Explaining representational shifts by selective attention, selective memorization, and random chance

Toshihiko Matsuka (matsukat@muscat.L.chiba-u.ac.jp)
Hidehito Honda (hito@muscat.L.chiba-u.ac.jp)
Sou Matsuura (matsuura@muscat.L.chiba-u.ac.jp)
Department of Cognitive and Information Science
Chiba University
Chiba, JAPAN

Abstract

Recent studies in category learning have shown that there are shifts in category representation. In the present study, we develop three models categorization that consisted of different learning objectives to examine cognitive mechanism underlying the representational shifts. The results of simulation indicated that the representational shift observed in Johansen & Palmeri (2002) can be explained by selective attention, selective exemplar memorization, or mere random chance. Although these models could not be differentiated based on classification generalization patterns, a detail examination of acquired model coefficients were conducted in order to design future studies.

Introduction

Recent studies in category learning and usage have shown that there are shifts in category representation. For example, Johansen and Palmeri (2002) showed that participants' category generalization patterns in early learning stages were more consistent with rule-like representations, but as learning progressed they exhibited generalization pattern that were more consistent with an exemplar-based representation. Johansen and Palmeri (2002) hypothesized that the selective attention process was the key cognitive mechanism that produced the representational shifts. They developed and successfully tested a model that initially paid attention to a single feature dimension and then gradually distributed its attention to achieve accurate classification. In contrast, Bourne and his colleague (Bourne, Healy, Kole & Graham, 2006; Bourne, Healy, Parker & Richard, 1999) suggested that representational shifts can occur both rule-to-exemplar and exemplar-to-rule fashion, depending on cognitive demands of tasks being performed. In particular, Bourne et al. (1999, 2006) claimed that memorabilities of strategies plays an important role in representational or strategy shifts.

In the present study, we developed three models of categorization that differed in their learning objectives. The first model was built on the basis of Johansen and Palmeri's (2002) idea that selective attention causes representational shifts. In particular, the model tries to acquire accurate categorization strategies while paying attention to a small number of dimensions. The second model was built on the basis of Bourne et al. (1999,2006) claim that the strategy memorability is the key process in representational shifts. This model tries to acquire accurate strategies while memorizing and using a smaller number of exemplars. The last model was somewhat different from previous two models in that it assumes that representational shifts occur by random chance. Given that this model was built on the basis of a stochastic optimization method, and it assumes that less-accurate simpler strategies (i.e., rule) are more likely to be realized in earlier stages of learning, while more complex strategies that bear higher classification accuracy (i.e., exemplar-like representation) are more likely to be maintained and applied in latter stages of learning.

Computational Models

Overview

In the present paper, we used ALCOVE (Kruschke, 1992) as the model of categorization, and CLEAR framework (Matsuka, Sakamoto, Chouchourelou & Nickerson, 2008) as the model of learning. CLEAR framework is a straightforward application of stochastic optimization method, namely Evolutional Strategy to human learning models. We refer CLEAR-augmented ALCOVE to as CaALCOVE. Three variants of CaAlCOVE were tested in the present study, namely standard CaALCOVE, attention-penalizing CaALCOVE, and (exemplar) memorization-penalizing CaALCOVE. The difference among those models is their learning objectives and we test how the learning objectives affect acquisition of different types of internal representations.

Categorization Algorithm - ALCOVE

In ALCOVE (Kruschke, 1992), 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 (i) is scaled by non-negative selective attention weights, a_i. 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 exemplar j and output node k:

$$O_k^{(n)}(x) = \sum_{j=1}^{J} w_{kj}^{(n)} \left[\exp \left(-\beta \cdot \sum_{i=1}^{I} a_i^{(n)} |\psi_{ji} - x_i| \right) \right] \quad (1)$$

where superscript n indicates n-th categorization strategy being utilized. Given that CaALCOVE's learning algorithm are based on stochastic optimization, dimensional attention weights takes the following form to obtain reasonable stability:

$$a_i^{(n)} = \left(1 + \exp\left(-D_i^{(n)}\right)\right)^{-1} \quad (2)$$

where D_i is a pseudo-attention weight. In CaALCOVE, Ds (not as) are updated in learning.

