Learning to Generalize through Predictive Representations: A Computational Model of Mediated Conditioning

Elliot A. Ludvig and Anna Koop

Department of Computing Science
University of Alberta
Edmonton, AB T6G 2E8
{elliot,anna}@cs.ualberta.ca

Abstract. Learning when and how to generalize knowledge from past experience to novel circumstances is a challenging problem many agents face. In animals, this generalization can be caused by mediated conditioning—when two stimuli gain a relationship through the mediation of a third stimulus. For example, in sensory preconditioning, if a light is always followed by a tone, and that tone is later paired with a shock, the light will come to elicit a fear reaction, even though the light was never directly paired with shock. In this paper, we present a computational model of mediated conditioning based on reinforcement learning with predictive representations. In the model, animals learn to predict future observations through the temporal-difference algorithm. These predictions are generated using both current observations and other predictions. The model was successfully applied to a range of animal learning phenomena, including sensory preconditioning, acquired equivalence, and mediated aversion. We suggest that animals and humans are fruitfully understood as representing their world as a set of chained predictions and propose that generalization in artificial agents may benefit from a similar approach.

The texture of our experience is often dotted by aversions and affinities that are only indirectly related to rewarding or punishing outcomes. For example, if I have a near-death experience in an ambulance on the way to the hospital, I am likely to shudder next time I hear a siren go by, even if the ambulance was not playing its siren during my traumatic episode. Or if I get really sick at a restaurant before ordering, I will certainly think twice about eating their food in the near future. In these two examples, stimuli that were never directly experienced in the offending situations still gain some of the residual response that memory of the initial situation provokes. In the animal learning literature, this indirect learning has been termed *mediated conditioning* and repeatedly reproduced in the laboratory with notable instances including sensory preconditioning [1,2], as in the ambulance example, acquired equivalence [3,4], and mediated aversion [5,6], as in the restaurant example. Generalization between stimuli based on their experienced history seems *prima facie* like a valuable asset to an animal or human (or animat) facing novel stimuli or situations and may even form part of the basis for categorization and conceptual knowledge [7].

M. Asada et al. (Eds.): SAB 2008, LNAI 5040, pp. 342–351, 2008.

This learned generalization hints at a larger puzzle that has troubled researchers in both machine learning and cognitive science: When should what you learn in one situation generalize to what you do in another? Generalization is a cornerstone of adaptive behavior that allows agents to take advantage of previous experience beyond the particulars of the original learning context. In the psychological literature, most studies of generalization have focussed on how responding generalizes amongst physically similar stimuli (e.g., tones of different frequencies). Mediated conditioning, however, presents an instance whereby long-term equivalences can be established between physically distinct stimuli, merely because of the animal's experience with the consequences, antecedents, and associates of those stimuli [4]. In this paper, we propose that the computational formalism of *predictive representations* [8,9,10] from reinforcement learning provides an efficient and effective mechanism for the mediated conditioning exhibited by many animals and humans. In this predictive representation (PR) approach to adaptive learning, stimuli are represented as the constellation of predicted future observations, rather than as composites of their physical properties. We leverage this idea to develop a real-time PR model and show how this reinfrocement-learning model explains the learned generalization observed in mediated conditioning experiments.

1 Predictive Representation Model

The key insight behind our model is that stimuli are represented as a collection of chained predictions about future observations [9]. This predictive representation for a stimulus implies that generalization will occur readily between stimuli that share similar predictions about the future—in a strong parallel to the manner that generalization occurs most readily between stimuli that share physical properties. These PRs play a similar role to the images or associatively activated representations in other theories of animal conditioning [4,6,11,12].

