Attention and Reinforcement Learning: Constructing Representations from Indirect Feedback

Fabián Cañas (canas@colorado.edu) & Matt Jones (mcj@colorado.edu)
University of Colorado, Department of Psychology & Neuroscience
Boulder, CO 80309 USA

Abstract

Reinforcement learning (RL) shows great promise as a theory of learning in complex, dynamic tasks. However, the learning performance of RL models depends strongly on how stimuli are represented, because this determines how knowledge is generalized among stimuli. We propose a mechanism by which RL autonomously constructs representations that suit its needs, using selective attention among stimulus dimensions to bootstrap off of internal value estimates and improve those same estimates, thereby speeding learning. Results of a behavioral experiment support this proposal, by showing people can learn selective attention for actions that do not lead directly to reward, through internally generated feedback. The results are cast in a larger framework for integrating RL with psychological mechanisms of representation learning.

Keywords: Reinforcement Learning; attention; generalization

Introduction

Humans have an incredible capacity to learn new and complex tasks in dynamic environments. In recent years, Reinforcement Learning (RL) has emerged as a theoretical framework that may explain how such powerful learning takes place (e.g., Sutton & Barto, 1998). Reinforcement learning draws on a synthesis of machine learning and neuroscience and offers a set of computational principles for describing learning of dynamic tasks. RL has led to major advances in the ability of machines to learn difficult tasks such as backgammon and autonomous helicopter flight (Tesauro, 1995; Bagnell & Schneider, 2001). RL has also received much interest in neuroscience, based on findings that phasic dopamine signals have similar properties to the internal feedback computed by RL algorithms (Schultz, Dayan, & Montague, 1997). This correspondence suggests that RL offers a useful model of biological learning.

Despite the promise of this framework, the learning performance of RL algorithms strongly depends on the representations on which they operate. RL works by learning which action to perform in each state of a task's environment. In realistically complex tasks with large state spaces, learning about every state individually is impossible, and instead the learner must generalize knowledge among states. Generalization is closely tied to similarity (Shepard, 1987), which in turn depends on how stimuli or situations are represented. Therefore the efficacy of generalization depends on how a task is internally represented. Most often in machine-learning applications, representations are pre-supplied by the modeler based on features that are carefully crafted to capture the most important aspects of the task being learned (e.g., Tesauro, 1995). In psychological contexts, stimuli are chosen so that the subject's representation is transparent, and consequently the question of how the representation is learned is neglected (Schyns, Goldstone, & Thibaut, 1998).

A great deal of psychological research in domains other than RL focuses on how people learn representations to facilitate learning, inference, and decision-making. The aim of our general research program is to explore how such mechanisms might interact with RL, and in particular how RL can build its own representations to bootstrap learning. In the present paper we focus on selective attention, building on models from the literature on category learning (Kruschke, 1992). In a companion paper (Jones & Cañas, 2010), we provide a formal framework for integrating representation learning with RL and implement a specific computational model based on selective attention. Here, we present a behavioral experiment that support the thesis that RL can drive representational learning. Our results show that the internally generated feedback signals at the core of RL can direct shifts of attention toward those stimulus dimensions that are most diagnostic of optimal action.

The remainder of this paper begins with background on RL and modeling of attention learning in categorization. We then outline our proposal for how RL and attention learning can bootstrap off of each other. We then report the results of a sequential decision-making experiment designed to test this specific proposal. Implications are discussed for the role of attention in more complex and temporally extended tasks, prescriptions for training in such tasks, and interactions between representation learning and declarative memory.

Reinforcement Learning

RL is a computational framework for learning dynamic tasks based on feedback from the environment. RL models represent a task as a set of environmental states together with a set of available actions in each state. The action selected at each step determines the immediate reward as well as the ensuing state. This general framework accommodates nearly any psychological task, from simple conditioning to elaborate planning (Sutton & Barto, 1998).

RL works by estimating values of states and actions, which reflect predictions of total future reward. From any given state, the action with the highest estimated value represents a best guess of the choice that will lead to the highest long-term reward. The key to learning value estimates, which lies at the heart of all RL models, is an internally generated feedback signal known as Temporal Difference (TD) error. TD error represents the discrepancy between the estimated value of an action prior to its execution and a new estimate based

on the immediate reward and the value of the ensuing state.

