

# A Cognitive Model of Multi-objective Multi-concept Formation

Toshihiko Matsuka<sup>1</sup>, Yasuaki Sakamoto<sup>1</sup>,  
Jeffrey V. Nickerson<sup>1</sup>, and Arieta Chouchourelou<sup>2</sup>

<sup>1</sup> Stevens Institute of Technology, Hoboken, NJ 07030, USA

<sup>2</sup> Rutgers University, Newark, NJ 07102, USA

**Abstract.** The majority of previous computational models of high-order human cognition incorporate gradient descent algorithms for their learning mechanisms and strict error minimization as the sole objective of learning. Recently, however, the validity of gradient descent as a descriptive model of real human cognitive processes has been criticized. In the present paper, we introduce a new framework for descriptive models of human learning that offers qualitatively plausible interpretations of cognitive behaviors. Specifically, we apply a simple multi-objective evolutionary algorithm as a learning method for modeling human category learning, where 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. In addition, unlike gradient descent, our model assumes that humans entertain multiple hypotheses and learn not only by modifying a single existing hypothesis but also by combining a set of hypotheses. This learning-by-combination has been empirically supported, but largely overlooked in computational modeling research. Simulation studies show that our new modeling framework successfully replicated observed phenomena.

## 1 Introduction

The ability to categorize plays a central role in high-order human cognition. By compressing the vast amount of available information, categorization allows humans (and other intelligent agents) to process, understand, and communicate complex thoughts and ideas by efficiently utilizing salient information while ignoring other types. Given the importance of categories as the building units of human knowledge [1], researchers in cognitive science and related fields have shown a great deal of interest in understanding and explaining human category learning.

Human category learning research has been strongly associated with computational modeling as a means to test various theories of how humans encode, organize, and use knowledge. Previous modeling efforts, however, have limited emphases because their principal focus was on describing categorization processes (i.e., forward algorithm specifying how various types of information are integrated to classify items to proper categories), largely overlooking the learning processes (i.e., backward algorithm specifying how category representations are modified based on experiences). In fact, a central debate in the categorization discourse is on how categories are represented (e.g., [2] [3]). The majority of modeling work employed a single learning method, namely

gradient descent (e.g., [2] [3]) with strict (categorization) error minimization as the sole objective of learning.

Recently, Matsuka [4] raised a concern that the lack of emphasis on learning processes will lead to a limited understanding of the nature of human concept formation. One limitation of using the traditional learning algorithms in human category learning research is that it assumes that every human learner modifies his or her category knowledge by local optimization to reduce misclassification irrespective of the contextual factors (e.g., the learner's goals). The result is little variability in learning among the simulated learners. A typical gradient descent optimization method can be considered a normative rather than descriptive account of human learning processes [4].

Unlike a typical gradient descent method, human learning is variable, not based exclusively on error minimization. For example, humans tend to develop simpler category representations instead of complex representations that are slightly more diagnostic in discriminating members of different categories [5]. Moreover, given two stimulus dimensions or features (e.g., size and color) that are equally diagnostic in distinguishing across categories, humans learn to attend to a single, randomly selected diagnostic dimension when making a classification judgment. They ignore the other dimension that is redundant but equally diagnostic [5], leading to individual differences in learning. Unlike human learning, which is often arbitrary and tends to be biased toward simplicity, the gradient descent algorithm always "correctly" updates the category representations, resulting in allocating attention to the two diagnostic dimensions.

The purpose of this work is to introduce a new framework for descriptive models of human learning that offers qualitatively plausible interpretations of cognitive behaviors. We apply a simple evolutionary algorithm as a learning method for modeling human category learning. Here, we show that the evolutionary algorithm can account for variations in human attention allocation during category learning that existing gradient descent models cannot. We conclude with a discussion of how our new framework might be extended to incorporate sensitivity to various contextual factors, such as learners' goals, that humans display.

## 2 Modeling Category Learning with Evolutionary Algorithm

Our new framework is called CLEAR for Category Learning with Evolutionary Algorithm. CLEAR models human learning processes by evolutionary processes where multiple chromosomes compete with one another for survival (i.e., optimization). Each chromosome, consisting of a set of genes, is a particular solution to a problem. The interpretation of each gene is model specific. For example, as in the current modeling approach, a gene may represent a coefficient that corresponds to an attention weight allocated to a stimulus dimension.

