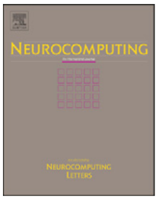
Neurocomputing 71 (2008) 2446–2455

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Toward a descriptive cognitive model of human learning

Toshihiko Matsuka [a,*], Yasuaki Sakamoto [b], Arieta Chouchourelou [c], Jeffrey V. Nickerson [b]

[a] *Department of Cognitive and Information Science, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan*
[b] *Stevens Institute of Technology, Center for Decision Technologies, Hoboken, USA*
[c] *Cyprus College, School of Humanities and Social Sciences, Nicosia, Cyprus*

ARTICLE INFO

Available online 7 May 2008

Keywords:
Cognitive modeling
Human learning
Evolutionary computation

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 important observed psychological phenomena.

1. Introduction

When we think, reason, and communicate, we have a tendency to use knowledge at a high level of abstraction. For example, instead of describing specific features—"a hairy animal with pointed teeth", we use the word "dog" to categorically represent the entity. Although we may lose some information through abstraction, the use of categorical knowledge is essential: otherwise our capacity-limited cognitive systems become overloaded [10]. By compressing the vast amount of available information, a cognitive process called *categorization* allows us to process, understand, and communicate complex thoughts and ideas by efficiently utilizing salient and relevant information while ignoring other types of information. This is why cognitive scientists argue categorization plays a central role in high-order human cognition.

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 their own scope because their principal focus remains on describing categorization processes, largely overlooking the importance of the learning processes. More precisely, previous studies have laid emphasis on forward algorithms that specify how various types of information are integrated in order to classify items into proper categories, but have de-emphasized backward algorithms that explain how category representations are modified based on experience.

Recently, this limited emphasis on learning processes in categorization has increasingly become recognized as an important issue to be examined in order to better understand the nature of human category learning or concept formation (e.g., [15,18]). The central issue is that a majority of learning algorithms embedded in existing cognitive models can be considered a normative account of human learning processes (how we *should* think) rather than a descriptive account (how we *really* do think) [18–20]. Most existing cognitive models employ a single learning method—a gradient descent method with strict (categorization) error minimization as the sole objective of learning—without careful evaluation of its descriptive property as a model of a human mind [18,20]. For instance, one limitation of using traditional learning algorithms in human category learning research is that it assumes that everyone carries out *normatively* oriented optimization process to reduce misclassification irrespective of the contextual factors such as the learner's goals (see [13]), essentially disregarding the situational adaptability of human mind. Another limitation of these normative accounts is that they disregard the possibility that both stochasticity and arbitrary decision making are part of human learning. As a result

---

[*] Corresponding author. Tel.: +81 43 290 3578; fax: +81 43 290 2278.
E-mail address: matsuka@cogsci.L.chiba-u.ac.jp (T. Matsuka).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.12.039

of these limitations, there is little variability in learning among simulated learners [18–20].

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 [21]. 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 [21], leading to individual differences in learning. Unlike human learning, which is often arbitrary and tends to be biased toward simplicity [11], the standard gradient descent algorithm always "correctly" updates the category representations, resulting in allocating attention to the two diagnostic dimensions [18] (but see [20]).

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 algorithms

Our new framework is called CLEAR for Category Learning with Evolutionary AlgoRithms. It models human learning processes by evolutionary processes where multiple notions (i.e., one complete set of a model's coefficients, equivalent to chromosomes in evolutionary computation) compete with others for survival. Each notion, consisting of a set of notional attributes, is a particular type of knowledge about a category. The interpretation of each notional attribute or simply attribute is model specific. For example, as in the current modeling approach, an attribute may represent a coefficient that corresponds to an attention weight allocated to a stimulus dimension. The hierarchical organization of knowledge or information in the CLEAR framework is shown in Table 1.

CLEAR assumes that human learning involves consideration of multiple notions—sets of model coefficients—according to their usefulness in a given task. Each of these notions determines which attributes of the categories are cognitively salient and which can be ignored. The capability of maintaining a population of notions gives CLEAR the chance to discover more profound and thorough concept. The integration of this cognitive machinery is an innovative contribution to cognitive modeling as virtually all previous categorization models only allow one notion for one category, neglecting the potential of humans to posses diverse and context-sensitive notions.

The assumption that humans entertain a range of notions is consistent with the results from human laboratory experiments [4]. For example, Anderson and Pichert [4] 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 were assigned. More importantly, 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 notions that activate different pieces of information. Likewise, Bourne et al. [6] observed that those who exhibited categorization tendencies that were consistent with exemplar use in training showed more rule use in the later transfer task. The results suggest that subjects had in fact acquired multiple notions in the study, and that a slight change in situational characteristics altered knowledge utility, which in turn led to different notions being selected.

Recent empirical and simulation studies suggest the inadequacy of error minimization as a sole mechanism of real human concept formation [18,21]. Real-world human learning is more likely to be explained as a process of multi-objective optimization of utility of knowledge, where objectives are not only defined by categorization errors, but also by each learner's intra- and inter-contextual factors. For example, one laboratory experiment showed that in ordinary situations humans have a preference for simpler, yet sufficiently accurate concepts over complex but marginally more accurate concepts, suggesting their objectives in learning the concept were defined by both accuracy and simplicity [21]. Likewise, individuals' subjective goals or motivations influence their paths and outcomes of concept learning (e.g., [5,17]). A simulation study [18] showed that only models that are sensible to contextual characteristics reliably replicated some important empirical phenomena. Accordingly, CLEAR assumes that learning is driven by an optimization of the subjectively and contextually defined utility of knowledge being acquired, rather than by a simple classification error minimization routine. This is another novel contribution of our new framework.

2.1. Learning via evolutionary algorithm

CLEAR utilizes the evolution strategy (ES) method for its learning processes. As in a typical ES application, it assumes three key processes in learning: *crossover*, *mutation*, and (survivor) *selection*. In the *crossover* process, the randomly selected notions form a pair and exchange information on notional attributes, creating a new notion. In human cognition, the crossover process corresponds to knowledge (re)combination, in which new notions are created by merging ideas from existing effective notions. In

Table 1
Hierarchy of knowledge or information in different domains

Evo. comp.	Human mind	Models (general)	CLEAR
Gene	Attribute	Coefficient	a_i, w_{kj} σ_θ
Chromosome	Notion	Vector of coefficients (i.e., one complete set of a model's coefficients)	$(\mathbf{a}'^{(n)}, \mathbf{w}'^{(n)}, \boldsymbol{\sigma}'^{(n)}) \in \mathbf{z}'^{(n)}$
Population	Concept	Matrix of coefficients (i.e., a population of coefficients)	$\mathbf{Z} = \begin{bmatrix} \mathbf{z}'^{(1)} \\ \mathbf{z}'^{(2)} \\ \vdots \\ \mathbf{z}'^{(n)} \end{bmatrix}$

the *mutation* process, each attribute (i.e., coefficient) is randomly altered. A mutation can be considered as a knowledge modification accomplished by revising each notion with a randomly generated hypothesis about each attribute. 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—in other words, notions are chosen on the basis of subjectively and contextually defined knowledge utility. Those selected notions will be kept in CLEAR's memory trace (the population space), while non-selected notions become obsolete or are forgotten.