For the mathematically inclined, TD error is defined as

$$\delta = r_t + \gamma \cdot V(s_{t+1}) - Q(a_t, s_t).$$

Here, s_t represents the current state (at time t), a_t is the action selected, and $Q(a_t, s_t)$ is the estimated value of that action. The immediate reward received is denoted r_t, and $V(s_{t+1})$ is the estimated value of the ensuing state. The temporal discount parameter, γ, represents the degree to which the learner values immediate versus delayed rewards.

A critical question for all RL models concerns the relationship between value estimates (Q or V) for different states. The simplest approach is to learn values for all states independently, but for most realistic tasks with large state spaces this approach is unfeasible. Effective learning therefore requires generalization, or the use of knowledge about one stimulus or situation to make inferences or choose actions for other, similar stimuli. A number of methods have been proposed for implementing generalization in RL, and in all cases, the pattern of generalization depends strongly on the way in which states are represented. Representations relying on different features produce different patterns of similarity and hence different generalization. Learning will be most effective if generalization somehow respects the structure of the task, such that the learner pools knowledge about states with similar consequences but discriminates between states that are meaningfully different.

Representation

The various mechanisms for representation learning that have been identified in cognitive psychology all have potential application to RL as means for speeding learning through enhancing generalization. Our work thus far has focused on principles derived from research on category learning. Much of the literature on human category learning aims to understand how humans develop powerful internal representations that facilitate learning and inference of conceptual knowledge. The mechanisms that have been studied include selective attention (Kruschke, 1992; Nosofsky, 1986); feature discovery (Schyns et al., 1998), prototype formation (Smith & Minda, 1998); hybrid rule-exemplar representations (Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994); clustering representations that grow with task complexity (Anderson, 1991); mutable representations that evolve among exemplars, prototypes, and rules (Love & Jones, 2006; Love, Medin, & Gureckis, 2004); and conceptual networks based on causal knowledge (Murphy & Medin, 1985). In this paper, we examine the interaction of RL and selective attention.

Though attention has been studied under many guises in psychology, its implications for learning and generalization have been primarily explored in categorization and animal conditioning. In these literatures, attention has been proposed to act by reshaping generalization gradients (Sutherland & Mackintosh, 1971; Nosofsky, 1986). The generalization gradient is an empirical function that describes how strongly subjects generalize between stimuli as a function of how much those stimuli differ. This function is monotonically decreasing, but it decreases more rapidly along attended dimensions than unattended dimensions (Jones, Maddox, & Love, 2005). Thus subjects generalize less between stimuli when they are attending to the dimensions those stimuli differ on. An alternative view is that the generalization gradient is fixed and isotropic, but the perceptual scaling of individual stimulus dimensions is adjustable. Attention to a dimension serves to stretch the perceptual space so that stimuli differing on that dimension are less similar and thus produce less generalization (Nosofsky, 1986).

Theories of selective attention in category learning propose that people learn to reallocate their attention to improve performance. ALCOVE, a model of categorization with learnable selective attention, has an attention weight for each dimension that determines the degree of generalization along that dimension (Kruschke, 1992). The attention weights are learned by gradient descent on classification error, driven by external feedback. This process leads attention to shift to more predictive dimensions, which leads to less generalization along these dimensions and greater generalization along irrelevant dimensions. Selective attention can thus be thought of as a mechanism for representational learning, which facilitates future learning of the task by adapting generalization.

Incorporating Attention into RL

The previous two sections suggest a natural integration between RL and attention learning. RL's major focus is in updating value estimates by computing sophisticated feedback signals from temporal patterns of reward, but current RL models do not address how value estimates are represented. In contrast, theories of category learning focus on how representations are created that allow for effective generalization, but learning is driven by simple updating rules based on external feedback. We propose a natural unification, in which the feedback signals and updating rules from RL drive the representation-learning mechanisms identified in the categorization literature. This integration makes RL significantly more flexible and autonomous, and therefore possibly more aligned with biological learning.

The critical empirical question we explore operationalizes the idea that RL can adapt its own representation through learned selective attention. Specifically, we investigate whether attention learning can be driven by internally generated TD-error signals in the same way that has been observed with explicit external feedback (Nosofsky, 1986). In a companion paper (Jones & Cañas, 2010), we present a formal model that embodies this hypothesis, by synthesizing the learning mechanisms of ALCOVE (Kruschke, 1992) and Q-learning, a well-studied RL model (Watkins & Dayan, 1992). The formalism of the integrated model shows a tight and mathematically elegant synthesis of the two mechanisms, which we believe offers a strong candidate explanation of

how biological RL processes build their own representations. Here we present an experiment that tests that explanation, by assessing the human capacity for attentional learning via internal value and error signals as opposed to direct external feedback.