CLEAR assumes that human learning involves consideration of multiple solutions (i.e., chromosomes) according to their usefulness in a given task. Each of these solutions determines which aspects of the categories are psychologically salient and which can be ignored. The use of a population of solutions in CLEAR is an innovative contribution to cognitive modeling as virtually all previous categorization models have optimized a single solution using gradient descent on prediction errors (e.g. [2]).

The assumption that humans entertain a range of solutions is consistent with the results from human laboratory experiments [6]. For example, Anderson and Pichert [6] asked people to read a story about a house from the perspective of either a burglar or home-buyer. The story contained pieces of information relevant to one perspective but irrelevant to the other. For example, a color television set was relevant to the burglar but not to the home-buyer. Alternatively, a leaking roof was relevant to the home-buyer but not the burglar. In a free recall task conducted after learning about the story, people recalled more information relevant to the perspective they took than irrelevant information. More important, shifting people's perspective after the initial recall allowed people to recall information that they could not recall previously, suggesting that people indeed have multiple solutions that activate different pieces of information.

Although CLEAR always has multiple solutions in its mind, CLEAR, like the participants in Anderson and Pichert [6], opts for and applies a single solution with the highest predicted utility (e.g., accuracy, score, etc.) to make one response at a time (e.g., categorize an input instance). The functions for estimating the utility for each solution is described in a later section. In the next section, we describe the processes that evolve into solutions in CLEAR.

## 2.1 Learning Via Evolutionary Algorithm

CLEAR utilizes the Evolution Strategy (ES) method (e.g. [7]) for its learning processes. CLEAR, as in a typical ES application, assumes three key processes in learning: *crossover*, *mutation*, and (survivor) *selection*. In the *crossover* process, the randomly selected chromosomes form a pair and exchange gene information, creating a new pair of chromosomes. In human cognition, the crossover process can be interpreted as conceptual combination, in which new solutions are created based on merging ideas from existing solutions that are useful (e.g., creative discovery). In the *mutation* process, each gene (i.e., coefficient) is randomly altered. A mutation can be considered as a modification of a solution by randomly creating a new hypothesis. In the *selection* process, a certain number of solutions are deterministically selected on the basis of their fitness in relation to the environment for survival. Those selected solutions (i.e., chromosomes) will be kept in CLEAR's memory trace (i.e., population space), while non-selected solutions become obsolete or are forgotten.

Unlike previous modeling approaches to category learning research which modify a single solution (i.e., a single set of coefficients), CLEAR maintains, modifies, and combines a set of solutions. The idea of having a population of solutions (as opposed to having an individual solution) is important because it allows not only the selection and concept combination (i.e., crossover processes) in learning, but also the creation of diverse solutions, making learning more robust. Thus, unlike previous models, CLEAR assumes that humans have the potential to maintain a range of solutions and are able to apply a solution most suitable for a particular set of situational characteristics. To our knowledge, these capabilities (i.e., using diverse solutions and learning-by-combination) have not been addressed in the category learning modeling discourse. This may warrant future research. The utility of having homogeneous versus heterogeneous solutions likely depends on situational factors (e.g., a motivation to test a range of strategies) that will vary from one context to another [8].

Another important feature of CLEAR is that it allows the hypothetical error surface to be non-smooth or discontinuous. This characteristic has not been successfully incorporated in cognitive modeling using the gradient descent optimization method. CLEAR, because of the stochastic nature of its optimization method, can incorporate multi-objective functions in learning that are consistent with the complexity of human learning and the possibly discontinuous nature of knowledge utility hypersurface.

Finally, another significant contribution of CLEAR to cognitive modeling is adaptive search strategies, which can be interpreted as analogous to learning rate. As in many recent ES, CLEAR incorporates a self-adoption mechanism for modifying solutions (i.e., coefficient mutation), dynamically altering range of search areas. This mechanism allows CLEAR to be sensitive to the topology of knowledge utility hypersurface (e.g., searching within a smaller area if solutions are close to an optimum). This, in turn, makes CLEAR sensitive to changes in learning objectives. Virtually all previous cognitive models incorporate either static or time-decreasing learning rate.

