

A Cognitive Model That Describes the Influence of Prior Knowledge on Concept Learning

Toshihiko Matsuka and Yasuaki Sakamoto

Stevens Institute of Technology, Hoboken, NJ 07030, USA
{tmatsuka, ysakamot}@stevens.edu

Abstract. It is well known that our prior knowledge and experiences affect how we learn new concepts. Although several formal modeling attempts have been made to quantitatively describe the mechanisms about how prior knowledge influences concept learning behaviors, the underlying cognitive mechanisms that give rise to the prior knowledge effects remains unclear. In this paper, we introduce a computational cognitive modeling framework that is intended to describe how prior knowledge and experiences influence learning behaviors. In particular, we assume that it is not simply the prior knowledge stored in our memory trace influencing our behaviors, but it is also the learning strategies acquired through previous learning experiences that affect our learning behaviors. Two simulation studies were conducted and the results showed promising outcomes.

1 Introduction

A body of empirical studies in human concept acquisition indicates that our learning behaviors are strongly influenced by our own prior knowledge and experiences (e.g. [1]). Several formal modeling attempts have been made to quantitatively evaluate hypotheses about how our prior knowledge influences our cognitive behaviors associated with concept learning (e.g. [2]). Yet, most of them take the form of a retrospective model only allowing post-hoc data fitting, a model that does not offer a priori predictions about the observed effects of prior knowledge (but see [3]). Furthermore, these models of prior knowledge tend to emphasize the forward processes (i.e., how stored information is used) without specifying any learning algorithms (i.e., how information is updated) and thus cannot describe the underlying cognitive mechanisms that yield prior knowledge effects.

In the paper, we introduce a computational cognitive model with a learning algorithm that is intended to describe how prior knowledge and experiences influence learning behaviors. One unique contribution of our work is that we apply a hybrid meta-heuristic optimization method to describe how prior knowledge affects biases in acquired knowledge as well as biases in learning strategies (e.g. biases in hypothesis generation). Our new cognitive model, PIKCLE (PrIor Knowledge on Concept LEarning), integrates the effect of prior knowledge on concept learning. A novel characteristic of PICKLE is its assumption that it is not simply the prior knowledge stored in our memory trace influencing our learning behaviors, but it is also the learning strategies acquired through previous learning experiences that affects our learning behaviors. Another significant aspect of PICKLE is that it may generate insightful a priori predictions about the role

of prior learning experiences on subsequent learning, as we demonstrate in the current simulation studies.

Note that the terms “category learning” and “concept learning” will be used interchangeably throughout this paper, because conceptual knowledge is considered to be categorically organized.

2 Categorization Algorithm

Rather than introducing new forward algorithms we chose to apply PIKCLE’s learning processes to SUPERSET’s [4] [5] categorization processes, because our main objective is to introduce a learning model that explains the effects of prior knowledge on concept acquisition. SUPERSET is a computational model of category learning based on the assumption that humans have a very flexible knowledge representation system, capable of adapting a situationally “optimal” representation scheme (e.g. rules, prototypes, or exemplars).

Unlike previous modeling approaches to category learning research which modify a single notion (i.e., a single complete set of coefficients, $\theta^n: \{\mathbf{w}^n, \mathbf{D}^n, \mathbf{C}^n\} \in \theta^n$), SUPERSET maintains, modifies, and combines a set of notions. The idea of having a population of notions (as opposed to having a single notion) is important because it allows not only the selection and concept combination in learning, but also the creation of diverse notions, making learning more robust. Thus, unlike previous categorization models, SUPERSET assumes that humans have the potential to maintain a range of notions and are able to apply a notion that is most suitable for a particular set of situational characteristics.

