how useful or predictive the shared structure might be. For example, one could list many relational patterns that arise in chess games but that are not especially useful for choosing a move or for predicting the course of the game. In previous work, we have found that implementing structural alignment and schema induction in a rich and structured artificial environment results in discovery of many frequent but mostly useless schemas (Foster et al., 2012). An alternative, potentially more powerful model of analogical learning would involve feedback from the environment, so that the value of an analogy or schema is judged partially by how well it improves predictions of reward or other important environmental variables. For example, the concept of a fork in chess is an important schema not (only) because it is a recurring pattern in chess environments, but because it carries information about significant outcomes (i.e., about sudden changes in each player's chances of winning). A natural framework for introducing this sort of reward sensitivity into theories of analogy is that of RL, which we review next.

Reinforcement Learning

RL is a mathematical and computational theory of learning from reward in dynamic environments. An RL task is characterized by an agent embedded in an environment that exists in some *state* at any given moment in time. At each time step, the agent senses the state of its environment, takes an action that affects what state occurs next, and receives a continuous-valued reward that depends on the state and its action (Sutton & Barto, 1998). This framework is very general and can encompass nearly any psychological task in which the subject has full knowledge of the state of the world at all times (i.e., there are no relevant hidden variables).

Most RL models work by learning *values* for different states or actions, which represent the total future reward that can be expected from any given starting point (i.e., from any state or from any action within a state). These values can be learned incrementally, from *temporal-difference (TD) error* signals calculated from the reward and state following each action (see Model section). There is strong evidence that the brain computes something close to TD error, and thus that RL captures a core principle of biological learning (Schultz et al., 1997).

In principle, this type of simple algorithm could be used to perfectly learn a complex task such as chess, by experiencing enough games to learn the true state values (i.e., probability of winning from every board position) and then playing according to those values. However, a serious shortcoming of this naive approach is that it learns the value of each state independently, which can be hopelessly inefficient for realistic tasks that typically have very large state spaces. Instead, some form of generalization is needed, to allow value estimates for one state to draw on experience in other, similar states.

Many variants of RL have been proposed for implementing generalization among states (e.g., Albus, 1981; Sutton, 1988). Here we pursue a direct and psychologically motivated form of generalization, based on similarity (Jones &

Cañas, 2010; Ormoneit & Sen, 2002). We assume the model has a stored collection of exemplar states, each associated with a learned value. The value estimate for any state is obtained by a similarity weighted average over the exemplars' values; that is, knowledge from each exemplar is used in proportion to how similar it is to the current state. This approach is closely related to exemplar-generalization models in more traditional psychological tasks such as category learning (Nosofsky, 1986). It can also be viewed as a subset of kernel methods from machine learning (Shawe-Taylor & Cristianini, 2004), under the identification of the kernel function with psychological similarity (Jäkel et al., 2008).

A critical consideration for all learning models (including RL models) is how well their pattern of generalization matches the inherent structure of the task. If generalization is strong only between stimuli or states that have similar values or outcomes, then learning will be efficient. On the other hand, if the model generalizes significantly between stimuli or states with very different outcomes, its estimates or predictions will be biased and learning and performance will be poor. The kernel or exemplar-similarity approach makes this connection explicit, because generalization between two states is directly determined by their similarity. As we propose next, analogy and schema induction offer a sophisticated form of similarity that is potentially quite powerful for learning complex tasks with structured stimuli.

Analogical RL

The previous two sections suggest a complementary relationship between analogy and RL, which hint at the potential for a computationally powerful, synergistic interaction between these two cognitive processes. We outline here a formal theory of this interaction. The next two sections provide a mathematical specification of a partial implementation of this theory, and then present simulation results offering a proof-in-principle of the computational power of this approach.

The first proposed connection between analogy and RL is that structural alignment yields an abstract form of psychological similarity that can support sophisticated generalization (Gick & Holyoak, 1980; Markman & Gentner, 1993). Incorporating analogical similarity into the RL framework could thus lead to rapid learning in complex, structured environments. For example, an RL model of chess equipped with analogical similarity should recognize the similarity between the two positions in Figure 1 and hence generalize between them. Consequently the model should learn to create forks and to avoid forks by the opponent much more rapidly than if it had to learn about each possible fork instance individually.