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:

$$P(C|x) = \frac{\exp\left(\phi \cdot O_c^{(\nu)}(x)\right)}{\sum_k \exp\left(\phi \cdot O_k^{(\nu)}(x)\right)}. \tag{3}$$

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

Although CaALCOVE would always have multiple strategies in mind, it opts for and applies a single strategy with the highest predicted utility, indicated by the superscript ν, to make one response at a time (e.g., categorizing an input instance).

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

Learning Algorithms - CLEAR

Overview of Learning Algorithm In CLEAR framework (see Matsuka, et al., 2008 for detailed discussion about its effectiveness and descriptive validity), Evolution Strategy (ES) method was used as learning processes. As in a typical ES application, we assumed three key processes in learning: *crossover*, *mutation*, and (survivor) *selection*. In the *crossover* process, randomly selected categorization strategies are combined to form a new strategy. In human cognition, the crossover process can be interpreted as conceptual combination in which new strategies are created based on merging ideas from existing useful strategies. In the *mutation* process, each model coefficient is randomly altered, which can be interpreted as strategy modification by randomly adjusting local attributes. In the *selection* process, a certain number of strategies are deterministically selected on the basis of their "usefulness." in relation to the situational characteristics. Those selected strategies will be kept in CaALCOVE's memory trace, while non-selected strategies become obsolete or are forgotten.

Unlike previous modeling approaches to category learning research, which modify a single strategy (i.e., a single set of coefficients), CaALCOVE maintains, modifies, and combines a set of strategies. The idea of having a population of strategies (as opposed to having an individual strategy) is important because it allows not only the selection and concept combination in learning, but also the creation of diverse strategies, making learning more robust. Thus, unlike previous models, CaALCOVE assumes that humans have the potential to maintain a range of strategies and are able to apply a strategy most suitable for a particular set of situational characteristics. In CaALCOVE framework, one individual could simultaneously have multiple representation schemes.

Although CaALCOVE always has multiple strategies in its knowledge space, it opts for and applies a single strategy 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 strategy is described in a later section.

Hypotheses Combinations In CaALCOVE, randomly selected pairs of strategies exchange information to create a new strategy. For the sake of simplicity, we use the following notation $\{\mathbf{w}^{(n)}, \mathbf{D}^{(n)}\} \in \boldsymbol{\theta}^{(n)}$. CaALCOVE utilizes discrete recombination of coefficients and intermediary recombination of the coefficient for self-adaptation. Thus,

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

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 strategies produced reaches the memory capacity of CaALCOVE.

Hypotheses Modifications After the recombination process, CaALCOVE randomly modifies its strategies, using a self-adapting strategy. Thus,

$$\sigma_{\theta_l}^{(n)}(t+1) = \sigma_{\theta_l}^{(n)}(t) \cdot \exp(N(0,\gamma)) \tag{5}$$

$$\theta_l^{(n)}(t+1) = \theta_l^{(n)}(t) + N(0, \sigma_{\theta_l}^{(n)}(t+1)) \tag{6}$$

where t indicates time, l indicates coefficients, γ defines search width (via σ'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 strategies, CaALCOVE selects a limited number of strategies to be maintained in its memory. In CaALCOVE, the survivor selection is done deterministically, selecting best 10 strategies on the basis of estimated utility of strategies or knowledge. The function defining utility of knowledge is described in the next section.

Knowledge Utility Estimation

The utility of each strategy or a set of coefficients determines the survivor selection process in CaALCOVE, which occurs twice. During categorization, it selects a single strategy with the highest predicted utility to make a categorization response (referred to as concept utility for response or UR hereafter). During learning, it selects best fit strategies to update its knowledge (utility for learning or UL hereafter). In both selection processes, the strategy utility is subjectively and contextually defined, and a general function is given as: $U(\boldsymbol{\theta}^n) = \Upsilon\left(E(\boldsymbol{\theta}^n), Q_1(\boldsymbol{\theta}^n), ..., Q_L(\boldsymbol{\theta}^n)\right)$ where Υ is a function that takes concept inaccuracy (i.e., *E*) and *L* contextual factors (i.e., *Q*) and returns an estimated strategy utility value (Note that learning is framed as a minimization problem).