Unlike previous modeling approaches to category learning research which modify a single notion (i.e., a single set of coefficients), CLEAR maintains, modifies, and combines a set of notions. The idea of having a population of notions (as opposed to having an individual notion) is important because it allows not only the knowledge selection or concept combination in learning, but also the creation of diverse notions, making learning more robust. Thus, our framework allows simulated individuals to have the potential to maintain a range of notions and the ability to apply a notion most suitable for a particular set of situational characteristics. The utility of having heterogeneous versus homogeneous notions, however, likely depends on situational factors (e.g., a motivation to test a range of strategies) that will vary from one context to another [17].

Another important feature of CLEAR is that it allows the hypothetical error (or knowledge utility) surface to be non-smooth or discontinuous. Some human empirical studies (e.g., [7]) have suggested that a human's concept space might have a non-smooth or discontinuous property. 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 the knowledge utility hypersurface.

Finally, the other notable contribution of CLEAR to cognitive modeling is the presence of adaptive search strategies, which are analogous to learning rates. As in many recent ES, our model incorporates a self-adoption mechanism for modifying notions—coefficient mutation—which dynamically alters the range of search areas. This mechanism allows CLEAR to be sensitive to the topology of the knowledge utility hypersurface. For example, CLEAR will search within a smaller area if notions are close to an optimum. This, in turn, makes it sensitive to changes in learning objectives. Virtually all previous cognitive models are insensitive to the topology, incorporating either static or time-decreasing learning rates.

2.2. Categorization processes in CLEAR

Rather than introducing new forward algorithms, we apply CLEAR's learning processes to ALCOVE's [14] 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 Eq. (1), each exemplar's activation in ALCOVE, scaled by specificity, β (which determines generalization gradient), is based on the inverse distance between an input, x, and a stored exemplar, ψ_j, in multi-dimensional representational space where each dimension is scaled by non-negative selective attention weights, a. The exemplar activations are then fed forward to the k-th output node (e.g., output for category k), O_k, weighted by w_{kj}, which determines the strength of association between each exemplar j and each output node k:

$$O_k^{(n)}(x) = \sum_j w_{kj}^{(n)} \left[\exp\left(-\beta \cdot \sum_i a_i^{(n)} |\psi_{ji} - x_i| \right) \right] \tag{1}$$

where superscript n indicates n-th notion 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:

$$P(C|x) = \frac{\exp(\phi \cdot O_c^{(v)}(x))}{\sum_k \exp(\phi \cdot O_k^{(v)}(x))} \tag{2}$$

where ϕ controls decisiveness of the classification response, and the superscript v indicates the notion adopted to make a categorization response.

Although CLEAR-augmented ALCOVE would always have multiple notions in mind, it, like the participants in Anderson and Pichert's empirical studies [4], opts for and applies a single notion with the highest predicted utility, indicated by the superscript v, to make one response at a time (e.g., categorize an input instance).

In the traditional ALCOVE model, a single notion consisting of attention (i.e., a_i) and association weights (i.e., w_{kj}) are updated by a gradient descent method to minimize the classification error. CLEAR-augmented ALCOVE optimizes multiple notions on the basis of their utility using an evolutionary computing method. We now describe the algorithms for optimizing the utilities of notions.

2.3. Learning process in CLEAR

Since CLEAR is based on an evolutionary strategy optimization method, no special encoding is involved: its genotype space is identical to its phenotype space [9]. There are, however, two special coefficients involved in CLEAR; the attributes (coefficients) for a self-adapting strategy, σ_w that defines a search width for association weights (this is analogous to a learning rate for association weights), and σ_a for dimensional attention weights. As their names suggests, σ_w and σ_a are also learned and adjusted throughout the learning process, as CLEAR searches for "optimal" learning rates while remaining sensitive to the topology of the knowledge utility hypersurface.

For the sake of simplicity, we use the following notation:

$$\mathbf{z}'^{(n)} = \langle w_{11}^{(n)}, \ldots, w_{KJ}^{(n)}, a_1^{(n)}, \ldots, a_I^{(n)}, \sigma_w^{(n)}, \sigma_a^{(n)} \rangle \tag{3}$$

where K is the number of categories, J, the number of memorized exemplars, I, the number of inputs' feature dimension, and the superscript n indicates notions.

2.3.1. Knowledge combinations

In CLEAR, randomly selected pairs of notions exchange information to combine knowledge. In particular, CLEAR utilizes discrete recombination of ALCOVE coefficients and intermediary recombination of the coefficient for self-adaptation. Thus, parent notions $\mathbf{z}^{(\mathrm{p1})}$ and $\mathbf{z}^{(\mathrm{p2})}$ would produce a child notion $\mathbf{z}^{(\mathrm{c})}$ in the following manner:

$$z_l^{(\mathrm{c})} = \begin{cases} z_l^{(\mathrm{p1})} & \text{if UNI} \leqslant 0.5 \\ z_l^{(\mathrm{p2})} & \text{otherwise} \end{cases} \tag{4}$$

where UNI is a random number drawn from the uniform distribution and the subscript l indicates learnable coefficients (e.g., association weights; see Eq. (3)). For a self-adapting strategy, $\sigma_l^{(\mathrm{c})} = 0.5 \cdot (\sigma_l^{(\mathrm{p1})} + \sigma_l^{(\mathrm{p2})})$. This combination process continues until the number of children notions produced reaches the memory capacity of CLEAR.