Figure 1 presents a schematic of the PR model, illustrating how these representations fit into the full learning scheme. Prediction generation in the model is a two-step process: On a given time step, the observations (stimuli) are first used to generate interim predictions for every potential stimulus. These interim predictions are then combined with the same initial observations to generate a new set of final predictions for that time step. These final predictions determine behavior, so, for example, in a simple learning task where a light is followed by a tone and then by food, the light would lead to a prediction of the tone which would lead to a prediction of the food. As a result, after learning, the light would also (indirectly) lead to a (weaker) prediction of the food and thereby elicit some of the associated conditioned responding.

We approach these tasks as a reinforcement-learning prediction problem, except that we calculate a separate value function for every stimulus—not only rewards. More formally, on every time step t, a value function V_t is computed for every potential observation as a semi-linear function of the vector $\mathbf{x}_t$ of the observation/prediction values x_t^i and the vector $\mathbf{w}_t$ of the learned weights w_t^i:

$$V_t = \sigma(\mathbf{w}_t^T \mathbf{x}_t) = \sigma(\sum_{i=1}^{n} w_t^i x_t^i) \tag{1}$$

where the squashing function, $\sigma(x) = \frac{1}{1+e^{-x}}$, is used to keep the value of V between 0 and 1. Half the components of the vector $\mathbf{x}_t$ are binary, indicating whether a particular stimulus was present on that time step (1) or not (0). Such a simplification is not strictly necessary, and real-valued noisy observations are surely possible, but not considered here for ease of exposition. The other half of the components are real-valued elements that correspond to predictions (see Fig. 1). The key to our model is that this computation is performed twice on each time step. The first iteration uses only the binary observations (with all predictions set to 0) to calculate an interim prediction. On the second iteration, this interim prediction becomes part of the stimulus representation and is used to generate the final prediction that is compared to future experience.

Learning in the PR model occurs on the ensuing time step when new observations are encountered, through the temporal-difference (TD) learning algorithm [10,13]. With this learning rule, an error δ_t is formed for each potential outcome, which is the difference between the current prediction and the sum of the new observations and resultant new predictions (as discounted by γ):

$$\delta_t = x_{t+1} + \gamma \tilde{V}_{t+1} - V_t \ . \tag{2}$$

Note that $\tilde{V}_{t+1}$ is the prediction as calculated using the vector of the new observations and predictions, $\mathbf{x}_{t+1}$, and the weight vector before being updated, $\mathbf{w}_t$, through the same two-step process described above. The discount factor, γ, determines the temporal horizon of the prediction. A low γ makes the model short-sighted, focusing the

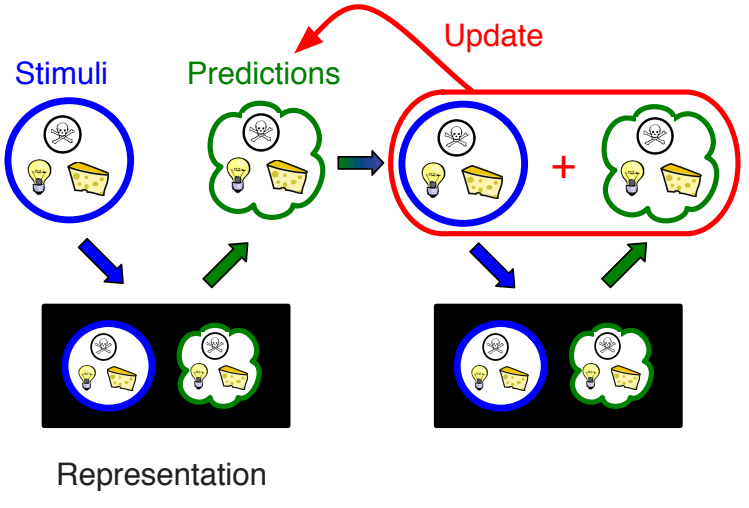

Fig. 1. Schematic of the PR Model. Observations (*blue circles*) at one time step are first used to generate interim predictions (*boxed green clouds*) for every stimulus. The same observations are then combined with these interim predictions to generate the final predictions (*larger green clouds*). Finally, on the next time step, the new observations and resultant new predictions are combined to update (*red arrow*) the weights based on the discrepancy between the predictions from the previous time step and these observed outcomes.

prediction on the near future; a higher γ extends the temporal window which the model is trying to predict.