The SUPERSET framework, like several other models of the human categorization process, assumes that humans hold memorized internal reference points (i.e., rules, prototypes, or exemplars) and utilize similarities or conformities between the input stimulus and the reference points as evidence to probabilistically assign the input stimulus to an appropriate category. The similarity (i.e., s_j) between input x and j -th reference point (i.e., R_j) in SUPERSET are formulated as:

$$s_j^n(x) = \exp \left[-\beta \left[\sum_{i=1}^I \frac{(R_{ji}^n - x_i)^2}{1 + \exp(-D_{ji}^n)} + \sum_{i=1}^{I-1} \sum_{m=i+1}^I 2C_{jim}^n [R_{ji}^n - x_i][R_{jm}^n - x_m] \right] \right] \quad (1)$$

where D_j and C_j are R_j ’s dimensional and correlational selective attention, respectively. A free parameter β defines an overall similarity gradient. Superscript n is an index for different notions. Subscripts i and m indicate feature dimensions, and I is the number of feature dimensions. Note that it is assumed that $C_{jim}^n = C_{jmi}^n$, $C_{jim}^2 \leq (1 + \exp(-D_{ji}^n))^{-1} \cdot (1 + \exp(-D_{jm}^n))^{-1}$. The correlational attention weights can take a negative value, where its signum indicates direction of attention field while its magnitude indicates the strength of attention.

The psychological similarities are fed forward to the output category nodes using the following function:

$$O_k^n(x) = \sum_j w_{kj}^n s_j^n(x) \quad (2)$$

where w_{kj} is a learnable association weight between reference point j and category node k . The probability that x being classified as category A is calculated using the following choice rule:

$$P(A|x) = \frac{\exp(\phi \cdot O_A^o(x))}{\sum_k \exp(\phi \cdot O_k^o(x))} \quad (3)$$

where ϕ controls decisiveness of classification response, and superscript o indicates the notion adopted to make a categorization response (i.e., the most useful notion).

3 PIKCLE

3.1 Qualitative Descriptions of Learning Mechanisms

PIKCLE assumes that human learning involves the consideration of multiple notions according to their usefulness in a given task. Each of these notions determines which aspects of the categories are psychologically salient and which can be ignored. PICKLE assumes that the definition of the learning objective is not based solely on the accuracy of knowledge, but also on the subjectively and contextually determined utility of knowledge being acquired.

PIKCLE's learning algorithm is built on the basis of the Evolution Strategy (ES) optimization method. PIKCLE, as in a typical ES application, assumes three key processes in learning: *crossover*, *mutation*, and (survivor) *selection*. In the *crossover* process, the randomly selected notions (i.e., one complete set of SUPERSET coefficients indicated by superscript n) form a pair and exchange elements (i.e. coefficients) of notions, creating a new notion. In human cognition, the crossover process can be interpreted as conceptual combination, where new notions are created based on merging ideas from existing effective notions. In the *mutation* process, each element of notion (i.e., coefficient) is pseudo randomly altered (see section below for detail). A mutation can be considered as a modification of a notion by randomly generating a new hypothesis. Note that each element of a notion (i.e., SUPERSET's coefficient) is associated with a unique dynamically altering random hypothesis generator function, allowing it to be sensitive to the topology of the hypersurface of objective function (e.g., searching within a smaller area if coefficients are close to an optimum). In the *selection* process, a certain number of notions are deterministically selected on the basis of their fitness in relation to the environment for survival. Those selected notions (i.e., a set of coefficient) will be kept in PIKCLE's memory trace, while non-selected notions become obsolete or are forgotten.

However, unlike typical ES application, PIKCLE also assume a systematic bias caused by prior knowledge and experiences in concept modification. That is, while there is still a stochastic process involved in the knowledge *mutation* process, there is a force or momentum created by prior experiences that systematically influences the likelihood of particular patterns of notions (coefficient configurations) to emerge in the process.

3.2 PIKCLE Learning Algorithm

For the sake of simplicity, we use the following notation $\{\mathbf{w}^n, \mathbf{D}^n, \mathbf{C}^n\} \in \theta^n$.