Experiment

The goal of the present experiment was to determine whether internal TD-error signals can drive attention learning in the absence of any immediate overt reward. The task consisted of a two-step decision process in which the action on the first step probabilistically determined the stimulus on the second step. Only after the second action did the subject receive feedback about reward.

The second stage of the task was a simple decision task with two possible stimuli and two possible actions. A different action was optimal (i.e., maximized reward) for each of these intermediate stimuli. Once this mapping was learned, one intermediate stimulus led to a higher reward than the other. RL predicts that once subjects learned the optimal actions on this second step, they would learn to assign differential values to the two intermediate stimuli. These values would in turn be used for computing a TD-error signal for actions in the first step, thereby allowing subjects to learn an action policy that maximizes the probability of obtaining the higher-valued intermediate stimulus.

The stimulus for the first choice varied on two continuous dimensions, one of which was more predictive of the outcome of the first action (i.e., the intermediate stimulus) and hence of which choice was optimal. The key question was whether learning the first action through TD error would also lead to learning of selective attention between stimulus dimensions, such that subjects would shift attention to the more relevant dimension. The stimulus set of the first step was designed so as to allow assessment of subjects' attentional allocation based on their patterns of errors, as described below.

Methods

150 undergraduate students from the University of Colorado, Boulder served as the experimental subjects in exchange for course credit.

Subjects were instructed they would pretend to be mushroom farmers. On each trial, they were presented with an image of a mushroom spore and asked to choose between two locations for growing the spore, Sun and Shade. This action determined the intermediate stimulus, a pair of blue or orange mushrooms. They were then given the option to sell the mushrooms to either a Troll or a Goblin, who paid them in gold coins. The structure of the task is outlined in Figure 1.

The stimulus in the first stage was a yellow spore shape, consisting of a circular center measuring 2.3 cm in diameter and radial spines arranged evenly around the center. The spines ranged from 8 mm to 260 mm in length and varied in number between 20 and 100. Spores were uniformly sampled from a circular region inscribed within this two-dimensional stimulus space. The spore was presented in the center of

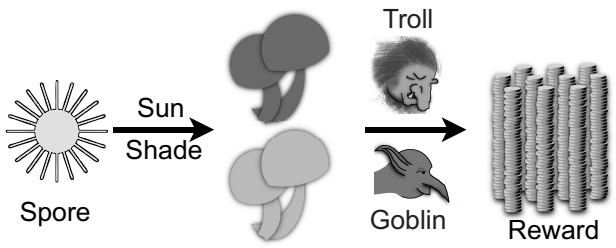

Figure 1: An overview of the task.

an LCD monitor over a black background. The subject selected an action by pressing either S (Sun) or C (Cave) on the keyboard. After this first response was given, the spore disappeared and a pair of cartoon mushrooms appeared in the center of the screen. The subject selected the second action by pressing T (Troll) or G (Goblin). The reward was then presented as stacks of gold coins with a numeric value underneath. The mushrooms and the chosen creature remained on the screen while the reward was displayed.

The transition after each step was animated, lasting 1200 ms between the first response and intermediate stimulus, and 970 ms between the intermediate stimulus and the reward. The reward remained on the screen for 800 ms. A blank screen separated the reward from the beginning of the next trial for 200 ms.

The reward structure for the second step was defined as shown in Table 1. Each mushroom color was associated with a different optimal action. Under these actions, one mushroom (henceforth referred to as the "good" mushroom) afforded a higher reward.

Table 1: Reward Structure of the Second Stage

	Creature Sold to	
Mushroom Color	Goblin	Troll
Blue	[200, 220]	[400 420]
Orange	[300, 320]	[100 120]

Note: Reward on each trial was sampled uniformly from the range shown.

The transition dynamics for the first step were defined as follows. For each action, the probability of one mushroom color versus the other was a logistic function of the dimension values of the spore, given by $p = 1/(1+exp(A(30L+10N)))$, where L and N represent the length and number of the spines, scaled to range from -1 to 1, and A represents the action on the first step, coded here as ± 1. The coefficients for L and N were counterbalanced between subjects, so that L was the more relevant dimension for half the subjects and N was more relevant for the other half. The effect of this design was to create an optimal decision bound, at an angle of 18.4° to one of the two axes, such that the action that maximized the probability of obtaining the good mushroom was determined by which side of the boundary each spore lay on.