The prediction calculation is updated on every time step by changing each weight, w^i, according to the TD error for that potential outcome:

$$w^i_{t+1} = w^i_t + \alpha\delta_t x^i_t \tag{3}$$

where α is a step-size parameter that influences the learning rate. In the model, events that have high salience, such as shocks, have a large step size, whereas less salient events, such as tones and lights, have a lower step size (and thus learning rate). For simplicity, we chose to only update the weights from the second iteration and force the weights in the first iteration to be identical to the corresponding weights in the second iteration. Other versions of this PR model with deeper (multi-layer or recurrent) predictions or with multiple cascading or independent learning updates are certainly possible and may even capture further empirical phenomena not considered here.

An important element of the PR model is that all experimental situations are modeled as real-time. In previous models of mediated conditioning and related phenomena [12,15], the flow of experience was often divided into discrete trials and punctate events—a structure which is not immediately apparent in the real world. The PR model allows stimuli to exist for multiple time steps, thereby predicting the continuation of themselves, a feature that is vital in explaining sensory preconditioning and mediated aversion (see Tables 2 and 3).

2 Results

We demonstrate successful performance of the PR model on three animal learning tasks that seem to involve mediated conditioning: acquired equivalence [3], sensory preconditioning [2], and mediated aversion [5]. For each of these experiments, we simulated the PR model with 100 trials in the first stage, 3 trials in the second stage, and a single test trial in the final stage. Sensory stimuli all lasted for 15 time steps, while food reward, shock, and illness lasted 3 time steps; an inter-trial interval of 60 time steps separated trials. In all simulations, the discount factor γ was .98, and the step size α was .4 for shock and illness, .3 for food reward, and .05 for other stimuli. Weights were initialized to 0 and capped at 3.

2.1 Application: Acquired Equivalence

When two stimuli are repeatedly followed by the same outcome, they often come to be treated more similarly in the future; that is, these stimuli acquire an equivalence relation [4,16]. For example, Honey and Hall [3] presented rats with three different stimuli (A, B, and C): A and B were always followed by food reward (f) while C was never rewarded (see upper part of Table 1). Rats then received parings of stimulus A with an electric shock (sh). When subsequently tested with stimuli B and C, rats showed significantly greater conditioned fear to stimulus B, which shared a common history with

Table 1. Experimental details and model interpretation of the acquired equivalence experiment from Honey and Hall [3]. Each column is a different stage of the experiment. A, B, and C are different sounds; f = food; sh = shock; pr = prediction.

Stage 1	Stage 2	Test	Result
Experiment: Honey and Hall [3]			
A→f	A→sh	B	B > C
B→f		C	
C			
PR Model Explanation:			
A→pr(f)	A→pr(f)→sh	B→pr(f)	B→pr(sh)
B→pr(f)	Therefore:	& pr(f)→pr(sh)	
	A→pr(sh)		
	pr(f)→pr(sh)		

shocked stimulus A. This transfer of conditioned fear to the stimulus that shared a training history with the shocked sound is the hallmark of an acquired equivalence relationship. This acquired equivalence by common consequences has also been demonstrated in pigeons [14] and humans [7,17].

Figure 2 presents simulation results from the PR model on this acquired equivalence task. As with animals, the model produced a greater prediction of shock (equivalent to more conditioned fear) with the stimulus (B) that shared a training history with the shocked stimulus (A). The lower part of Table 1 gives an intuitive account of how our model yields these results. After the first stage, both A and B produce a prediction of food. In the second stage of training, A produces a prediction of food, which is followed by shock. Thus, A produces a prediction of shock, and, here is the key point, the prediction of food also produces a prediction of shock. Finally, in the third stage, B still produces a prediction of food, which, in turn, produces a prediction of shock. This indirect prediction of shock is the basis of the acquired equivalence effect (and other forms of mediated conditioning) in our PR model.