Knowledge Combinations. In PIKCLE, randomly selected pairs of notions exchange information to combine notions. In particular, PIKCLE utilizes the discrete recombination of SUPESSET coefficients. Thus, parent notions θ^{p1} and θ^{p2} produce a child notion θ^c or

$$\theta_i^c = \begin{cases} \theta_i^{p1} & \text{if } UNI \leq 0.5 \\ \theta_i^{p2} & \text{otherwise} \end{cases} \quad (4)$$

where UNI is a random number drawn from the Uniform distribution. This combination process continues until the number of children notions produced reaches the memory capacity of PIKCLE.

In PICKLE, self-adapting strategy coefficients denoted by σ_i^n , which define the widths of random search areas for SUPERSET's coefficients, are also combined during this stage. Unlike θ^n , the intermediate crossover method is used for the self-adapting strategy coefficients. Thus, $\sigma_i^c = 0.5 \cdot (\sigma_i^{p1} + \sigma_i^{p2})$.

Knowledge Modifications. Two separate hypothesis generation mechanisms are involved in PIKCLE's knowledge modification phase. One is a traditional stochastic modification process, modifying each coefficient with a random number drawn from the Gaussian distribution (i.e., last term in Eq.5). The other is systematic hypothesis generation (i.e. middle term in Eq. 5) that takes relationships among different types of knowledge in its memory trace into account (e.g. interaction between attention weights allocated to feature dimensions 1 and 2, i.e., D_{j1} & D_{j2}). Specifically, one element of notion θ_i^n at time $t + 1$ is updated as follows:

$$\theta_i^n(t + 1) = (1 - \eta)\theta_i^n(t) + \eta A(\mathbf{v}, \mathbf{u}, \theta^n(t)) + N(0, \sigma_i^n(t + 1)) \quad (5)$$

where t indicates time, A is a simple autoassociative network with one hidden layer causing a systematic bias in knowledge modification, a free parameter η controls the relative influence of the systematic bias, and $N(0, \sigma_i)$ is a random number drawn from the Gaussian distribution with the corresponding parameters. The random search width for notion element l is given as:

$$\sigma_i^n(t + 1) = \sigma_i^n(t) \cdot \exp(N_G(0, \gamma_G) + N_l(0, \gamma_l)) \quad (6)$$

where $N_G(0, \gamma_G)$ is a random number drawn from the Gaussian distribution with the corresponding parameters that is applicable to all knowledge elements within a notion (G stands for Global) where a free parameter γ_G is fixed for all knowledge elements within and across notions. $N_l(0, \gamma_l)$ is another random number drawn from the Gaussian distribution that is only applicable to notion element l . Unlike γ_G , γ_l varies across knowledge elements, and thus each knowledge element (i.e., θ_l) has its own unique tendency of thoroughness in a hypothesis search process. Specifically, since we assume that useful knowledge elements would usually take non-zero values across a variety of knowledge types, we define γ_l to be a mean the absolute value of corresponding coefficients amalgamating all prior experiences. For example, if a person experienced that information on feature dimension "color" has been more useful than feature dimension "size" in classifying several different categories in the past, then the person is more likely to have thorough knowledge modification on the basis of the applicability of "color" than "size" of input stimuli in new categorization tasks.

The other important element of notion modification (Eq. 5) is a systematic bias given below:

$$A_l(\mathbf{v}, \mathbf{u}, \boldsymbol{\theta}^n(t)) = \sum_h v_{lh} \left[1 + \exp \left(- \sum_l u_{hl} \theta_l^n(t) \right) \right]^{-1} \quad (7)$$

Equation 7 is a simple autoassociative network (AAN) with one nonlinear hidden layer. The fundamental idea behind integrating AAN in the notion modification phase is that even with some stochastic changes in notions resulting from the random knowledge recombination process (i.e., Eq. 4) and the stochastic notion modification process (i.e., Eq. 6), PIKCLE has a tendency to update notions such that the configural pattern of a new notion set is similar to that of previously acquired and applied notions; it is influenced by successful prior experiences. Note that AAN can be considered as a form of data reduction system (e.g. PCA), and thus, through this ANN, PIKCLE might acquire “eigen-notions” that are applicable to various type of concepts and categories to some extent.