2.3.2. Knowledge modifications

After the recombination process, CLEAR randomly modifies its notional attributes, using a self-adapting strategy. Thus,

$$w_{kj}^{(n)}(t+1) = w_{kj}^{(n)}(t) + N(0, \sigma_w^{(n)}(t+1)) \tag{5}$$

where t indicates time, and $N(0, \sigma_w^{(n)})$ is a random number drawn from the normal distribution with the corresponding parameters, and

$$\sigma_w^{(n)}(t+1) = \sigma_w^{(n)}(t) \cdot \exp(N(0, \gamma)) \tag{6}$$

where γ is a constant defining a general search width. Similarly,

$$a_i^{(n)}(t+1) = a_i^{(n)}(t) + N(0, \sigma_a^{(n)}(t+1)) \tag{7}$$

$$\sigma_a^{(n)}(t+1) = \sigma_a^{(n)}(t) \cdot \exp(N(0, \gamma)) \tag{8}$$

As noted earlier, these self-adaptive learning rates are attributes of a notion (i.e., $\{\sigma_w^{(n)}, \sigma_a^{(n)}\} \in \mathbf{z}^{(n)}$), and thus they are also subject to CLEAR's learning process. Self-adaptive learning rates that yield a "better" notion are more likely to be retained and then applied in future stages of hypothesis generation. In other words, CLEAR's learning process involves a simultaneous optimization of association weights, selective dimensional attention weights, and learning rates on the basis of knowledge utility.

2.3.3. Selection of surviving hypotheses

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

2.4. Estimating utility

CLEAR assumes that learning is driven by an optimization of the subjectively and contextually defined utility of knowledge being acquired. The utility of each notion or a set of coefficients governs the two selection processes. During categorization, CLEAR 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, CLEAR 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 given as

$$U(\mathbf{z}^{(n)}) = \Upsilon(E(\mathbf{z}^{(n)}), Q_1(\mathbf{z}^{(n)}), \ldots, Q_L(\mathbf{z}^{(n)})) \tag{9}$$

where Υ is a function that takes concept inaccuracy (i.e., E) and L contextual factors (i.e., Q) and returns an estimated notion utility value. (Note that CLEAR's learning is framed as a minimization problem.) There are a virtually infinite number of contextual functions appropriately defined for Eq. (9). For example, in ordinary situations, humans prefer simpler notions to more complex ones. A function may require only small amount of diagnostic information to be processed, gaining simplicity while losing some accuracy. Weight decay models have this characteristic: $Q(z^{(n)}) = \lambda_w \sum_{kj} (w_{kj}^{(n)})^2 + \lambda_a \sum_i (a_i^{(n)})^2$ where λ's are weighting scalars. On the contrary, in a highly critical task such as medical diagnosis, many will prefer a notion with the highest accuracy, without regard to the complexity cost. Another example showing that accuracy alone cannot explain knowledge utility may be seen our everyday life: the acquisition of knowledge that is communicative (e.g., $Q(z^{(n)}) = 1 - |CORR(z^{(n)}, z^*)|$ where z^* indicates "social norm") is one important goal in social settings.

Note that functions for UR and UL do not have to be the same. For example, a doctor's UL would include "thoroughness" or "attentiveness", but the same doctor's UR may be highly influenced by "speediness" in an emergency scenario. Another example is that 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 (e.g., $Q(z^{(n)}) = (\sum_{i \neq n} dist(z^{(n)} - z^{(i)}))^{-1}$, where $dist$ is some function estimating distance between two notions). 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.

The last example illustrates another notable contribution of the use of a population of notions (a set of notions comprising a concept). The utility of knowledge can be characterized by a set of functions defined at different levels of knowledge hierarchy. For example, categorization accuracy is defined at the notion-level while knowledge diversity or thoroughness (having a variety of approaches to categorize) is defined at the concept-level, which is effectively a set of notions. Because existing models allow only one notion per concept, such a complex definition of knowledge utility has been unattainable.

2.5. Estimation of accuracy of hypothesis sets

Stochastic optimization method are typically implemented using batch-learning, because batch-learning provides a reliable and stable estimate of an objective function. We, however, believe learning occurs in an instance-by-instance basis in humans (e.g., [12]). CLEAR, therefore, should generate and test a set of notions for each training instance. But, if the accuracies of concepts are estimated on the basis of the current training instance alone, then CLEAR may over-generalize. This in turn can lead to oscillation, even though evolutionary strategy can be considered to have a memory effect (ideas persist in the gene pools).

Some empirical and theoretical studies have suggested that humans not only utilize performance on current instances but also utilize memorized past instances in learning (e.g., [12,22]). This cognitive machinery is integrated in CLEAR. Note that there are various theories and psychological phenomena associated with memory mechanisms in general human cognition (e.g., [1,30]) and in human categorization or category learning (e.g., [10]). However, since those mechanisms and phenomena are very complex and intertwined, and there is a subfield of cognitive modeling research surrounding these phenomena, it is beyond the scope of the present study to incorporate memory mechanisms of such sophistication. Instead, we choose to incorporate a rather simple memory retention mechanism that appears to capture ordinary human memory in general situations. Specifically, a modified version of Anderson and Schooler's [2,3] memory retention function is integrated in the CLEAR framework. This function simultaneously accounts for learning and forgetting mechanisms.

A general function of classification (in)accuracy, $E(\mathbf{z}^{(n)})$, can be formulated as

$$E(\mathbf{z}^{(n)}) = \sum_{g=1}^{G} \sum_{k=1}^{K} \Xi(d_k^{(g)}, x^{(g)})[d_k^{(g)} - O_k^{(n)}(x^{(g)})]^2 \tag{10}$$

where superscript g indicates a particular input–output pair, G is the number of unique training pairs, and the last term is the sum of squared error with d being the desired output. Ξ is a general memory retention function, defining the strength of retention of a particular input–output pair, that incorporates some theories on human memory, such as memory capacity, memory decay, and

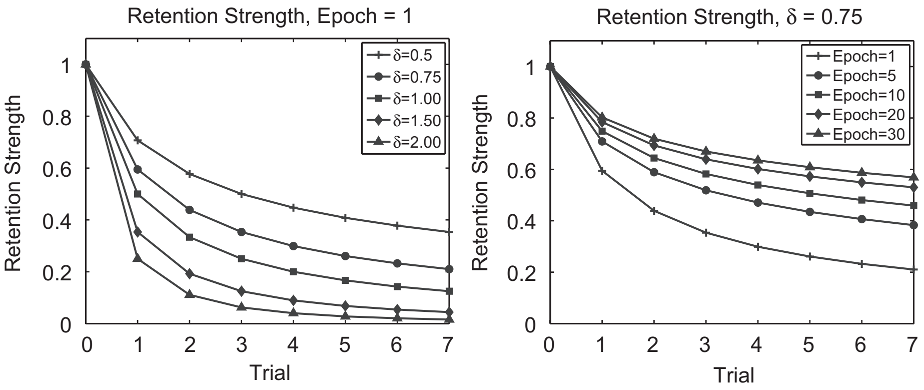


Fig. 1. Left panel: the retention strengths for different memory decay parameters in the first training epoch. Right panel: the retention strengths at different training epochs with the memory decay parameter (δ) being 0.75.

practice effect. This knowledge accuracy function can be interpreted in the following way: humans will realize the accuracy of a new hypotheses set not only by applying the hypotheses to the current training-instance-to-category relationship but also by applying the hypotheses to several previously encountered instances, retrospectively verifying the current hypotheses.