Subjects were randomly assigned to Length-relevant and

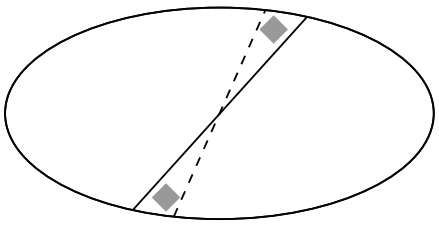

Figure 2: Predictions from selective attention in first step of task. Attention to the more relevant (horizontal) dimension leads to stretching of the stimulus space. Critical stimuli (grey) near ends of optimal decision bound (solid line) are predicted to lead to errors, producing rotation in best fit of linear classifier to subject's responses (dashed line).

Number-relevant conditions, which differed in which spore dimension was more predictive. The roles of the creatures, the colors of the mushrooms, and the labels for the first action were also counterbalanced between subjects. Each subject completed 240 trials (480 total decisions) in blocks of 40.

Predictions and Analysis

Our theory predicts subjects to shift attention to the more relevant spore dimension. Under the view of attention as a transformation of perceptual space, subjects' representations of the set of spores should become stretched along the more relevant dimension and compressed along the less relevant dimension, as shown in Figure 2. Consider the stimuli in the highlighted areas of the figure. Under the attention-altered representation, most of their neighbors lie on the opposite side of the optimal decision bound. Therefore, similarity-based generalization will lead to higher rates of suboptimal actions for these critical stimuli, as compared to matched stimuli on the other side of the optimal bound. The same prediction arises if one assumes subjects learn prototypes for spores associated to the two actions, because each critical stimulus is more similar to the opposite prototype (taken to be the centroid of the region on that side of the optimal bound). Therefore our predictions do not depend on an assumption of exemplar-based generalization.

To test this prediction, we used bivariate logistic regression to fit a linear classifier to each subject's responses. This classifier estimated a linear boundary in stimulus space that best divided the spores the subject chose to grow in the sun from those grown in the shade. To illustrate this analysis, Figure 3 shows the response distribution of a typical subject in the learning group (defined below). Open and closed circles represent stimuli for which the subject selected each of the two actions, the solid line represents the optimal bound, and the dashed line represents the output of the linear classifier. The prediction from selective attention, based on the analysis of expected errors described above, is that the boundary separating each subject's decisions will be rotated relative to the optimal boundary, as shown by the dashed line in Figure 2. Importantly, the estimation of a linear decision bound is a purely descriptive analysis that makes no commitment

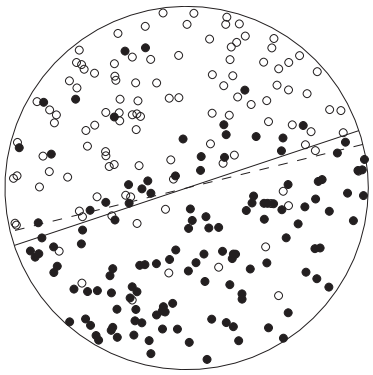

Figure 3: Distribution of responses on first step for a typical subject. Solid line shows optimal bound. Dashed line shows fit of linear classifier.

regarding psychological decision processes. In the companion modeling paper (Jones & Cañas, 2010), we fit a process model based on exemplar generalization, and it makes the same predictions.

In the absence of selective attention, the representation of the stimulus space would remain circular, and therefore by symmetry there should be no systematic bias in the subject's estimated decision bound. Therefore, testing for the predicted bias is a diagnostic way to determine whether our postulated attention-learning mechanism is operating.

Results

On average, subjects made the correct action on the second step of the task on 89.4% of trials. Figure 4 shows the distribution, across subjects, of the proportion of good mushrooms obtained following the first step. The histogram shows a clear bimodality, wherein many subjects performed at chance for the first step, but a significant number were able to learn effective actions.

As explained below, we only predict selective attention for subjects who learn the first stage of the task. Therefore we analyzed the responses of subjects who performed above 70% on the first stage. This cutoff was based on a visual inspection of Figure 4 to safely exclude subjects who were performing at chance. A total of 30 subjects performed at or above 70% on the first step of the task, 11 in the Length-relevant condition and 19 in the Number-relevant condition.