The second proposed connection is that the TD error computed by RL models, for updating value estimates, can potentially drive analogical learning by guiding schema induction. Instead of forming schemas for whatever relational structures are frequently encountered (or are discovered by analogical comparison of any two states), an analogical RL model can be more selective, only inducing schemas from analogies that significantly improve reward prediction. Such analogies indicate that the structure common to the two analogue states may have particular predictive value in the current task, and hence that it might be worth extracting as a standalone concept. For example, if the model found a winning fork move by analogical comparison to a previously seen state involving a fork, the large boost in reward could trigger induction of a schema embodying the abstract concept of a fork.

The proposed model thus works as follows (see the next section for technical details). The model maintains a set of exemplars E, each with a learned value, $v(E)$. To estimate the value of any state s, it compares that state to all exemplars by structural alignment, which yields a measure of analogical similarity for each exemplar (Forbus & Gentner, 1989). The estimated value of the state, $\tilde{V}(s)$, is then obtained as a similarity-weighted average of $v(E)$. After any action is taken and the immediate reward and next state are observed, a TD error is computed as in standard RL. The exemplar values are then updated in proportion to the TD error and in proportion to how much each contributed to the model's prediction, that is, in proportion to $\mathrm{sim}(s,E)$.

Additionally, whenever the structural alignment between a state and an exemplar produces a sufficient reduction in prediction error (relative to what would be expected if that exemplar were absent), a schema is induced from that analogy. The schema is an abstract representation, defined on token (placeholder) objects, and it contains only the shared information that was successfully mapped by the analogy. The schema is added to the pool of exemplars, where it can acquire value associations directly (just like the exemplars do). The advantage conferred by the new schema is that it allows for even faster learning about all states it applies to (i.e., that contain that substructure). For example, rather than learning by generalization among different instances of forks, the model would learn a direct value for the fork concept, which it could immediately apply to any future instances. A consequence of the schema induction mechanism is that the pool of concrete exemplars comes to contain more and more abstract schemas. Thus the model's representation transitions from initially episodic to more abstract and conceptual.

Analogical RL thus integrates three principles from prior research: RL, exemplar generalization, and structural alignment of relational representations. Because each of these principles has strong empirical support as a psychological mechanism, it is plausible that they all interact in a manner similar to what we propose here. Thus it seems fruitful to explore computationally what these mechanisms can achieve when combined.

Model

The simulation study presented below uses a variant of RL known as *afterstate learning*, in which the agent learns values for the possible states it can move into (Sutton & Barto, 1998). This is a reasonable and efficient method for the task we use here—tic-tac-toe, or noughts & crosses—because the

agent's opponent can be treated as part of the environment and is the only source of randomness. Our main proposal regarding the interaction between RL and analogical learning is not limited to this approach.

The operation of the model is illustrated in Figure 2. On each time step, the model identifies all possible actions and their associated afterstates. For each afterstate s, it computes an analogical similarity, K, to each exemplar, E, by structural alignment. Each possible mapping $M : s \to E$ is evaluated according to

$$\Phi(M) = \beta \cdot \sum_{o \in s} \mathrm{sim}\,(o, M(o)) + \sum_{r \in s} \mathrm{sim}\,(r, M(r)) \cdot \left[1 + \sum_{i=1}^{n_r} I_{\{M(\mathrm{child}_i(r)) = \mathrm{child}_i(M(r))\}} \right]. \quad (1)$$