In CLEAR, because of the assumption of memory decay (and the practice effect in some cases), previously encountered instances, at different times, have a different influence on hypotheses accuracy estimates (e.g., [2,3]). In particular, the memory retention function, Ξ determines the effect of previously encountered instances in realizing the accuracy of new hypothesis sets. There are several theories and models applicable for the Ξ function, and a relatively simple yet robust example is the function introduced by Anderson and Schooler's [3]. Their model is particularly appealing in our framework, because it simultaneously accounts for both the power law of forgetting [28,30] and the power law of learning [23]. Phenomena of forgetting and learning of training exemplars are both important in ordinary category learning where a limited numbers of exemplars are typically presented multiple times in a repetitive training, which in turn causes humans to retain the training exemplars better in later training phases than in earlier phases (the practice or learning effect) while within a current training block the more recent exemplars are still retained better than the ones presented earlier (the recency effect—or forgetfulness).

By incorporating Anderson and Schooler's function and by making the simplifying assumption that the category structure is deterministic, the knowledge accuracy function can be implemented as follows:

$$E(\mathbf{z}^{(n)}) = \sum_{g=1}^{G}\sum_{k=1}^{K}\left(\frac{\sum_{\forall i|x^{(i)}=x^{(g)}}(\tau^{(i)}+1)^{-\delta}}{\sum_{\forall i|x^{(i)}=x^{(0)}}(\tau^{(i)}+1)^{-\delta}}\right) \times (d_k^{(g)} - O_k^{(n)}(x^{(g)}))^2 \tag{11}$$

where the middle term is the exemplar retention function (i.e., Ξ) defining the strength of retaining training exemplar $x^{(g)}$. The memory decay parameter, δ, in the exemplar retention function controls the speed of memory decay, and τ indicates how many instances were presented since $x^{(g)}$ appeared, with the current training being represented with "0". For example, if $x^{(g)}$ appeared one instance before the current trial, then $\tau = 1$. If $x^{(g)}$ repetitively appeared in first, 10th, and 20th instances before the current trial, then the numerator of the exemplar retaining function becomes $(1+1)^{-\delta} + (10+1)^{-\delta} + (20+1)^{-\delta}$.

The denominator in the exemplar retaining function normalized retention strengths by dividing the strength of the current instance (i.e., $x^{(0)}$). Thus, it controls the relative effect of the training exemplar in evaluating the accuracy of knowledge. Given the power law of forgetting and the power law of learning, E is strongly biased or influenced by training exemplars shown more recently in early training trials. but it evenly accounts for various exemplars in later training trials. Fig. 1 shows illustrative examples.

Table 2
Schematic representation of stimulus set used in Simulation Study 1

Stimulus features			Category types					
Dim1	Dim2	Dim3	T1	T2	T3	T4	T5	T6
1	1	1	A	A	A	A	A	A
1	1	2	A	A	A	A	A	B
1	2	1	A	B	A	A	A	B
1	2	2	A	B	B	B	B	A
2	1	1	B	B	B	A	B	B
2	1	2	B	B	A	B	B	A
2	2	1	B	A	B	B	B	A
2	2	2	B	A	B	B	A	B

3. Simulations

3.1. Simulation 1

In Simulation 1, we simulated a classical study of categorization [29] which is often used as a benchmark [24]. The stimulus structures are shown in Table 2. There were a total of eight training instances defined by three binary feature dimensions (i.e., shape, color, and size). In the study, human subjects were trained to learn to classify those instances into the correct categories with corrective feedback. Shepard et al. [29] created six category structures by varying the complexity of correct categorization strategies. The results of previous empirical studies showed [24,29] that Type 1 (T1) was the easiest to learn to classify, followed by T2–T6 being the most difficult (see Fig. 2 left panel), where the differences in difficulty for T3–T5 were not statistically significant.

T1 was easiest to learn, probably because it only requires a simple one-dimensional rule for a correct categorization. T2 can be considered as XOR-logic, described by Dimensions 1 and 2. T3–T5 are one-dimensional rules with two exceptions (one for each category) where recognition of the exceptions requires consideration of all three feature dimensions. T6 was the most complex as it requires memorization of many if not all exemplars. The results of the original study [29] and its replication [24] have

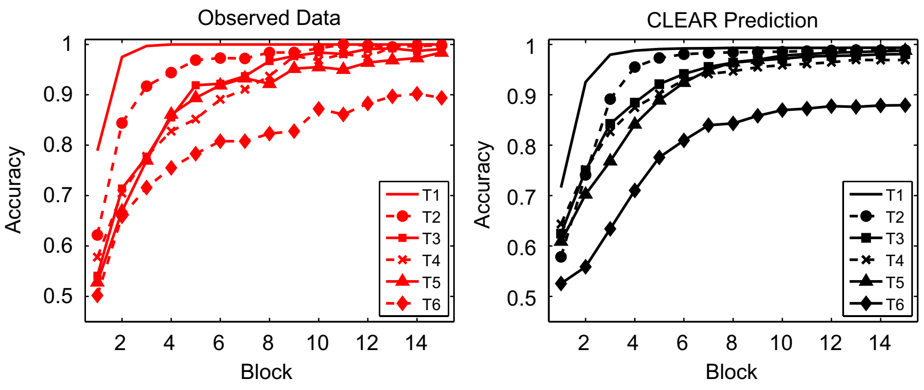


Fig. 2. Left panel: observed empirical results reported in Nosofsky et al. [24]. Right panel: CLEAR's predictions of the same categorization tasks.

been interpreted in the following way: category learners can selectively allocate attention to stimulus features on a dimension-by-dimension basis, and can learn to allocate attention in an optimal or near-optimal manner across stimulus dimensions.

3.1.1. *Methods*

Nosofsky et al. [24] collected data on learning curves for those six category structures, which are shown in Fig. 2. In Simulation 1, we tried to replicate those learning curves with CLEAR.