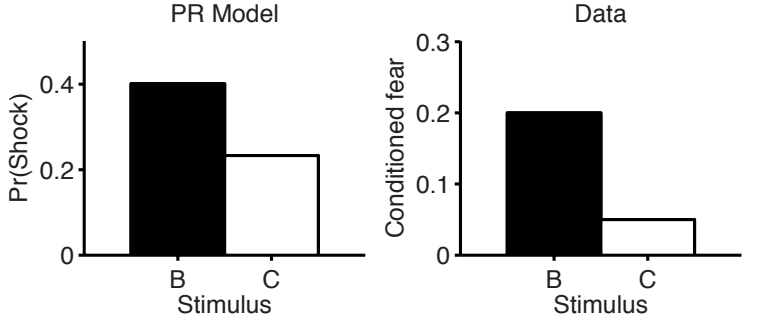

Fig. 2. PR model simulation results (*left*) and empirical data (*right*) from an acquired equivalence experiment. Data are re-plotted from Honey and Hall [3]. Stimulus B shared a training history with the shocked stimulus, while stimulus C did not.

2.2 Application: Sensory Preconditioning

Sensory preconditioning is another experimental situation wherein one stimulus gains an association with reward through the mediation of a second stimulus [1,2]. The upper portion of Table 2 displays the experimental design for a typical sensory preconditioning procedure. Animals are first trained with pairings of two previously neutral stimuli (A, B) as well as a third, unpaired stimulus (C). In the second stage, one of the paired stimuli (B) is then followed by a mild shock (sh). Finally, in the test stage, the other two stimuli are presented alone, and animals display greater conditioned fear to the paired stimulus (A) than the unpaired stimulus (C). The link established between B and A by their training history (in Stage 1) results in greater subsequent generalization between the two stimuli. This effect can be further augmented by presenting the stimuli simultaneously rather than sequentially in Stage 1 (bracketed conditions in Table 2).

Figure 3 displays the empirical data (right) and corresponding simulation results (left) from the real-time PR model in a sensory preconditioning procedure. As with real animals, in the model, sensory preconditioning results in greater generalization to the paired stimulus (A) from the first stage, most markedly for the simultaneous training case [2]. The lower portion of Table 2 schematizes how the PR model explains this generalized responding to stimulus A in the test stage. After the first stage of training, stimulus A produces a prediction of stimulus B. Because stimuli last for multiple time steps in the real-time PR model, all stimuli also learn to produce self-predictions. In the second stage, stimulus B produces a prediction of itself, so both the stimulus and its prediction are followed by the shock. As a result, both B and the prediction of B lead to predictions of shock. In the final, test stage, stimulus A leads to a prediction of B, which leads to a prediction of shock and the associated conditioned response. The simultaneous case shows greater sensory preconditioning than the sequential version in the PR model because, in the first stage, in the simultaneous case, the model additionally learns that stimulus B predicts stimulus A (bracketed value in Table 2). This additional

Table 2. Experimental details and model interpretation for a sensory preconditioning experiment. Each column is a different stage of the experiment. For clarity, only predictions directly pertinent to the explanation of the primary effect are included. Bracketed items refer to the simultaneous version of the task. A, B, and C are different stimuli; sh = shock; pr = prediction.