This expression takes into account object similarity, by comparing each object o in s to its image in E; relational similarity, by comparing each relation r in s to its image in E; and parallel connectivity, by having similarity between mutually mapped relations "trickle down" to add to the similarity of any mutually mapped role-fillers (Forbus & Gentner, 1989). The sim function is a primitive (object- and relation-level) similarity function, β determines the relative contribution of object similarity, n_r is the number of roles in relation r, $\mathrm{child}_i(r)$ is the object filling the i^{th} role of r, and $I_{\{P\}}$ is an indicator function equal to 1 when proposition P is true. Analogical similarity is then defined as the value of the best mapping (here the θ parameter determines specificity of generalization):

$$K(s, E) = \exp\left(\theta \cdot \max_M \Phi(M) \right). \quad (2)$$

The activation $a(E)$ of each exemplar is determined by normalizing the analogical similarities, and the estimated value of s, $\tilde{V}(s)$, is computed as a similarity-weighted average of the exemplar values $v(E)$ (Figure 2). Thus the estimate is based on the learned values of the exemplars most similar to the candidate state.

Once values $\tilde{V}(s)$ have been estimated for all candidate afterstates, the model uses a softmax (Luce-choice or Gibbs-sampling rule) to select what state to move into (here τ is an exploration parameter):

$$\Pr[s_t = s] \propto e^{\tilde{V}(s)/\tau}. \quad (3)$$

Learning based on the chosen afterstate s_t follows the SARSA rule (Rummery & Niranjan, 1994), after the model chooses its action on the next time step. This produces a TD error, which is then used to update the exemplar values by gradient descent (see Equations for δ and $\Delta v(E)$ in Figure 2).

The model also grows its representation in two ways. First, it begins with no exemplars, and on each trial adds the state it moves to as a new exemplar with probability inversely proportional to the number of exemplars already in the model. This recruitment policy leads the exemplar pool to grow with the square root of time, which seems to give good performance.

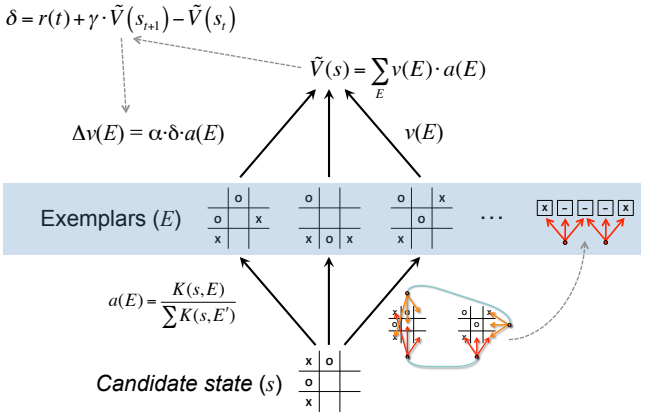

Figure 2: Model operation. Each candidate afterstate is evaluated by analogical comparison to stored exemplars, followed by similarity-weighted averaging among the learned exemplar values. Learning is by TD error applied to the exemplar values. On some trials, especially useful analogies produce new schemas that are added to the exemplar pool. In the example here, s and E both have guaranteed wins for X by threatening a win in two ways. The induced schema embodies this abstract structure. Dots with red arrows indicate ternary "same-rank" relations. r = reward; γ = temporal discount parameter; α = learning rate; other variables are defined in the text.

The more important form of representation learning in the model is schema induction. Schema induction has not been implemented yet, but Figure 2 shows how it is expected to work. Following learning after each trial, the model determines how much each exemplar contributed to reducing prediction error, by comparing δ to what it would have been without that exemplar. If the reduction is above some threshold, the analogical mapping found for that exemplar (lower right of figure) produces a schema that is added to the exemplar pool (far right). The schema is given a value of v initialized at $\tilde{V}(s_t)$. This schema value is updated on future trials just as are the exemplar values. Acquisition of new schemas in this way is predicted to improve the model's pattern of generalization, tuning it to the most useful relational structures in a task.