In CaALCOVE, the predicted (in)accuracy of a strategy during categorization is estimated based on a retrospective verification function, which assumes that humans estimate the accuracies of the strategies by applying the current strategies to previously encountered instances retrieved from a memory trace . Thus,

$$E(\boldsymbol{\theta}^{(n)}) = \sum_{g=1}^{G} \sum_{k=1}^{K} \Xi \left(d_k^{(g)}, x^{(g)} \right) \left[d_k^{(g)} - O_k^{(n)} \left(x^{(g)} \right) \right]^2 \tag{7}$$

where superscript g indicates a particular input-output pair, G is the number of unique training pairs, and the exemplar retention function Ξ returns the retrieval strength g-th input-output pair. The last term is the sum of squared error with d being the desired output.

By assuming category structures being deterministic, the following exemplar retention function, based on Anderson and Schooler's learning-forgetting function (1991), is used in the present simulation study. Thus,

$$\Xi(d^{(g)}, x^{(g)}) = \frac{\sum\limits_{\forall i | x^{(i)} = x^{(g)}} (\tau^{(i)} + 1)^{-\delta}}{\sum\limits_{g} \sum\limits_{\forall i | x^{(i)} = x^{(g)}} (\tau^{(i)} + 1)^{-\delta}} \tag{8}$$

Memory decay parameter, δ, 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." Thus, τ = 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 strategies. $E(\boldsymbol{\theta})$ 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 (Anderson & Schooler, 1991; Newell & Rosenbloom, 1981).

Simulation

Three variants of CaAlCOVE were tested in the present study, namely standard CaALCOVE, attention-penalizing CaALCOVE, and (exemplar) memorization-penalizing CaALCOVE. The difference among those models is their learning objectives (i.e., knowledge utility functions).

Standard CaALCOVE

The learning objective function for the standard CaALCOVE was given as Eq 7. This model assumes that the representational shifts during category learning occur by mere random chance. That is, it assumes that simpler categorization strategies (i.e., rule-like representation) are more likely to be hypothesized and thus heavily utilized in earlier stages of learning. But as learning progresses, more complex and accurate strategies based on exemplar-like representation will be realized and tested, simply because creations of a larger number of hypotheses allow sufficient exploration of the solution space.

Attention-penalizing CaALCOVE

Attention-penalizing CaALCOVE (CAL-AP, hereafter) assumes that strong selective attention causes the representational shift. CAL-AP allocates most of its attention to a smaller number dimensions in earlier stages of learning, but it gradually allocates its attention to other dimensions to achieve more accurate categorization. That is, it penalizes distributed attention in earlier stages, but the penalization weakens as learning progresses. The underlying idea of CAL-AP is basically the same as the model proposed by Johansen & Parmeli (2002).

The knowledge utility function for CAL-AP is given as follows;:

$$U\left(\boldsymbol{\theta}^{(n)}\right) = E\left(\boldsymbol{\theta}^{(n)}\right) + \lambda_a \sum_i \frac{\left(a_i^{(n)}\right)^2}{\left(a_i^{(n)}\right)^2 + \sum\limits_{l=1}^{I} \left(a_l^{(n)}\right)^2} \tag{9}$$

This function encourages CAL-AP to pay attention to a smaller number of feature dimensions, or it penalizes CAL-AP when it selectively pays attentions to many dimensions. Note that the knowledge utility is estimated based on selective attention weight a, but not pseudo-selective attention weight D. λ_a is a scalar that balances the trade-off between categorization accuracy and the attention penalization. The value for λ_a decreases as learning progresses, like the annealing function used in Johansen & Palmeri (2002).

Memorization-penalizing CaALCOVE

Memorization-penalizing CaALCOVE (CAL-MP, hereafter) assumes that selective memorization and usage of particular exemplars causes the representational shift. That is, it assumes that a smaller numbers of exemplars are memorized and utilized in earlier stages of learning, causing CaALCOVE to exhibit categorization pattern that is consistent with a rule-like representation. As in CAL-AP, this model also weakens its penalization weight as learning progresses.