The coefficients within ANN (i.e., \mathbf{v}, \mathbf{u}) are learned using a standard online version of gradient descent. We incorporate a *pseudorehearsal* method [6] for learning in ANN, assuming that this ANN serves as long term memory units in the PIKCLE framework. This allows ANN to avoid catastrophic forgetting (e.g. [7] [8]).

In summary, two separate notion modification processes are involved in the knowledge modification phase in PIKCLE. Those two processes are influenced by prior knowledge and experiences acquired in previous concept learning tasks; but the ways in which prior knowledge affect the two processes are different. One type of notion modification is stochastic updating in which previously useful notion elements (i.e. model coefficients) are more thoroughly searched in a current learning task. The other type is systematic updating that—with the influence of previously acquired knowledge—causes PIKCLE to generate new notion sets that resembles previously acquired effective notions.

Selection of Surviving Knowledge. After creating new sets of notions, PIKCLE selects a limited number of notions to be maintained in its memory. In PIKCLE, the survivor selection is achieved deterministically, selecting best notions on the basis of estimated utility of concepts or knowledge. The function defining utility of knowledge is described in the next section.

3.3 Estimating Utility

The utility of each notion or a set of coefficients determines the selection process in PIKCLE, which occurs twice. During categorization, PIKCLE selects a single notion with the highest predicted utility to make a categorization response (referred to as concept utility for response or UR hereafter). During learning, PIKCLE selects best fit notions to update its knowledge (utility for learning or UL hereafter). In both selection processes, the notion utility is subjectively and contextually defined, and a general function is as follows:

$$U(\boldsymbol{\theta}^n) = \mathcal{U}(E(\boldsymbol{\theta}^n), Q_1(\boldsymbol{\theta}^n), \dots, Q_L(\boldsymbol{\theta}^n)) \quad (8)$$

where \mathcal{T} is a function that takes concept inaccuracy (i.e., E) and L contextual factors (i.e., Q) and returns an estimated utility value for a corresponding notion (Note that PIKCLE's learning is framed as a minimization problem). There are virtually infinite contextual functions appropriately defined for Eq. 8 (e.g. concept abstraction, domain expertise and knowledge commonality). For example, in ordinary situations, humans prefer simpler notions (e.g. requiring a smaller amount of diagnostic information to be processed) to complex ones, as long as both are sufficiently accurate, whereas in highly critical tasks (e.g. medical diagnosis), many might choose a notion with the highest accuracy disregarding complexity.

Note that functions for UR and UL do not have to be the same. For example, domain experts often know multiple approaches to categorize objects. This ability appears to be a very important characteristic and thus be a part of their UL. However, "knowledge diversity" is only relevant for selecting a population of notions (for survival), but not for selection of a particular notion to make a categorization response. Thus, knowledge diversity should not be considered for UR.

In PIKCLE, the predicted (in)accuracy of a notion during categorization is estimated based on a retrospective verification function [9], which assumes that humans estimate the accuracies of the notions by applying the current notions to previously encountered instances with a memory decay mechanism. Thus,

$$E(\theta^n) = \sum_{g=1}^G \sum_{k=1}^K \left(\frac{\sum_{\forall i|x^{(i)}=x^{(g)}} (\tau^{(i)} + 1)^{-\delta}}{\sum_{\forall i|x^{(i)}=x^{(0)}} (\tau^{(i)} + 1)^{-\delta}} \right) \left(d_k^{(g)} - O_k^{(n)}(x^{(g)}) \right)^2 \quad (9)$$