Simulation

The model was tested on its ability to learn tic-tac-toe. Each board position was represented by treating the nine squares as objects of types 0 (blank), 1 (focal agent's), and 2 (opponent's), and defining 8 ternary "same-rank" relations for the rows, columns, and diagonals. Thus a player wins by filling all squares in any one of these relations. Object similarity was defined as 1 for matching object types and 0 otherwise. Similarity between relations was always 1 because there was only one type of relation. Reward was given only at the end

of a game, as +1 for the winner, -1 for the loser, or 0 for a draw. After the game ended, it moved to a special terminal state with fixed value of 0. For simplicity, all free parameters of the model ($\beta, \theta, \alpha, \gamma, \tau$) were set to a default value of 1.

Three variations of the model were implemented, differing in their levels of analogical abstraction. The Featural model was restricted to literal mappings between states (upper-left square to upper-left square, etc.). This model still included generalization, but its similarity was restricted to the concrete similarity of standard feature-based models. The Relational model considered all 8 mappings defined by rigid rotation and reflection of the board. This scheme was used in place of searching all 9! possible mappings for every comparison, to reduce computation time. Finally, the Schema model extended the Relational model by starting with two hand-coded schemas, 111 and 022. The first of these is a single same-rank relation bound to three instances of the player's own token. Thus moving into a state satisfying this schema produces an immediate win. Likewise, moving into a state satisfying the second schema risks an immediate loss. The model was given no information about these schemas (i.e., v was initialized to 0 for both), but it was capable of learning values for them. The purpose of this model was to test the utility of having schemas that capture task-relevant structures. Logically this question is separate from that of how such schemas are acquired, although we have addressed that question elsewhere (Foster et al., 2012), and we plan to integrate a solution into the present model soon.

Each model variant was trained in blocks of 10 games of self-play followed by a pair of testing games against an ideal player (playing first in one game and second in the other). Learning occurred only during training. In testing games, the model was given one point for each non-losing move it made (i.e., moves from which it could still guarantee a draw), for a maximum of 9 points per pair of testing games.

Average learning curves are shown in Figure 3A for 50 independent copies of each model over 5000 blocks (50,000 training games). Figure 3B shows results for single copies of the Relational and Featural models over 30,000 blocks. These results show that the Featural model does eventually learn, but the Relational model learns an order of magnitude faster, and the Schema model learns another order of magnitude faster than the Relational model.

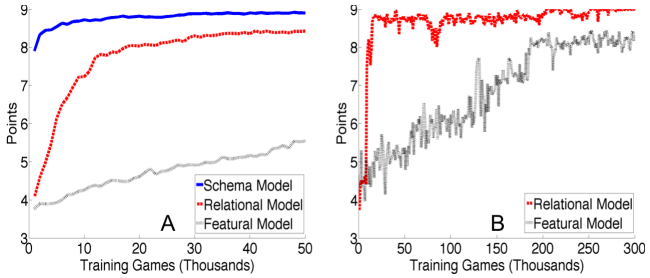

Figure 3: Learning curves. A: 50 copies of each model. B: Single copies of the two slower models over extended training.

Discussion

The results presented here constitute a proof-of-principle that analogy and schema induction can be productively integrated with a learning framework founded on RL and similarity-based generalization. This integration leads to a model exhibiting sophisticated, abstract generalization derived from analogical similarity, as well as discovery of new higher-order relational concepts driven by their ability to predict reward.

The basic modeling framework used here applies not just to analogical similarity and schema induction, but to other forms of representational learning as well. Kernel-based RL offers a powerful and general theory of representation learning, because it can be integrated with any form of representation that yields a pairwise similarity function. Its TD error signal can drive changes in representation via the objective of improving generalization. In previous work, we have applied this idea to learning of selective attention among continuous stimulus dimensions (Jones & Cañas, 2010). The current model offers a richer form of representation learning, in that it acquires new concepts rather than reweighting existing features.

The analogical RL model also builds on other models of relational learning. Tomlinson & Love (2006) propose a model of analogical category learning, with essentially the same similarity and exemplar generalization mechanisms adopted in the present model. Our model adds to theirs in that it applies to dynamic tasks and in that it grows its representation through schema induction. Van Otterlo (2012) has developed methods for applying RL to relational representations of the same sort used here, although the approach to learning is quite different. His models are not psychologically motivated and hence learn in batches and form massive conjunctive rules, with elaborate updating schemes to keep track of the possible combinations of predicates. In contrast, the present approach learns iteratively, behaves probabilistically, and grows its representation more gradually and conservatively. This approach is likely to provide a better account of human learning, but a more interesting question may be whether it offers any performance advantages from a pure machine-learning perspective.