The knowledge utility function for CAL-AP is given as follows;

$$U\left(\boldsymbol{\theta}^{(n)}\right) = E\left(\boldsymbol{\theta}^{(n)}\right) + \lambda_w \sum_{kj} \frac{\left(w_{kj}^{(n)}\right)^2}{\left(w_{kj}^{(n)}\right)^2 + \sum\limits_{l=1}^{I} \left(w_{kj}^{(n)}\right)^2} \tag{10}$$

This function encourages CAL-MP to form a smaller set of active links (i.e., a link whose relative value is higher in its magnitude than other links) from exemplars and category nodes, or it penalizes CAL-MP when it associates categories with many exemplars (in terms of the relative values). In other words, Eq 10 promotes CAL-MP to maintain a smaller number of useful exemplars. Thus, when the memorization penalization weight (λ_w) is high, CAL-MP is more likely to acquire a rule-link representation. In contrast, if its penalization weight is small (e.g. knowledge accuracy outweighs selective memorization of exemplars) then it would acquire an exemplar-like representation. As in CAL-AP, the value for λ_w decreases as learning progresses.

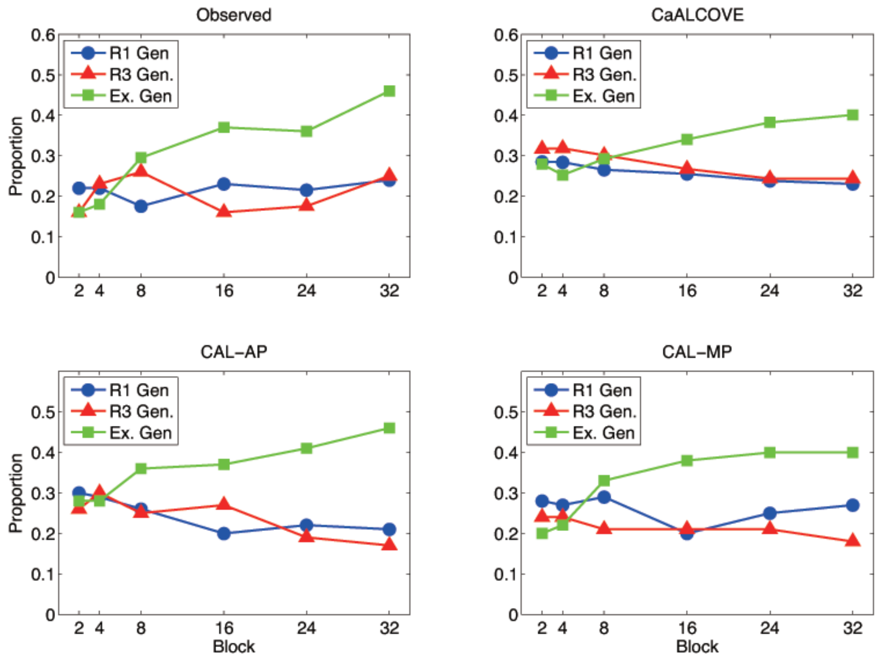

Figure 1: Results of Simulation. Generalizations profiles for Observed data (upper left), standard CaALCOVE (upper right), attention-penalizing CaALCOVE (lower left), and memorization-penalizing CaALCOVE (lower right). R1 Gen. indicates generalization pattern that is consistent with the categorization rule based on Dimension 1, and R3 Gen based on Dimension 3. Ex. Gen indicates generalization pattern that is consistent with exemplar usage.

Method

The basic simulation procedures followed those of the original study (Johansen & Palmeri, 2002). Table 1 shows schematic representation of stimulus set, which was adapted from Medin & Schaffer (1978). The models were run in a simulated training procedure with 32 trial blocks, where each block consisted of a random presentation of the nine unique training exemplars (see Table 1) exactly once, in order to learn the correct classification responses for the stimulus set. After, 2nd, 4th, 8th, 16th, 24th, and the last training blocks, the transfer tests were conducted using testing exemplars (i.e., T1 - T7). As in the original study, the responses to only T1, T2, T4, T5, and T6 were considered in the present simulation. The model parameters were arbitrary selected: For all three models β = 1.5, ϕ^1 = 1. For standard CaALCOVE, δ=1, γ=1, while those for other models were 0.5 and 0.5, respectively. Within each model, knowledge utility for learning (UL) and knowledge utility for response (UR) were assumed to be identical. There were a total of 100 simulated subjects for each model.