The basic training procedures follow that of Nosofsky et al.'s [24]. CLEAR was run in a simulated training procedure with 16 trial blocks, where each block consisted of random presentations of the eight unique training exemplars exactly twice, in order to learn the correct classification responses. There are a total of 100 simulated subjects for each category structures. The model configurations and parameters were identical for all six conditions. The model parameters were selected arbitrarily: overall similarity gradient $(\beta) = 1$, decisiveness of categorization response $(\phi) = 5$, memory decay $(\delta) = 1$, overall mutation width $(\gamma) = 0.5$. The memory size for CLEAR was fixed at 10 (i.e., possessing 10 notions at a time).

3.1.2. *Results and discussion*

The right panel of Fig. 2 shows the results of Simulation 1. CLEAR successfully replicated the general trend of the observed data. T1 was the easiest to learn followed by T2, {T3, T4, T5}, and then T6. T3–T5 were similarly difficult to learn. The predicted classification accuracies at the last block closely matched those of the observed data. However, it slightly under-predicted the difficulty of T3–T5 category structures in early training blocks. The prediction may improve if the model parameters were more carefully selected.

In summary, CLEAR successfully replicated one benchmarking category learning task set [24]. Although CLEAR's instance-based learning algorithm is driven by stochastic processes, it was able to learn these categories with a rather small number of training trials. The utilization of multiple notions and the retrospective verification machinery (Eq. (11)) are the key cognitive processes in CLEAR in the present simulation. A predecessor of CLEAR employing a Monte Carlo optimization method without multi-notion nor retrospective verification requires much more training trials in order to learn categories [18,19].

3.2. *Simulation 2*

In Simulation 2, we replicated a recent study that suggests that human concept formation might be driven by the optimization of multi-faceted knowledge utility [21]. Table 3 shows the schematic representation of the stimulus set used in the present simulation. Note that Dimensions 1 and 2 are redundant and are also perfectly correlated with the category membership, each being a necessary and sufficient diagnostic dimension.

Table 3
Schematic representation of stimulus set used in Simulation Study 2

Stimulus set Category	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

The left column of Fig. 3 shows the observed data, including classification accuracy and dimensional attention weights, where the amount of attention allocated to each feature was operationally defined by the feature viewing time. In the empirical study, the majority of subjects were able to categorize the stimuli almost perfectly and its aggregated results suggest that on average subjects paid attention to both of the correlated dimensions approximately equally (Fig. 3, top left panel). More interestingly, when the attention data were analyzed per individual, the majority of subjects in the empirical study [21] chose to pay attention primarily if not exclusively to either one of them, as shown in Fig. 3 (bottom left). Thus, the majority of subjects seemed to have at least two learning objectives: (1) to minimize categorization error and (2) to minimize knowledge complexity. Paying attention to only one dimension in this categorization task leads to a reduction of knowledge complexity, because it allows CLEAR to reduce the number of feature dimensions to be processed. This in turn can also reduce the number of unique exemplars to be memorized and used. That is, if an individual learned to pay attention to Dim1, then only one unique exemplar needs to be processed for each category.

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 notions. Although it is uncertain whether these subjects were deliberately acquiring multiple notions or not,

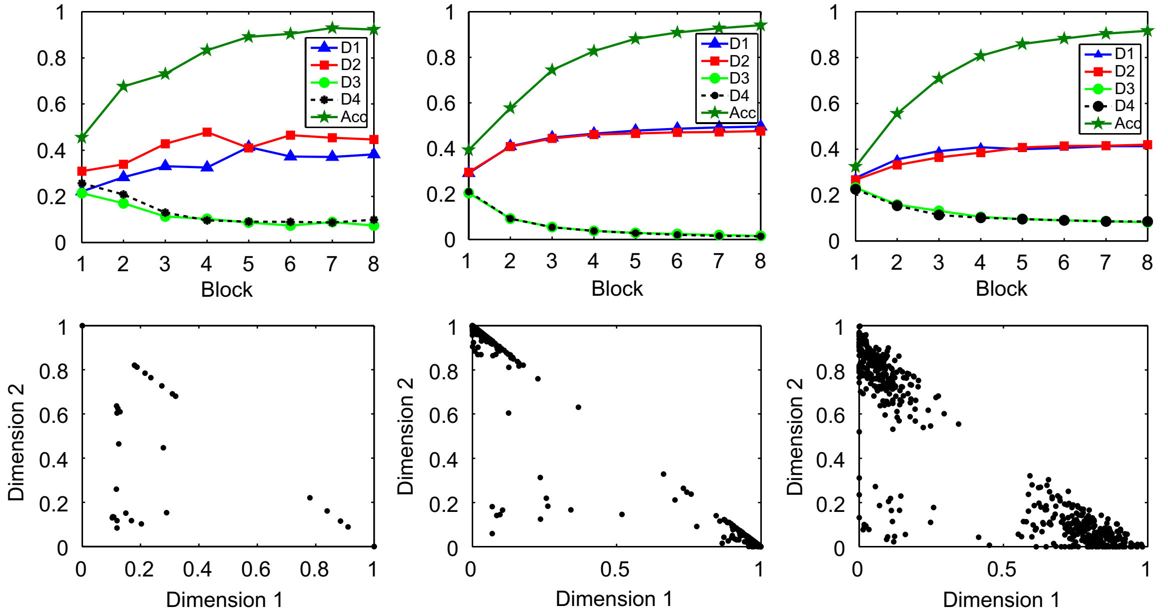


Fig. 3. Left column: observed empirical results. Middle column: SA, right column MSA. The graphs in the top row show classification accuracies and the amounts of relative attention allocated to the four feature dimensions. The scatter plots compare relative attention allocated to Dimensions 1 and 2 for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

Table 4
Schematic representation of stimulus set used in Simulation Study 3

Stimulus set Category	Dim1	Dim2	Dim3	Dim4	Dim5
A	1	1	1	2	1
A	1	1	2	1	1
A	1	2	1	2	1
A	2	1	2	1	1
B	1	2	1	1	1
B	2	1	2	2	1
B	2	2	1	1	1
B	2	2	2	2	1

we simulated learners who posses multiple ways to categorize input stimuli—that is, learners who maintain diverse notions.