Stage 1	Stage 2	Test	Result
Experiment: Rescorla [2]			
A→B [AB]	B→sh	A	[A] > A > C
C		C	
PR Model Explanation:			
A→pr(A),pr(B)	B→pr(B)→sh	A→pr(A),pr(B)	A→pr(sh)
B→**pr(B)**,[pr(A)]	[B→pr(A)→sh]	& pr(B)→pr(sh)	[A→pr(sh)]
	Therefore:	[& pr(A)→pr(sh)]	
	B→pr(sh)		
	pr(B)→pr(sh)		
	[**pr(A)→pr(sh)**]		

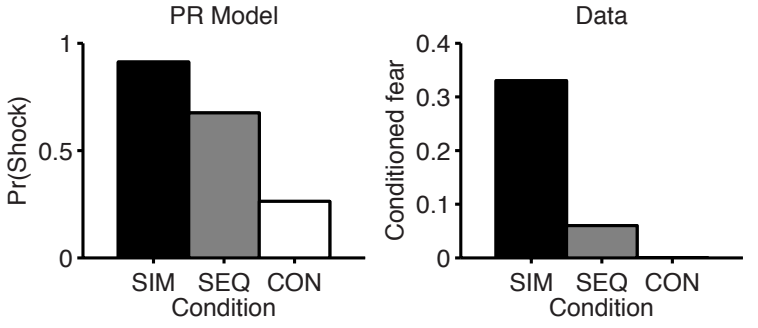

Fig. 3. PR model simulation results (*left*) and empirical data (*right*) from a sensory preconditioning experiment. Data are re-plotted from Figure 4 in Rescorla [2] as degree of response suppression. SIM = simultaneous; SEQ = sequential; CON = control.

prediction results in the prediction of A directly preceding shock in Stage 2, leading to greater generalization in Stage 3 because of the self-prediction of stimulus A.

2.3 Application: Mediated Aversion

A final set of empirical phenomena that nicely illuminate properties of this PR model are the series of mediated aversion experiments [5,6]. Table 3 shows the design for a typical experiment: Animals are first trained with 2 neutral stimuli (A, B) each paired with one of 2 different foods/flavours (f1, f2). In the second stage, animals are presented one of the two stimuli followed by injection with lithium chloride (LiCl), an illness-inducing agent. On the final, test stage, animals are presented with the 2 foods/flavours and will typically preferentially eat from the food whose associate was not paired with illness in the second stage.

Table 3. Experimental details and PR model interpretation for a mediated aversion experiment. Each column is a different stage of the experiment. For clarity, only predictions directly pertinent to the explanation of the primary effect are included. A, B = stimuli; f1, f2 = foods/flavours; G1, G2 = 2 groups of animals; pr = prediction; LiCl = Lithium Chloride, an illness-inducing agent.

Stage 1	Stage 2	Test	Result
Experiment: Holland [5]			
A→f1 B→f2	G1: A→LiCl & G2: B→LiCl	f1	G1>G2
PR Model Explanation:			
A→pr(A),pr(f1) f1→pr(f1)	A→pr(f1)→LiCl Therefore: A→pr(LiCl) **pr(f1)→pr(LiCl)**	**f1→pr(f1)** & pr(f1)→pr(LiCl)	$f1\rightarrow_{pr(LiCl)}$

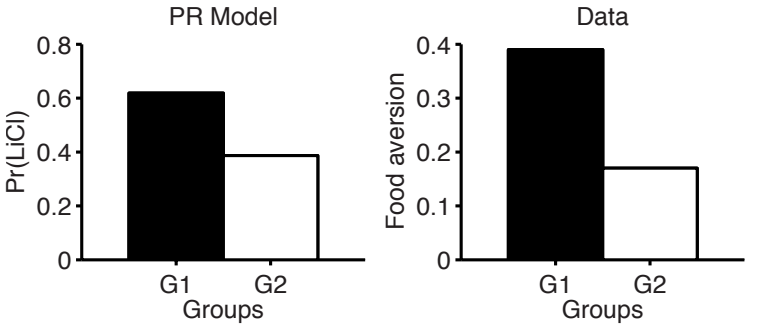

Fig. 4. PR model simulation results (*left*) and empirical data (*right*) from a mediated aversion experiment. Data are re-plotted from Figure 1 in Holland [5] as percent decrease in consumption of food f1 for the groups that had the same food (G1) or a different food indirectly devalued (G2).