## 2.2 Categorization Processes in CLEAR

Rather than introducing new forward algorithms, we apply CLEAR's learning processes to ALCOVE's [2] categorization processes. ALCOVE is a computational model of category learning that assumes that humans store every studied instance or exemplar in memory. We chose ALCOVE because of its popularity and demonstrated predictive capability using relatively simple mechanisms in perceptual classification research.

In ALCOVE, categorization decision is based on the activations of stored exemplars. As shown in Equation 1, each exemplar's activation in ALCOVE, scaled by specificity,  $c$  (which determines generalization gradient), is based on the inverse distance between an input,  $x$ , and a stored exemplar,  $\psi_j$ , in multi-dimensional representational space, scaled by dimensional selective attention weights,  $\alpha$ . The exemplar activations are then fed forward to the  $k$ -th output node (e.g., output for category  $k$ ),  $O_k$ , weighted by  $w$ , which determines the strength of association between each exemplar  $j$  and each output node  $k$ :

$$O_k^m(x) = \sum_j w_{kj}^m \left[ \exp \left( -c \cdot \sum_i \alpha_i^m |\psi_{ji} - x_i| \right) \right] \quad (1)$$

where superscript  $m$  indicates  $m$ -th solution being utilized.

The probability of categorizing input instance  $x$  to category  $C$  is based on the activation of output node  $C$  relative to the activations of all output nodes:

$$P(C) = \frac{\exp(\phi \cdot O_C^o(x))}{\sum_k \exp(\phi \cdot O_k^o(x))}. \quad (2)$$

where  $\phi$  controls decisiveness of classification response, and superscript  $o$  indicates the solution adopted to make a categorization response.

In the traditional ALCOVE model, attention ( $\alpha$ ) and association weights ( $w$ ) are updated by gradient descent that minimizes the actual and the predicted value (i.e., Eq. 2). In the current work, a solution consists of these coefficients (i.e., attention and association weights). CLEAR optimizes a set of these solutions. We now describe the functions for estimating the utility of a solution (i.e., attention and association weights).

### 2.3 Learning Process in CLEAR

Since CLEAR is based on Evolutionary Strategy, its genotype space is identical to its phenotype space, except the genes (coefficients) for self-adapting strategy,  $\sigma$ . In CLEAR, there are two coefficients for self-adaptation;  $\sigma_w$  for defining search area for  $w$ , and  $\sigma_\alpha$  for  $\alpha$ . For the sake of simplicity, we use the following notation  $(\mathbf{w}^m, \boldsymbol{\alpha}^m) \in \boldsymbol{\theta}^m$ .

**Hypotheses Combinations.** In CLEAR, randomly selected pairs of solutions are exchanging information to combine knowledge. In particular, CLEAR utilizes discrete recombination of ALCOVE coefficients and intermediary recombination of the coefficient for self-adaptation. Thus, parent solutions  $\boldsymbol{\theta}^{p1}$  and  $\boldsymbol{\theta}^{p2}$  would produce a child solution  $\boldsymbol{\theta}^c$ , where  $\theta_i^c = \theta_i^{p1}$  if  $\text{UNI} \leq 0.5$  or  $\theta_i^{p2}$  otherwise, where UNI is a random number drawn from the Uniform distribution. For self-adapting strategy,  $\sigma_i^c = 0.5 \cdot (\sigma_i^{p1} + \sigma_i^{p2})$ . This combination process continues until the number of children solutions produced reaches the memory capacity of CLEAR.

**Hypotheses Modifications.** After the recombination process, CLEAR randomly modifies its solutions, using a self-adapting strategy. Thus,

$$\sigma_w^m(t+1) = \sigma_w^m(t) \cdot \exp(N(0, \gamma)) \quad w_{kj}^m(t+1) = w_{kj}^m(t) + N(0, \sigma_w^m(t+1)) \quad (3)$$

$$\sigma_\alpha^m(t+1) = \sigma_\alpha^m(t) \cdot \exp(N(0, \gamma)) \quad \alpha_i^m(t+1) = \alpha_i^m(t) + N(0, \sigma_\alpha^m(t+1)) \quad (4)$$

where  $t$  indicates time,  $\gamma$  defines search width (via  $\sigma$ 's), and  $N(0, \sigma)$  is a random number drawn from the Normal distribution with the corresponding parameters.