3.2.1. 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 classes of knowledge. Eqs. (12) and (13) describe knowledge utility functions for learning (ULs) for SA and MSA, respectively. For both models, Eq. (12) is used for their UR, as knowledge diversity should have no effect in selecting a notion to make a response. The knowledge utility for SA (and UR of MSA) is given as

$$U_{a+s}(\mathbf{z}^{(n)}) = E(\mathbf{z}^{(n)}) + \lambda_w \sum_k \sum_j (w_{kj}^{(n)})^2 + \lambda_a \sum_i \left[1 + (a_i^{(n)})^{-2} \cdot \sum_l^I (a_l^{(n)})^2\right]^{-1} \quad (12)$$

where the first term is defined as in Eq. (11). The second term is the weight decay function, regularizing **w**. The third term is the relative attention elimination function, reducing the number of attended dimensions, thus regularizing **a**. λ's are scalars weighting different contextual factors. The knowledge UL for MSA is given as

Table 5
Results of Simulation Study 3

D1	73.7%	D1 ∩ D2	57.4%	D1 ∩ D2 ∩ XOR	54.8%
D2	76.0%	D1 ∩ XOR	71.1%	D1 ∪ D2 ∪ XOR	100.0%
XOR	96.6%	D2 ∩ XOR	73.2%	{D1 ∩ XOR} ∪ {D2 ∩ XOR}	90.2%

The proportions of simulated subjects who acquired corresponding categorization strategies. D1: a strategy where the most attention ($a_1 > 0.70$) is allocated Dimension 1; D2: a strategy based on Dimension 2 ($a_2 > 0.70$); XOR: a strategy in which both Dimensions 3 and 4 are sufficiently attended ($a_3 > 0.35$ and $a_4 > 0.35$).

$$U_{a+s+d}(\mathbf{z}^{(n)}) = U_{a+s}(\mathbf{z}^{(n)}) \cdot \left[1 + \sum_m \xi(f(m,n))\right] \quad (13)$$

where the last term controls diversity by penalizing the presence of "similar" concepts. The function f indicates the distance between notion m and n, and ξ is defined as

$$\xi(f(m,n)) = \begin{cases} 1 - \left(r^{-1} \cdot \sqrt{\sum_l (z_l^{(n)} - z_l^{(m)})^2}\right)^2 & \text{if } \sqrt{\sum_l (z_l^{(n)} - z_l^{(m)})^2} \leqslant r \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

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 [9].

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 [21]. There were a total of eight training blocks, each of which was organized as a random presentation of 12 unique exemplars. The model parameters were selected arbitrarily; $c = 1.25$, $\phi = 5$, $\delta = 1$, $\gamma = 0.5$, $\lambda_w = 0.01$, $\lambda_a = 0.75$, $r = 1.5$. 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 notions at a time). There were a total of 500 simulated subjects for both models.

3.2.2. Results and discussion

Fig. 3 shows the results of the simulation study. Both SA and MSA successfully replicated the observed learning curve and aggregated attention learning curves (Fig. 3 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 notions, we investigated all notions 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 notion whose UR value was the lowest) as a *dominant* dimension and the other diagnostic dimension 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., notions 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 notions than SA. With different parameter configurations, we were able to observe greater degrees of differences in the amounts of attention allocated to *dominant* and *recessive* dimensions. However, with those parameter configurations, CLEAR also exhibited high degrees of attentional shifts, resulting in too many alternations of the dominant and recessive dimensions (e.g., Dim 1 being the dominant dimension at time t, Dim 2 being the dominant at time $t+1$, and then Dim 1 become the dominant at time $t+2$ again) than empirical data. Since such a highly frequency of attentional shifts were unrealistic, we disregarded these parameter configurations.

Possible reasons for the smaller degree of difference was that the notion 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 notions. However, we are uncertain about the descriptive validity of such learning mechanisms as human cognitive processes.

3.3. Simulation 3

The results of Simulation 2 suggest the potential advantage of applying cognitive information processing methods to solve real-world machine learning problems. In real-world problems, there are many instances where only fragmented information is available at hand for classification, and the way in which the fragmentation occurs is usually unknown a priori. One classification approach for such situations is to pay attention to many diagnostic feature dimensions even if they are highly correlated. This allows an individual to classify instances using information on features that are available. Although this may reduce the influence of missing features on classification accuracy (for training samples), it certainly increases the complexity of the classifier or the knowledge base. This is a problematic consequence: as the complexity of a classifier increases, there is a greater risk of over-generalization (e.g., [32]).

The other classification approach for potentially uncertain situations is to maintain several simpler classification strategies and use one that is applicable for a given situation. With this approach, each notion is kept sufficiently simple, and thus it is less likely to be subject to a generalization problem due to knowledge (or model) complexity. In addition, with a proper level of knowledge diversity, this approach can effectively classify many instances in potentially uncertain situations with incomplete information. With the CLEAR framework, acquiring multiple strategies to solve a classification problem can be achieved with only one single learning task, as shown in Simulation 2.

In Simulation 3, we further investigate CLEAR's potential using a different stimulus structures. Table 4 shows the schematic representation of the stimulus set. Note that Dimension 1 is partially diagnostic; information on Dimension 1 alone delivers 75% accuracy. The same is true for Dimension 2, but note that D1 and D2 are statistically independent. Dimensions 3 and 4 form an XOR logic and they cojointly provide a perfect categorization strategy. Dimension 5 is non-diagnostic constant, and was included to see if CLEAR inappropriately pay attention to this dimension. The main aim of Simulation 3 is to see whether CLEAR is capable of acquiring multiple categorization approaches or notions where separate notions offer different patterns of benefits. For example, a categorization approach based exclusively on D1 is simple but only moderately accurate, whereas a categorization approach based on the XOR logic (D3 and D4) is complex but perfectly accurate. This perfect categorization strategy is generally preferred, but it is effective if and only if features on D3 and D4 are available. If either is missing, an imperfect strategy (e.g., D1 or D2) would result in moderately accurate yet best achievable categorization.

3.3.1. Methods

In Simulation 3, we created MSA-type learners whose learning objective is to acquire diverse simple and accurate categorization strategies (i.e., Eq. (13)). The model was run in a simulated training procedure with corrective feedback. There were a total of 30 training blocks, each of which was organized as a random presentation of eight unique exemplars. The model parameters were selected arbitrarily; $c=5$, $\phi=5$, $\delta=1$, $\gamma=0.5$, $\lambda_w=0.1$, $\lambda_a=0.75$, $r=0.5$. The memory sizes for the model were 10 (i.e., possessing 10 notions at a time). There were a total of 1000 simulated subjects.

3.3.2. Results and discussion

Table 5 shows the results of Simulation 3. The mean classification accuracy in the last training block was 0.984. Approximately 97% of simulated subjects acquired the XOR strategy ($a_3>0.35$ and $a_4>0.35$). A moderate proportion (i.e., 55%) of subjects acquired all three strategies, while a quite high proportion of subjects (approximately 90%) acquired at least one pair of simple and complex classification strategies (i.e., D1 and XOR or D2 and XOR). The frequency of simultaneous acquisition of the three strategies may be increased by increasing the memory capacity of CLEAR (in this simulation CLEAR holds only 10 notions at a time).