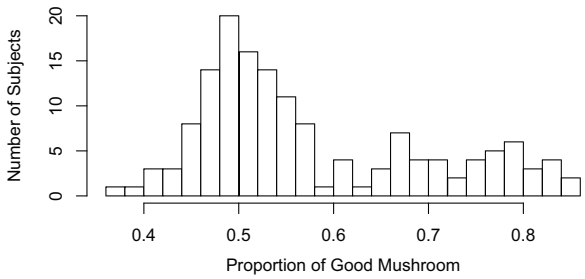

Figure 4: Distribution of performance on first step of task.

A linear classifier was fit to the first-step responses of each subject in the learning group. Figure 5 shows the orientations of the resulting decision bounds, indicated by dots on the circumference of the stimulus region. The mean orientation for each group is shown as a dashed line, and the optimal bound as a solid line. The Number-relevant condition is shown in black and the Length-relevant condition in grey. The mean orientation of the decision bound for subjects in the Length-relevant condition was 7.96° from the Number axis. This value was significantly different from the optimal bound ($18.4°; t_{10} = -2.99, p = .014$) as well as from zero ($t_{10} = 2.29, p = .045$). The mean orientation for the Length-relevant condition was 7.33° from the Number axis. This too was significantly different from the optimal bound ($t_{18} = -3.25, p = .004$) and from zero ($t_{18} = 2.14, p = .046$).

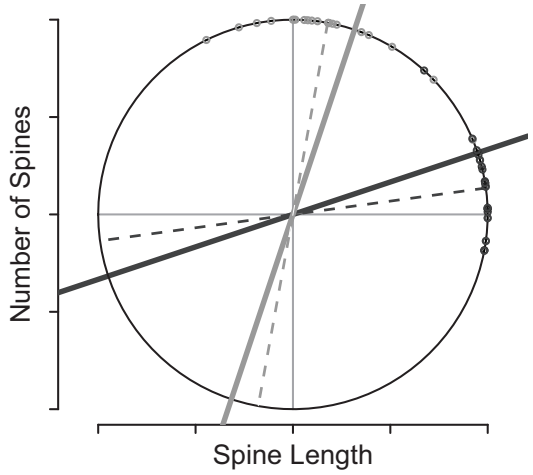

Figure 5: Orientations of empirical decision bounds for subjects in learning group. Small circles = subjects; dashed lines = means; heavy solid lines = optimal bounds; black = Number-relevant; grey = Length-relevant.

Discussion

The results of the decision-bound analysis confirm that subjects made more errors on the critical stimuli. This prediction follows directly from the assumption of selective attention to the more relevant dimension. Because actions on the first step only led to colored mushrooms and not nominal reward, our results support the proposal that attention learning can be driven by internal value estimates and error signals.

We only predicted selective attention for higher-performing subjects for three reasons. First, our theory only predicts attention to be learned once some amount of learning has taken place in associating stimuli to appropriate actions. Attention learning essentially works as a bootstrapping method operating by altering generalization and thus requires some amount of reliable knowledge to begin with in order for adaptation of generalization to have a useful effect. Second, because our theory predicts a bidirectional relationship between attention and value learning, those subjects who exhibit more selective attention should perform better on the task. Therefore, performance acts as a cue to indicate which subjects are more likely to exhibit a measurable effect. The third reason is purely methodological, in that the linear classifier requires a systematic set of responses in order to estimate a meaningful decision bound.

An alternative to our proposal of attention learning is that subjects simply disregarded one dimension of the stimulus entirely. This more strategic explanation is still consistent with our general theory of representation learning driven by RL, but the mechanism would be incompatible with continuous adjustment of attention weights. Regardless, the data rule out this explanation. The fact that the mean bound orientations were reliably different from zero (i.e., the less relevant axis) implies that subjects were sensitive to the less relevant dimension (they were just less sensitive to it than to the primary dimension). Another possibility is that some subjects disregarded one dimension and others disregarded the other, with most subjects in each condition disregarding the less relevant dimension. However, this explanation predicts a bimodal distribution of bound orientations at the subject level, which is clearly not present.

General Discussion

We have shown that humans can learn to shift attention in a dynamic task where reward is not given immediately following the decision that attention acts on. This finding tightly aligns with the internal TD-error signals that RL relies on, and it shows that direct external feedback is not required in order to learn selective attention.