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

### 2.4 Estimating Utility

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

$$U(\boldsymbol{\theta}^m) = \mathcal{T}(E(\boldsymbol{\theta}^m), Q_1(\boldsymbol{\theta}^m), \dots, Q_L(\boldsymbol{\theta}^m)) \quad (5)$$

where  $\mathcal{T}$  is a function that takes concept inaccuracy (i.e.,  $E$ ) and  $L$  contextual factors (i.e.,  $Q$ ) and returns an estimated solution utility value (Note that CLEAR's learning is framed as a minimization problem). There are virtually infinite contextual functions appropriately defined for Eq. 5 (e.g. concept abstraction, domain expertise and knowledge

commonality). For example, in ordinary situations, humans prefer simpler solutions (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 solution 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 and such an 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 solutions (for survival), but not for selection of a particular solution to make a categorization response. Thus, knowledge diversity should not be considered for UR.

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

$$E(\theta^m) = \sum_{g=1}^G \left[ \left( \frac{\sum_{\forall i|x^{(i)}=x^{(g)}} (\tau^{(i)} + 1)^{-D}}{\sum_g \sum_{\forall i|x^{(i)}=x^{(g)}} (\tau^{(i)} + 1)^{-D}} \right) \sum_k^K \left[ d_k^{(g)} - O_k^m(x^{(g)}) \right]^2 \right] \quad (6)$$

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  $x^{(g)}$ . Memory decay parameter,  $D$ , in the exemplar retention function controls speed of memory decay, and  $\tau$  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.

### 3 Simulations

In order to investigate the capability of CLEAR to replicate observed empirical data, a simulation study was conducted. In particular, we simulated an empirical study that suggests that human concept formation might be driven by more than minimization of classification error [5]. Table 1 shows the schematic representation of the stimulus used in the empirical study and the present simulation study. (Note: Dimensions 1 and 2 are redundant and are also perfectly correlated with the category membership, each being a necessary and sufficient diagnostic dimension). The left column of Fig. 1 shows the observed data. The majority of subjects in the empirical study [5] chose to pay attention primarily if not exclusively to either one of them, as shown in Fig.1 (bottom

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

Category	Stimulus Set			
	Dim1	Dim2	Dim3	Dim4
A	1	1	3	4
A	1	1	4	1
A	1	1	1	2
B	2	2	2	1
B	2	2	3	2
B	2	2	4	3
C	3	3	1	3
C	3	3	2	4
C	3	3	3	1
D	4	4	4	2
D	4	4	2	3
D	4	4	1	4

left), where the amount of attention allocated to each feature was operationally defined by the feature viewing time. Thus, the majority of subjects seemed to have at least two learning objectives: (1) to minimize categorization error and (2) to minimize knowledge complexity. In the present study we simulated this multi-objective learning phenomenon with CLEAR.

Another interesting phenomenon reported in that study is that some subjects who learned to pay attention to either dimension exclusively reported that they realized that there was another diagnostic dimension, indicating the possibility of possessing multiple concepts or solutions. Although it is uncertain whether these subjects were deliberately acquiring multiple solutions or not, we simulated this type of learners as well.

**Methods:** There were two types of CLEAR learners involved in the present simulation study, namely SA who tries to acquire simple accurate classes of knowledge, and MSA whose learning objective is to acquire multiple simple accurate class of knowledge. Equations 7 and 8 describe knowledge utility functions for learning (UL) for SA and MSA, respectively. For both models, Eq.7 is used for their UR, as knowledge diversity should have no effect in selecting a solution to make a response. The knowledge utility for SA (and UR of MSA) is given as:

$$U_{a+s}(\theta^m) = E(\theta^m) + \lambda_w \sum_k \sum_j w_{kj}^2 + \lambda_\alpha \sum_i \left[ 1 + \alpha_i^{-2} \cdot \sum_l^I \alpha_l^2 \right]^{-1} \quad (7)$$

where the first term is defined as in Eq.6, the second term is a weight decay function regularizing  $w$ , the third term is relative attention elimination function reducing the number of dimensions attended, and  $\lambda$ 's are scalars weighting different contextual factors. The knowledge utility for learning (UL) for MSA is given as:

$$U_{a+s+d}(\theta^m) = U_{a+s}(\theta^m) \cdot \left[ 1 + \sum_n \xi(\delta(m, n)) \right] \quad (8)$$

where the last term controls diversity (penalizing having "similar" concepts),  $\delta$  indicates the distance between solution  $m$  and  $n$ , and  $\xi$  is defined as

$$\xi(\delta(m, n)) = \begin{cases} 1 - (r^{-1} \cdot \sqrt{\sum_i (\theta_i^m - \theta_i^n)^2})^2, & \text{if } \sqrt{\sum_i (\theta_i^m - \theta_i^n)^2} \leq r \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

where  $r$  is a parameter defining the penalizing radius. The diversity controlling function is an application of a modified version of the fitness sharing method [7].