In short, CLEAR successfully learned multiple classification strategies that were comprises distinct patterns of knowledge utility. It learned a "complex" but very accurate notions whose applicability may be restricted because of its complexity (it is applicable if and only if D3 and D4 are available). It also learned at least one simple but only moderately accurate notion that is less restrictive. As a whole, CLEAR acquired robust knowledge that is functional in various situations with only one single learning task.

4. Discussion

One main objective in developing the CLEAR framework is to enhance the descriptive characteristics of the embedded learning algorithm, which has been largely overlooked in previous modeling research. There are, however, two potential threats to its descriptive validity. This section describes the potential threats and our responses to the threats.

4.1. *Precision of knowledge utility estimates*

The most cognitively demanding process integrated in CLEAR is probably the estimation of knowledge utility. This apparently complex process might be a challenge to its descriptive validity. That is, some may wonder if humans really do carry out such a process. We believe that human learning involves the estimation of knowledge utility, but at a rather low level of precision. Given that CLEAR hold multiple notions and that the selection of notions is done deterministically using quasi-qualitative (relational) information, we expected CLEAR to achieve successful learning with imprecise estimations of knowledge utility. Maintaining multiple notions can act as buffer allowing CLEAR to have some extra "unfit" notions. In addition, CLEAR selects the first n out of N notions to maintain in its memory trace on the basis of the rank order of the notion's utility estimates. The use of the relational information (i.e., the rank order of notions) makes CLEAR's learning approach indulgent with a coarse computational machinery. In other word, the imprecision in quantitative differences among the utility estimates is less critical in learning in the present framework. Thus, as long as the rank order of notions is adequately correct, CLEAR should acquire situationally suitable knowledge.

In order to evaluate this assertion, we replicate Simulation 2 with noise perturbed knowledge utility estimates. Here, we only considered "SA" learners. For each sub-utility functions (i.e., accuracy, association simplicity, and attentional simplicity), we added zero-mean Gaussian noise. The variances of the noise distributions were determined by that of values estimated for the three sub-utility functions. The results showed that the simulated learners in most cases successfully acquired simple (paying attention primarily to one diagnostic dimension) and accurate (95 + % correct) knowledge.

Although the calculation of accurate knowledge utility values may be complex, we believe that humans do not necessarily examine knowledge at that level of precision. Rather, we believe that humans use (subjective) coarse estimates of knowledge utility in learning. As shown in this auxiliary simulation, those coarse estimates can and often do result in successful knowledge acquisition. In this regard, although our implementation of the learning algorithm may appear complex, we believe that its underlying cognitive demands remains sufficiently low. The lack of a need for precise information is very attractive and plausible in cognitive modeling, because it only requires computation with and processing of small bits of information.

4.2. *Random number generator and complexity of CLEAR-type learning*

The stochastic process plays a significant role in knowledge acquisition in CLEAR. The integration of random number generator functions, however, can be another threat to CLEAR's descriptive validity. Some might ask that can humans produce (quasi) random numbers and/or sequences, and if so, can it be done without overly complicating the computation operations involved category learning? Here, we discuss the neural circuit that is potentially involved in random number generation, as well as the cognitive demand associated with this process.

It has been shown that there is a brain area called the lateral magnocellular nucleus (LMAN) that involves generation of (quasi) random pattern in zebra finch, a type of songbirds [25,26]. In particular, a neural pathway that extends from LMAN to the robust nucleus of the arcopallium was identified as the key neural circuit for the birds' vocal learning which closely resembles a stochastic optimization process such as simulated annealing. The birds produce highly variable outputs as juveniles, but gradually learns to produce a stable song in a trial-and-error fashion, comparing their song and that of a tutor. This neural circuit is homologous to cortico-basal ganglia circuits in mammals. This in turn may suggest that the basal ganglia, which is considered to be more "primitive" than the brain areas associated with high-order cognition such as frontal lobe, plays a role in the generation of random patterns in humans.

Interestingly, one important function of the basal ganglia in mammal is learning [27], including reinforcement learning (RL) [8]. Note that CLEAR's learning algorithm resembles RL in a sense that it searches for a set of notions (i.e., policy in RL) with "optimal" expected utility (i.e., long-term reward in RL) in a probabilistic fashion. To be clear, we do not believe that all learning process in CLEAR is carried out inside the basal ganglia: the estimation of knowledge utility probably involves other areas associated with high-level cognition, such as the frontal lobe.

The cognitive effort associated with random number generation (perhaps in the basal ganglia) is probably low. This, along with the lack of a need for precise knowledge utility estimations, lead us to think that, in spite of its apparently complexity, the actual computational complexity of the learning process that we attempt to replicate with CLEAR is plausibly small. The learning is distributed where each computation involves low numbers of bits of information and simple operations.

5. Conclusion

We introduced the CLEAR framework, which is built on the basis of an evolutionary algorithm, in order to descriptively model human cognitive processes involved in learning of categories. We integrated two important yet often neglected characteristics of human cognition into CLEAR.

A widely accepted yet apparently unrealistic assumption about human category learning is mono-notion learning. That is, it has been assumed that learning involves an identification and maintenance of an "optimal" point in concept space. On the contrary, we believe learning involves an optimization of multiple notions, each of which may have different knowledge utility. This assumption is particularly important in the generation and management of diverse, novel, and constructive concepts by allowing us to combine different notions. In addition, possession of multiple notions permits us to exhibit very adaptive behaviors by allowing us to decide which notion to manifest. For example, when we communicate with children, we have tendency to use a notion with a higher degree of knowledge abstraction than one that would be used to communicate with adults. Such multi-notion learning is one novel contribution of CLEAR to cognitive modeling.

Human learning is compound and variable, not based exclusively on error minimization (e.g., [17,20,21]). We assume human learning is an optimization of a subjectively and contextually defined utility of concepts being acquired. That is, human learning is highly sensitive to internal and external situational characteristics, which we deem a key factor making human cognition adaptive. Specifically, we hypothesize that human learning involves some form of supervised learning which facilitates an acquisition of accurate knowledge, and some form of reinforcement learning which guides learning trajectories to meet situational needs. These two forms of learning cojointly result in the acquisition of contextually "optimal" knowledge or concepts. The second novel contribution of CLEAR is the modeling of this multi-facetted aspect of human learning.