In the present model, the activation of each exemplar elicited by a candidate state can be thought of as a feature of that state. The exemplar effectively has a "receptive field" within the state space, defined by the similarity function. This duality between exemplar- and feature-based representations is founded in the kernel framework (see Shawe-Taylor & Cristianini, 2004). The present model takes advantage of this duality, producing a smooth transition from an episodic, similarity-based representation to a more semantic, feature-based representation defined by learned schemas.

The model as currently implemented does have several limitations. Foremost, it does not yet include a mechanism for inducing new schemas. We and others have shown how schema induction can be successfully deployed in an open-

ended model in a complex environment (Doumas et al., 2008; Foster et al., 2012). We hope that building this type of mechanism into the analogical RL framework will produce a better-controlled, directed system capable of autonomously discovering genuinely new abstract concepts.

A second limitation of the current model is its slowness to learn, due to the nature of gradient descent operating in a large weight space. In contrast, human learning often shows understanding of new concepts in as little as one trial (Maas & Kemp, 2009). The theory of analogy via structure mapping seems like the best candidate for a process-level theory of such rapid learning, and we predict that the full analogical RL model with schema induction will show significant steps in that direction.

The present work is complementary to hierarchical Bayesian models that discover relational structure through probabilistic inference (Tenenbaum et al., 2011). Whereas our model builds up schemas from simpler representations, the Bayesian approach takes a top-down approach, defining the complete space of possibilities a priori and then selecting among them. The top-down approach applies to any learning model, because any well-defined algorithm can always be circumscribed in terms of its set of reachable states. This is a useful exercise for identifying inductive biases and absolute limits of learning, but it offers little insight into the constructive processes that actually produce the learning. These mechanistic questions are critical if the goal is to understand how the human mind discovers new, abstract concepts.

Acknowledgments

Supported by AFOSR Grant FA-9550-10-1-0177 to MJ.

References

Albus, J. S. (1981). *Brains, Behavior and Robotics*. Byte Books.

Bagnell, J. A., & Schneider, J. C. (2001). Autonomous helicopter control using reinforcement learning policy search methods. *IEEE Int Conf Robo*, 1615-1620.

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

Forbus, K. D., & Gentner, D. (1989). Structural evaluation of analogies: What counts. *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, 341–348.

Foster, J. M., Cañas, F., & Jones, M. (2012). Learning conceptual hierarchies by iterated relational consolidation. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 324–329.

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

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychol*, *12*, 306–355.

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

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychon B Rev*, *15*, 256–271.

Jones, M., & Cañas, F. (2010). Integrating reinforcement learning with models of representation learning. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 1258–1263.

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

Maas, A. L., & Kemp, C. (2009). One-shot learning with bayesian networks. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychol*, *25*, 431–431.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen*, *115*, 39–57.

Ormoneit, D., & Sen, S. (2002). Kernel-based reinforcement learning. *Mach Learn*, *49*, 161–178.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychol*, *7*, 573–605.

Rummery, G. A., & Niranjan, M. (1994). *On-line q-learning using connectionist systems* (Tech. Rep. No. CUED/F-INFENG/TR 166). Cambridge University.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

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

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Mach Learn*, *3*, 9–44.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. The MIT Press.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

Tesauro, G. (1995). Temporal difference learning and td-gammon. *Commun ACM*, *38*(3), 58–68.

Tomlinson, M. T., & Love, B. C. (2006). From pigeons to humans: Grounding relational learning in concrete examples. *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, 199–204.

Van Otterlo, M. (2012). Solving relational and first-order logical markov decision processes: A survey. *Reinforcement Learning*, *12*, 253–292.