Figure 4 shows how the PR model effectively captures the key result: In both the data and simulations, the group that had food f1 indirectly devalued showed a much greater prediction of illness than the second group. As sketched out in the lower part of Table 3, the PR model learns that stimulus A leads to a prediction of food f1 in the first stage. As a result, in the second stage, both stimulus A and the prediction of food f1 precede (and learn to produce predictions of) illness (LiCl administration). Finally, in the final test phase, food f1 leads to a prediction of itself which leads to a prediction of illness and the observed food aversion. Once again, self-prediction is an important component of the explanation, but this time in a different guise than with sensory preconditioning. In sensory preconditioning, self-prediction is the important feature in the second phase when stimulus B's self-prediction leads to the prediction of B producing a prediction of food (cf. Table 2). In the mediated aversion experiment, the crucial self-prediction occurs in the final phase when food predicts itself, leading to a prediction of illness (cf. Table 3).

3 Conclusions

In this paper, we have shown how mediated conditioning can be effectively modeled with our real-time PR network model. The PR model conceives of humans and animals as generating a network of chained predictions of future observations, which, in some ways, cashes out the "image" or "representation" of earlier theories of conditioning [4,6,11]. The selected empirical examples—acquired equivalence, sensory preconditioning, and mediated aversion—each illustrate additional properties of the model in explaining this form of learned generalization.

In their neural network model, Gluck and Myers [12] also address many of the same empirical phenomena. They suggest that redundancy compression and predictive differentiation are the two functions largely responsible for the increased generalization observed in mediated conditioning experiments. Here, we propose an alternate computational account, based on the notion that stimuli are represented as the chained predictions of all future observations. Similarities in this predictive space produce learned

generalization between stimuli. In addition, the real-time dynamics of our PR model proffers novel explanations for more phenomena, including the difference between simultaneous and successive sensory preconditioning (see Fig. 3). Our model also bears some similarity to Sutton's TD Models [10], which allow artificial agents to incrementally learn a full-world model for better planning.

Where in the brain might all these iterative predictions be computed? One possibility is suggested from the few studies that have examined lesion effects on these tasks. We know that acquired equivalence and sensory preconditioning are both dependent on the hippocampus and the surrounding entorhinal and perirhinal cortices [12,18,19,20,21]. Moreover, humans with hippocampal atrophy show deficits in the transfer (generalization) stage of an acquired equivalence task [17]. These results, taken together, hint that the medial temporal areas might be responsible for creating new predictive representations for use by reinforcement learning systems elsewhere in the brain (e.g., basal ganglia; see [22]). These predictive representations could also provide a unifying framework for knowledge creation [23], including spatial learning and object memory, two of the more common processes attributed to the hippocampus and perirhinal cortex, respectively.

In conditioning, animals clearly learn more than a simple association between a neutral cue and a rewarding stimulus. They learn a panoply of interrelations among all the different stimuli in their environment—relationships that can be exposed through clever experimental manipulations, as in the generalization tests central to acquired equivalence, sensory preconditioning, and mediated aversion. These three examples of mediated conditioning or learned generalization demonstrate the value of trying to model animal learning as a network of chained predictions. This predictive promiscuity, as captured by our PR model, helps animals learn and adapt more quickly when confronted with novel situations and stimuli. No doubt the empirical story in each of these cases is more nuanced than this brief exposition has allowed (for more details, see [4,6]), but our model captures the core effects and offers a framework for thinking about how the empirical exceptions might constrain future modeling attempts.

Acknowledgments. The authors would like to thank Rich Sutton for inspiration without constraint, the Alberta Ingenuity Fund and iCore for support, and Karen Skinazi for editing help.