Most previous category learning models have incorporated a form of optimization method that can be regarded as a normative rather than a descriptive account of human learning processes [18]. The present work is an initial attempt to develop a more descriptive account of human learning by applying a qualitatively plausible and quantitatively feasible algorithm. Rigorous research in descriptive models of human learning is indispensable in the structural elucidation of cognitive information flow and, more generally, in the advancement of cognitive science.

Acknowledgment

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

References

[1] J.R. Anderson, The Adaptive Character of Thought, Lawrence Erlbaum, Hillsdale, NJ, 1990.
[2] J.R. Anderson, D. Bothell, C. Lebiere, M. Matessa, An integrated theory of list memory, J. Mem. Lang. 38 (1998) 341–380.
[3] J.R. Anderson, L.J. Schooler, Reflections of the environment in memory, Psychol. Sci. 2 (1991) 396–408.
[4] R.C. Anderson, J.W. Pichert, Recall of previously unrecallable information following a shift in perspective, J. Verbal Learn. Verbal Behav. 17 (1978) 1–12.
[5] L.W. Barsalou, Ad hoc categories, Mem. Cognit. 11 (1983) 211–227.
[6] L.E. Bourne, A.F. Healy, J.T. Parker, T.C. Rickard, The strategic basis of performance in binary classification tasks: strategy choices and strategy transitions, J. Mem. Lang. 41 (1999) 223–252.
[7] G.H. Bower, T.R. Trabasso, Reversals prior to solution in concept identification, J. Exp. Psychol. 66 (1963) 409–418.
[8] J. Brown, D. Bullock, S. Grossberg, How the Basal Ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues, J. Neurosci. 19 (1999) 10502–10511.
[9] A.E. Eiben, J.E. Smith, Introduction to Evolutionary Computing, Springer, Berlin, 2003.
[10] W. Estes, Classification and Cognition, Oxford, New York, 1996.
[11] J. Feldman, The simplicity principle in human concept learning, Curr. Dir. Psychol. Sci. 12 (2003) 227–232.
[12] C. Gonzalez, J.F. Lerch, C. Lebiere, Instance-based learning in dynamic decision making, Cognitive Sci. 27 (2003) 591–635.
[13] D.H. Krantz, H.C. Kunreuther, Goals and plans in decision making, Judgement Decision Making 2 (2007) 137–168.
[14] J.K. Kruschke, ALCOVE: an exemplar-based connectionist model of category learning, Psychol. Rev. 99 (1992) 22–44.
[15] J.K. Kruschke, Locally Bayesian learning with applications to retrospective revaluation and highlighting, Psychol. Rev. 113 (2006) 677–699.
[17] A.B. Markman, G.C. Baldwin, W.T. Maddox, The interaction of payoff structure and regulatory focus in classification, Psychol. Sci. 16 (2005) 852–855.
[18] T. Matsuka, Simple, individually unique, and context-dependent learning method for models of human category learning, Behav. Res. Methods 37 (2005) 240–255.
[19] T. Matsuka, Modeling human learning as context dependent knowledge utility optimization, in: Advances in Natural Computation, ICNC '05, Lecture Notes in Computer Science, vol. 3601, Springer, Berlin, 2005, pp. 933–946.
[20] T. Matsuka, A. Chouchourelou, On the learning algorithms of descriptive models of high-order human cognition, in: Advances in Neural Networks, ISNN '06, Lecture Notes in Computer Science, vol. 3971, Springer, Berlin, 2006, pp. 41–49.
[21] T. Matsuka, J.E. Corter, Observed attention allocation processes in category learning, Q. J. Exp. Psychol. 61 (2008) 1067–1097.
[22] D.L. Medin, M.M. Schaffer, Context theory of classification learning, Psychol. Rev. 85 (1978) 207–238.
[23] A. Newell, P.S. Rosenbloom, Mechanism of skill acquisition and the law of practice, in: J.R. Anderson (Ed.), Cognitive Skill and their Acquisition, Lawrence Erlbaum, Hillsdale, NJ, 1981.
[24] R.M. Nosofsky, M.A. Gluck, T.J. Palmeri, S.C. McKinley, P. Glauthier, Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961), Mem. Cog. 22 (1994) 352–369.
[25] F. Nottebohm, The neural basis of birdsong, PLoS Biol. 3 (2005) 759–761.
[26] B.P. Olveczky, A.S. Andalman, S.F. Michale, Vocal experimentation in the Juvenile songbird requires a Basal Ganglia circuit, PLoS Biol. 3 (2005) 902–909.
[27] M.G. Parckard, B.J. Knowlton, Learning and memory functions of the Basal Ganglia, Annu. Rev. Neurosci. 25 (2002) 563–593.
[28] D.C. Rubin, A.E. Wenzel, One hundred years of forgetting: a quantitative description of retention, Psychol. Rev. 103 (1996) 734–760.
[29] R.N. Shepard, C.L. Hovland, H.M. Jenkins, Learning and memorization of classifications, Psychol. Monogr. 75 (1961).
[30] S. Sikstrom, Forgetting curves: implications for connectionist models, Cognitive Psychol. 45 (2002) 95–152.
[32] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Academic Press, San Diego, 1999.

Toshihiko Matsuka is an Associate Professor at Department of Cognitive and Information Science and Graduate School of Advanced Integrated Science, Chiba University. Before joining Chiba University, he was with Stevens Institute of Technology (assistant research professor) and Rutgers University Mind and Brain Analysis (RUMBA) Laboratory (postdoctoral associate). His research interests include computational modeling of high-order human cognition (e.g., concept formation) and behavioral studies on human learning. He received his Ph.D. in Measurement and Evaluation (Psychometrics) from Columbia University.

Yasuaki Sakamoto is a Research Assistant Professor at the Howe School of Technology Management, Stevens Institute of Technology. He studies human and machine learning in the Center for Decision Technologies with primary emphases on developing computer programs that behave in a similar fashion as humans and devising methods for improving human performances. His research is based on human experimentation and computational modeling. He received his Ph.D. in Psychology from the University of Texas at Austin, where he was a member of the Computation and Cognitive Processes lab. His undergraduate work in Psychology was done at Indiana University at Bloomington, where he worked in the Percepts and Concepts lab.

Arieta Chouchourelou is an Assistant Professor at School of Humanities and Social Sciences, Cyprus College. Her research interests include perception of human body movements, cognitive neuroscience, and cognitive modeling. She holds a Ph.D. in Psychology from Rutgers University.

Jeffrey V. Nickerson is an Associate Professor in the Howe School of Technology Management at Stevens Institute of Technology. He is the Director of the Center for Decision Technologies. His research interests include Human–Robot Interaction, Social Network Analysis, and Sensor Network Design. He holds a Ph.D. in Computer Science from New York University.