At its core, RL uses predictions or knowledge about later states to build predictions and knowledge about prior states. Application of an RL model to our task predicts that after learning the second stage of the task, one mushroom becomes internally represented as more valuable than the other. This internal value in turn acts as a proxy reward that drives learning in the first stage of the task. Our findings support the proposal that this internal proxy reward signal is also capable of driving attention learning.

An alternative to the interpretation that our subjects are using RL-like internal values for the intermediate stimuli is the possibility of an explicit system that learns about both stages of the task simultaneously after the external reward at the end of each trial. Fu and Anderson (2008) found evidence for such a mechanism in a task structurally similar to ours. Explicit learning based on declarative memory is not, however, incompatible with RL. RL as we have discussed thus far, in its most simple form, only updates estimates about the most recent state. However, specific mechanisms, termed eligibility traces, have been explored within RL to maintain information across time steps to facilitate learning (Sutton & Barto, 1998). Eligibility traces permit simultaneous updating of multiple prior eligible states. Declarative memory may play an important role in encoding these eligibility traces, and therefore Fu and Anderson's results do not preclude an underlying RL

mechanism for learning several steps of a task at once.

Furthermore, declarative memory is unlikely to have played a role in the first step of the present experiment. First, in Fu and Anderson's design (2008), there was a direct correlation between the action in the first step and the eventual reward, which could support direct learning of the first action. In our design, only the conjunction of the spore and the action taken on it was directly related to the possible outcomes after the second step. Second, the spores were drawn from a rich set varying on two continuous dimensions, whereas the second stage of the task was very simple. Therefore subjects likely learned values for the intermediate mushrooms, which could then be used as feedback for the first action, well before the relatively weak correlation between spore-action pairs and final reward could be learned. Third, we have shown that subjects' decision bounds were consistently tilted away from unidimensional rules, indicating that subjects learned the first action using implicit information-integration processes not amenable to declarative memory (Ashby & Maddox, 2005). Though our current work does not completely preclude other learning mechanisms, we sought to isolate mechanisms directly related to RL and TD error, and our results show good support for such mechanisms.

Although not tested directly, the behavior of the subjects who did not learn the first stage sufficiently in our task fits well into the learning framework we propose. Before the differential value of the mushrooms is learned, the feedback to all actions of the first step is constant, which drives attention to generalize across the entire spore space. It is possible that by the time some subjects learned the optimal actions for the second step, they may have learned to entirely disattend any variability of the spores. This inattention is self-perpetuating and prevents future learning.

The potential for learned inattention in dynamic tasks has interesting theoretical and practical implications, because it could make aspects of a task far removed from overt reward difficult to learn. From this perspective, it is clear that an understanding of the mechanisms of attention learning could be beneficial in designing human training programs, such as backward chaining to train intermediate value representations before earlier stages are encountered.

The primary question we examined here was whether TD error, and therefore RL, can have an influence not just on learning values of stimuli within a fixed representation, but whether the representation itself can be altered. Shifts in attention alter the similarity structure of a stimulus space and therefore typify the sort of changes in representation we predict RL to effect. That humans exhibited changes in representation in the service of learning a new task involving fine discrimination of stimuli suggests a rich interplay of representation learning and RL.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychol Rev*, *98*, 409–429.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu Rev Psychol*, *56*, 149-178.

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

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *J Exp Psychol Gen*, *127*, 107-140.

Fu, W., & Anderson, J. R. (2008). Dual learning processes in interactive skill acquisition. *J Exp Psychol-Appl*, *14*, 179-191.

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

Jones, M., Maddox, W. T., & Love, B. C. (2005). Stimulus generalization in category learning. *Proceedings of the* 27^{th} *Annual Conference of the Cognitive Science Society*, 1066-1071.

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psych Rev*, *99*, 22-44.

Love, B. C., & Jones, M. (2006). The emergence of multiple learning systems. *Proceedings of the* 28^{th} *Annual Conference of the Cognitive Science Society*, 507-512.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain: A network model of category learning. *Psychol Rev*, *111*, 309–332.

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

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

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychol Rev*, *101*, 53–79.

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

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behav Brain Sci*, *21*, 1-54.

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

Smith, J., & Minda, J. (1998). Prototypes in the mist: The early epochs of category learning. *J Exp Psychol Learn*, *24*(6), 1411–1436.

Sutherland, N., & Mackintosh, N. (1971). *Mechanisms of Animal Discrimination Learning*. NY: Academic Press.

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

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

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*, 279–292.