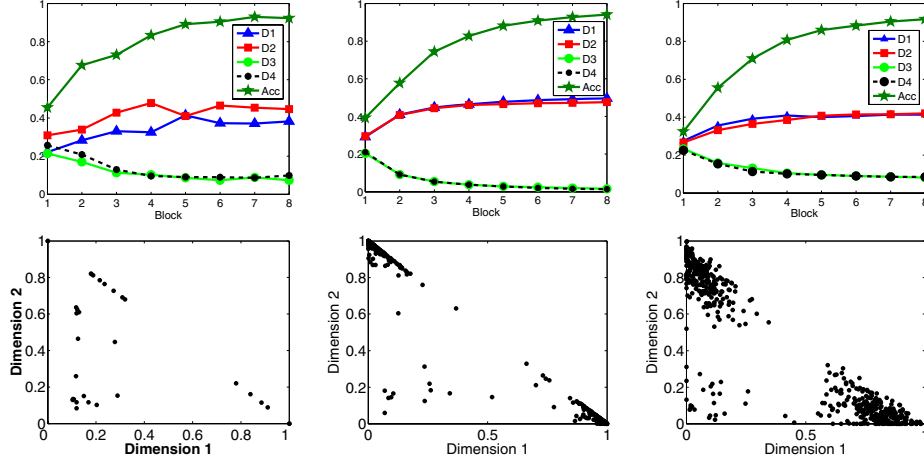
Both models were run in a simulated training procedure to learn the correct classification responses for the stimuli with corrective feedback. The basic training procedures followed that of the original study [5]. There were a total of 8 training blocks, each of which was organized as a random presentation of 12 unique exemplars. The model parameters were selected arbitrarily;  $c = 1.25$ ,  $\phi = 5$ ,  $D = 1$ ,  $\gamma = 0.5$ ,  $\lambda_w = 0.01$ ,  $\lambda_\alpha = 0.75$ ,  $r = 1.5$ . Note that the same parameter values were used for SA and MSA, except  $r$ , which was only applicable for MSA. The memory sizes for the two models were 10 (i.e., possessing 10 solutions at a time). There were a total of 500 simulated subjects for both models.

**Results & Discussion:** Figure 1 shows the results of the simulation study. Both SA and MSA successfully replicated the observed learning curve and aggregated attention learning curves (Fig. 1 top row). In replicating individual attention allocation patterns, SA seemed more consistent with the empirical data than MSA, indicating that (the majority of) human subjects in that study might have had simplicity and accuracy as their learning objectives, but not concept diversity.

To test if MSA had acquired diverse solutions, we investigated all solutions in MSA's memory trace and compared them with those of SA's. Let us define the diagnostic feature dimension selected to be attended by the manifesting concept (i.e., concept or solution whose UR value was the lowest) as a *dominant* dimension and the other as a *recessive* dimension. For MSA, the average difference in the relative amount of attention to be allocated to a recessive and dominant dimensions was 0.24 in the latent concepts (i.e., solutions whose UR values are not lowest), whereas that for SA was 0.01. The maximum differences were 0.67 and 0.21 for MSA and SA, respectively. While the differences may not be great, MSA successfully acquired more diverse solutions than SA. Although with different parameter configurations, we observed greater degrees of differences, there were too much attentional shifts, exhibiting too much alternation of the dominant and recessive dimensions than empirical data.

Possible reasons for the smaller degree of difference was that the solution diversity in MSA took into account both association and attention weights. We might have achieved a greater degree of diversity in attention allocation, if we only controlled attention distribution patterns. But, there was no clear empirical or theoretical justification for embedding such a cognitive mechanism, and thus we did not test the possibility. Furthermore, more explicit diversity control learning mechanisms, such as Pareto-optimal search methods, would find more diverse effective solutions, we are uncertain about their empirical or theoretical justification for such learning mechanisms as cognitive processes.