where g indicates particular training exemplars, G is the number of unique training exemplars, the last term is the sum of squared error with d being the desired output, and the middle term within a parenthesis is the (training) exemplar retention function defining the strength of the retaining training exemplar $x^{(g)}$ [10]. Memory decay parameter, δ , in the exemplar retention function controls speed of memory decay, and τ indicates how many instances were presented since $x^{(g)}$ appeared, with the current training being represented with "0." Thus, $\tau = 1$ indicates $x^{(g)}$ appeared one instance before the current trial. The denominator in the exemplar retaining function normalizes retention strengths, and thus it controls the relative effect of training exemplar, $x^{(g)}$, in evaluating the accuracy of knowledge or concept. $E(\theta)$ is strongly influenced by more recently encountered training exemplars in early training trials, but it evenly accounts for various exemplars in later training trials, simultaneously accounting for the Power Law of Forgetting and the Power Law of Learning [10].

4 Simulations

In order to investigate the capability of PIKCLE to exhibit cognitive behaviors that resemble real people, two simulation studies were conducted. We created two category structures shown in Table I. There are three binary feature dimensions in those categories.

Table 1. Schematic representation of stimulus set used in Simulation Study

Stimulus Set						
Dim1	Dim2	Dim3	Label - Set 1A	Label - Set 1B	Label - Set 2A	Label - Set 2B
1	1	1	A	C	E	G
1	1	2	A	C	E	H
1	2	1	A	C	F	H
1	2	2	A	D	F	G
2	1	1	B	C	F	G
2	1	2	B	C	F	H
2	2	1	B	D	E	H
2	2	2	B	D	E	G

4.1 Simulation 1

Previous empirical studies indicate that prior knowledge can strongly interfere with the acquisition of new concepts [11] [12]. For example, learning linearly separable categories can become a difficult task when people are informed that categories that are about to learn can be potentially linearly non-separable [12]. In Simulation 1, we sequentially organized two hypothetical concept learning tasks. Our simulated subjects learned Set1A where information on Dim 1 provides sufficient evidence of perfect classification, followed by learning of Set1B. Set 1B would require our simulated subjects to pay attention to all feature dimensions in order to classify all exemplars correctly.

Methods: There were three types of PIKCLE learners involved in the present simulation study, namely BP who has a bias in generating new hypotheses on the basis of prior knowledge and experience in concept learning ($\eta = 0.2$ in Eq. 5) and maintains acquired knowledge; FP who does not have the bias ($\eta = 0.0$) but still possesses previously acquired knowledge in their memory trace; and FF who does not have the bias and forgets acquired notions. For the sake of simplicity, we omitted correlational attention weights from SUPERSET (see Eq. 1). In addition, we also assumed that $D_{ji} = D_{li}, \forall j \& l$. Furthermore, as in typical SUPERSET implementation [4] [5], we defined the knowledge utility function for BP, FP, and FF to be:

$$U(\theta^n) = E(\theta^n) + \lambda_w \sum_k \sum_j (w_{kj}^n)^2 + \lambda_\alpha \sum_i \left[1 + (\alpha_i^n)^{-2} \cdot \sum_l (\alpha_l^n)^2 \right]^{-1} \quad (10)$$

where $E(\theta^n)$ is as given in Eq.9, and $\alpha_i^n = (1 + \exp(-D_i^n))^{-1}$. The second term (i.e., a weight decay function regularizing w) and the third term (i.e., the relative attention elimination function reducing the number of dimensions attended) regularize SUPERSET's knowledge complexity. λ 's are constant free parameters weighting different regularizing factors.