References

1. Brogden, W.J.: Sensory pre-conditioning. Journal of Experimental Psychology 25, 323–332 (1939)
2. Rescorla, R.A.: Simultaneous and successive associations in sensory preconditioning. Journal of Experimental Psychology: Animal Behavior Processes 6, 207–216 (1980)
3. Honey, R.C., Hall, G.: Acquired equivalence and distinctiveness of cues. Journal of Experimental Psychology: Animal Behavior Processes 15, 338–346 (1989)
4. Hall, G.: Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. Animal Learning & Behavior 24, 233–255 (1996)
5. Holland, P.C.: Acquisition of representation-mediated conditioned food aversions. Learning & Motivation 12, 1–18 (1981)

6. Holland, P.C.: Event representation in pavlovian conditioning: Image and action. Cognition 37, 105–131 (1990)
7. Hall, G., Mitchell, C., Graham, S., Lavis, Y.: Acquired equivalence and distinctiveness in human discrimination learning: Evidence for associative mediation. Journal of Experimental Psychology: General 132, 266–276 (2003)
8. Littman, M.L., Sutton, R.S., Singh, S.: Predictive representations of state. In: Advances in Neural Information Processing Systems, vol. 14, pp. 1555–1561 (2002)
9. Rafols, E.J.R., Ring, M.B., Sutton, R.S., Tanner, B.: Using predictive representations to improve generalization in reinforcement learning. In: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 835–840 (2005)
10. Sutton, R.S.: TD models: Modeling the world at a mixture of time scales. In: Proceedings of the 12th International Conference on Machine Learning, pp. 531–539 (1995)
11. Wagner, A.R.: SOP: A model of automatic memory processing in animal behavior. In: Spear, N.R., Miller, R.R. (eds.) Information processing in animals: Memory mechanisms, pp. 5–47. Erlbaum, Hillsdale (1981)
12. Gluck, M., Myers, C.: Hippocampal mediation of stimulus representation: A computational theory. Hippocampus 3, 491–516 (1993)
13. Sutton, R.S.: Learning to predict by the methods of temporal differences. Machine Learning 3, 9–44 (1988)
14. Bonardi, C., Rey, V., Richmond, M., Hall, G.: Acquired equivalence of cues in pigeon autoshaping: Effects of training with common consequences and with common antecedents. Animal Leaning & Behavior 21, 369–376 (1993)
15. Honey, R.C., Ward-Robinson, J.: Acquired equivalence and distinctiveness of cues: I. Exploring a neural network approach. Journal of Experimental Psychology: Animal Behavior Processes 28, 378–387 (2002)
16. Lawrence, D.H.: Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. Journal of Experimental Psychology 39, 770–784 (1949)
17. Myers, C.E., Shohamy, D., Gluck, M.A., Grossman, S., Kluger, A., Ferris, S., Golomb, J., Schnirman, G., Schwartz, R.: Dissociating hippocampal versus basal ganglia contributions to learning and transfer. Journal of Cognitive Neuroscience 15, 185–193 (2003)
18. Coutureau, E., Killcross, A.S., Good, M., Marshall, V.J., Ward-Robinson, J., Honey, R.C.: Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. Journal of Experimental Psychology: Animal Behavior Processes 28, 388–396 (2002)
19. Lazzaro, S.C., Gournani, K., Ludvig, E.A., Gluck, M.A.: Lesions of the entorhinal cortex abolish sensory preconditioning in rats. In: Society for Neuroscience Abstracts, 997.12 (2005)
20. Nicholson, D., Freeman, J.: Lesions of the perirhinal cortex impair sensory preconditioning in rats. Behavioral Brain Research 112, 69–75 (2000)
21. Port, R.L., Patterson, M.M.: Fimbrial lesions and sensory preconditioning. Behavioral Neuroscience 98, 584–589 (1984)
22. Schultz, W., Dayan, P., Montague, P.R.: A neural substrate of prediction and reward. Science 275, 1593–1599 (1997)
23. Koop, A.: Understanding experience: Temporal coherence and empirical knowledge representation. Master's thesis, University of Alberta (2007)