**Fig. 1.** Left column: Observed empirical results. Middle column: SA, Right Column MSA. The graphs in the top row show classification accuracies and the amounts of relative attention allocated to the four feature dimensions. The scatter plots compare relative attention allocated to Dimensions 1 and 2 for the last 3 blocks for empirical data and the last block for SA and MSA.

## 4 Conclusion

We introduced a CLEAR framework, which used an evolutionary algorithm as a learning method for computational models of human category learning. CLEAR, unlike gradient descent, uses multiple solutions (e.g., attention and association weights). Each solution can be regarded as a filtering device that determines which aspects of the stored information become salient and which should be ignored. The use of multiple solutions allows CLEAR to have a diverse set of solutions. Because different solutions filter different pieces of information, CLEAR can account for variability across learners and tasks that gradient descent method cannot.

As an initial step, we applied CLEAR to ALCOVE's categorization processes. ALCOVE is an exemplar model of category learning that assumes that humans store every previously encountered example in memory. CLEAR optimized multiple solutions, each of which highlighted different aspects of the stored exemplars. CLEAR was able to correctly replicate observed category learning phenomena that were characterized as multi-objective optimization. In addition, CLEAR was able to acquire diverse solutions during such multi-objective learning processes.

**Future Directions.** Future research might address how different tasks and goals are linked to different solutions in CLEAR. Research on this issue involves specifying the contextual factor in CLEAR's utility calculation. Human cognitive processing is greatly affected by situational factors, such as goals and motivations, and thus the contextual factor in CLEAR likely plays an important role in modeling human cognition. For example, even when people are learning about the same categories, their knowledge about the categories can vary when they engage in different tasks and have different goals

(e.g., [8]). Specifying the contextual factor in CLEAR to incorporate sensitivity to different tasks and goals may allow CLEAR to form ad hoc categories to achieve temporary goals, such as things to sell at a garage sale to dispose of unwanted possessions, like humans do [10].

**Final Note.** Most previous category learning models have used the gradient descent optimization method, which can be regarded as a normative rather than a descriptive account of human learning processes [4]. Unlike the gradient descent method, human learning varies across individuals and contexts, and is based on factors other than misclassification. The present work is an initial attempt to develop a more descriptive account of human learning by applying a simple evolutionary algorithm. More rigorous research in descriptive models of human learning is indispensable in the structural elucidation of cognitive information flow and, more generally, in the advancement in cognitive science.

## Acknowledgements

This research was supported in part by the Office of Naval Research, Grant # N00014-05-1-00632.

## References

1. Solomon, K. O., Medin, D. L., Lynch, E.: Concepts Do More Than Categorize. *Trends in Cognitive Sciences* 3 (1999) 99–105
2. Kruschke, J. K.: ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. *Psychological Review* 99 (1992) 22–44
3. Love, B. C., Medin, D. L., Gureckis, T. M.: SUSTAIN: A Network Model of Human Category Learning. *Psychological Review* 111 (2004) 309–332
4. Matsuka, T.: Simple, Individually Unique, and Context-Dependent Learning Method for Models of Human Category Learning. *Behavior Research Methods* 37 (2005) 240–255
5. Matsuka, T., Corter, J. E.: Process Tracing of Attention Allocation in Category Learning. (2006) Under review.
6. Anderson, R. C., Pichert, J. W.: Recall of Previously Unrecallable Information Following a Shift in Perspective. *Journal of Verbal Learning and Verbal Behavior* 17 (1978) 1–12
7. Eiben, A. E., Smith, J. E: *Introduction to Evolutionary Computing*, Berlin: Springer-Verlag (2003)
8. Markman, A. B., Baldwin, G. C., Maddox, W. T.: The Interaction of Payoff Structure and Regulatory Focus in Classification. *Psychological Science* 16 (2005) 852–855
9. Matsuka T., Chouchourelou, A.: A Model of Human Category Learning with Dynamic Multi-Objective Hypotheses Testing with Retrospective Verifications. In *Proc. IJCNN 2006*. (2006) Forthcoming.
10. Barsalou, L. W.: Ad Hoc Categories. *Memory and Cognition* 11 (1983) 211–227