All BP, FP, and FF were run in a simulated training procedure to learn the correct classification responses for the stimuli with corrective feedback. For each category set, there were a total of 10 training blocks, each of which was organized as a random presentation of the eight unique exemplars. The model parameters were selected arbitrarily;

Table 2. Results of Simulation 1

Types	Set 1A				Set 1B			
	Acc	Attn D1	Attn D2	Attn D3	Acc	Attn D1	Attn D2	Attn D3
BP	0.968	0.709	0.133	0.158	0.743	0.531	0.271	0.198
FP	0.969	0.686	0.149	0.165	0.895	0.402	0.413	0.185
FF	0.970	0.712	0.140	0.147	0.951	0.237	0.550	0.184

$c = 5, \phi = 5, \delta = 1, \gamma_G = 0.2, \lambda_w = 0.1, \lambda_\alpha = 1$. Note that the same parameter values were used for all three types of learners, except η . The identical parameter configuration was applied to learning of Set1A and Set1B. The memory sizes for the two models were 10 (i.e., possessing 10 notions at a time). There were a total of 100 simulated subjects for both models.

Predictions: We predict that all three type of learners will be able to learn Set1A very accurately in a similar manner. However, we predict that BP is more likely to be fixated at Dim 1 in the Set1B task, managing only mediocre accuracy performance. FP and FF, on the other hand, would shift his/her attention to Dim 2, thus resulting in a relatively highly accuracy rate.

Results: The results of Simulation 1 are shown in Table 2. As we predicted, all three types of learners resulted in similar accuracies and attention allocation patterns for the Set1A task. For the Set1B task, BP performed the worst, followed by FP then FF. And, as predicted, BP, because of the interference by prior knowledge, paid the most attention to Dim1 which was less diagnostic than Dim2. FP paid more attention to Dim 2 which was imperfect (but the most diagnostic dimension), but it also paid a similar amount of attention to Dim 1. FF paid the most attention to Dim2 and equally less attention to Dims 1 and 3, allowing it to be the best performer for the Set1B task. Note that although FF performed the best in Set1B, this does not necessarily mean that it is the best model in terms of the objective of the present research. Rather, it is probably the worst model, unable to exhibit the interference effect of prior knowledge [11] [12]. Instead, BP’s prediction is most consistent with the results of previous empirical studies.

4.2 Simulation 2

Previous empirical studies indicate that expectation and context affect our concept learning behaviors [11] [12] [13]. In Simulation 2, we let PIKCLE learn hypothetical learning tasks with two hypothetical contextual elements and then observed a knowledge generalization pattern.

Methods: Table 1 shows a schematic representation of category structure used in Simulation 2. There are two category types, Set 2A and Set 2B, both of which are defined by three a feature dimension plus one additional dimension (Dim 4, not shown) representing situational characteristics. Dim 4 was fixed at “1” for the Set2A task and “2” for the Set2B task. Note that both Set2A and Set2B are simple XOR-logic type categories (Dim1&Dim2 for Set2A and Dim2&Dim3 for Set2B). After successful learning,

we gave PIKCLE a generalization tasks to see how situational characteristic influence the choice of notion to be manifested. In particular, we estimated the amount of attention allocated to correlations between the key feature dimensions (Dim1&Dim2 for Set2A and Dim2&Dim3 for Set2B).

There were two types of PIKCLE learners involved in the present simulation study, namely BP and FF. The basic simulation method follows that of Simulation 1, except that there were a total of 50 training blocks; and that we applied a none-restrictive version of SUPERSET. There were a total of four conditions - BP learning Set2A first followed by Set 2B (i.e., BPA condition); BP learning Set2B first (BPB); FF learning Set2A first (FFA); FF learning Set2B first (FFB).

Results & Discussion. For both FFA and FFB, there was a very strong recency effect in that this particular PIKCLE implementation was not sensitive to situational characteristics and was “too adaptive.” A recency effect was also observed in BPA (0.11) and BPB (0.09) conditions, but the magnitudes were much smaller. BP was shown to be affected by prior knowledge and experiences and sensitive to situational characteristics (i.e., in PIKCLE, BP paid attention to the covariation between Dim1 & Dim2 when the situational characteristic or Dim4 is “1”, while it paid attention to the covariation between Dim2 & Dim3 when Dim4 is “2”). The results also indicate that BP is capable of maintaining diverse effective notions.

Although this simulation result seems easily anticipated and perhaps too simplistic given that PIKCLE contains an autoassociative network, it is worth noting that most computational models of concept learning take a form of radial basis function network, which usually does not allow different types of notional elements to directly interact with each other (e.g. predicting association weight between basis unit j and output unit k using association weight between l and m). PIKCLE, on the other hand, allows direct interaction among elements in a notion by integrating the effect of prior knowledge on concept learning.

5 Conclusion

It is well known that our prior knowledge and experiences affect how we learn new concepts (e.g.[1] [2]). However, the cognitive mechanisms that give rise to the prior knowledge effects remains unclear. In the present paper, we introduced a computational cognitive modeling framework that is intended to describe how prior knowledge and experiences influence learning behaviors. Our new model’s main contribution is that it is not simply the prior knowledge stored in our memory trace that influences our learning behaviors, but it is also the learning strategies acquired through previous learning experiences that affect our learning behaviors.

In addition, our new model PIKCLE can acquire general knowledge called “eigen-notions” that are applicable to various type of concepts and categories along with sensitivity to situational characteristics. There is an interesting psychological phenomenon called ad-hoc categories [14], i.e., people can instantaneously and spontaneously create a category, typically in the service of some goal [15]. The cognitive mechanism underlying the creation of ad-hoc categories may be described by the “eigen-notions.”

References

1. Murphy, G., Medin, D.: The role of theories in conceptual coherence. *Psychological Review* 92, 289–316 (1985)
2. Heit, E., Bott, L.: Knowledge selection in category learning. In: Medin, D.L. (ed.) *Psychology of Learning and Motivation* (1999)
3. Rehder, B., Murphy, G.L.: A Knowledge-Resonance (KRES) model of category learning. *Psychonomic Bulletin & Review* 10, 759–784 (2003)
4. Matsuka, T., Sakamoto, Y.: A Model of Concept Formation with a Flexible Representation System. In: *Advances in Neural Networks, ISNN07. Lecture Notes on Computer Science*, vol. 4491, pp. 1139–1147. Springer, Heidelberg (2007)
5. Matsuka, T., Sakamoto, Y.: Integrating a Flexible Representation Machinery in a Model of Human Concept Learning. In: *Proc. of IJCNN2007*. Forthcoming (2007)
6. Robins, A.: Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science* 7, 123–146 (1995)
7. French, R.M.: Catastrophic forgetting in connectionist network. *Trends in Cognitive Sciences* 3, 128–135 (1999)
8. Grossberg, S.: Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science* 11, 23–63 (1987)
9. Matsuka, T., Chouchourelou, A.: A Model of Human Category Learning with Dynamic Multi-Objective Hypotheses Testing with Retrospective Verifications. In: *Proc. IJCNN 2006*, pp. 3648–3656 (2006)
10. Anderson, J.R., Bothell, R., Lebiere, C., Matessa, M.: An integrated theory of list memory. *Journal of Memory and Language* 38, 341–380 (1998)
11. Pazzani, M.: The influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory & Cognition* 17, 416–432 (1991)
12. Wattenmaker, W.D., Dewey, G.L., Murphy, T.D., Medin, D.L.: Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology* 18, 158–194 (1986)
13. Barrett, S.E., Abdi, H., Murphy, G.L., McCarthy Gallagher, J.: Theory-based correlations and their role in children's concepts. *Child Development* 64, 1595–1616 (1993)
14. Barsalou, L.W.: Ad Hoc Categories. *Memory and Cognition* 11, 211–227 (1983)
15. Medin, D.L., Ross, B.H., Markman, A.B.: *Cognitive Psychology* (4th Ed.) Hoboken, NJ: Wiley (2005)