Results

Figure 1 shows observed and predicted generalization profiles. R1 Generalization (R1 Gen. in figure). indicates a generalization pattern that is consistent with the categorization rule based on Dimension 1 (i.e. responding AABBB to transfer stimuli T1, T2, T4, T5, and T6, see Johansen & Palmeri for detailed discussion about generalization patterns), and R3 Generalization is based on Dimension 3 (i.e. responding BBABA to the above mentioned stimuli). Exemplar Generalization (Ex. Gen.) indicates a generalization pattern that is consistent with exemplar usage (i.e., responding ABBBA to the stimuli). Note that as in the original simulation study, we calculated the proportion of R1 Generalization, for example, as the proportion of exact R1 Generalizations (AABBB) and those that differed by one response.

In general, all three models were successful in replicating the observed phenomena. All three models tended to exhibit the rule-based generalization patterns in earlier stages of learning and gradually shifted to the exemplar-based generalization pattern. The attention-penalizing CaALOVE (CAL-AP) seemed more successful than other models in showing increased exemplar usage. However, at the same time CAL-AP seemed least successful in not showing steady usages of the rules (i.e., usage of the rules did not declined in the observed data). It is rather intriguing that different learning objectives resulted in similar generalization patterns. The results of the present simulation suggest that representational shifts observed in Johansen & Palmeri (2002) may be explained by selective attention (i.e. CAL-AP), selective memorization of exemplars (CAL-MP), or mere random chance (CaALCOVE).

Discussion

Since the predicted generalization patterns alone could not differentiate the three models, we examined how each of the

[1] The value of ϕ does not affect model predictions

Table 1: Schematic representation of stimulus set used in simulations

Training					Transfer				
Cat	D1	D2	D3	D4	ID	D1	D2	D3	D4
A	1	1	1	0	T1	1	0	0	1
A	1	0	1	0	T2	1	1	1	1
A	1	0	1	1	T3	0	1	0	1
A	1	1	0	1	T4	0	0	1	1
A	0	1	1	1	T5	1	0	0	0
B	1	1	0	0	T6	0	0	1	0
B	0	1	1	0	T7	0	1	0	0
B	0	0	0	1					
B	0	0	0	0					

three models exhibited such generalization patterns by analyzing acquired selective attention and exemplar-to-category association weights.

Figure 2 shows acquired selective attention weights after 2nd (left column), 16th (middle column), and the last training block (right column). The top row shows acquires attention weights for standard CaALCOVE, the middle row for CAL-AP, and the bottom row for CAL-MP. Although the three models exhibited similar generalization patterns, the patterns of selective attention distributions differed somewhat greatly.

In early stages of learning, CAL-AP tended to allocate extreme attention weights to a smaller number of dimensions as its learning objective function suggested. CAL-MP, on the other hand, tended to evenly allocate its selective attention weights, while CaALCOVE exhibited intermediate behaviors. These tendencies generally hold throughout the learning processes.

A similar trend was obtained for association weights. CAL-MP tended to have a smaller number of active links between exemplars and categories, while CAL-AP tended to have a somewhat larger number of active links. CaALCOVE showed an intermediate pattern.

These analyses only confirm that the three models exhibited similar generalization patterns with different internal representation schema. We cannot infer which model(s) more accurately accounts for the representational shift, because there is no empirical data. However, these analyses are helpful in designing future empirical studies. For example, data on selective attention allocation pattern (e.g. Matsuka & Corter, 2008) would allow us to evaluate the three models examined in the present study, which in turn provides rich information for understanding cognitive mechanism underlying representational shifts.

Conclusions

Recent studies in category learning and usage have shown that there are representational shifts during category learning (Johansen & Palmeri, 2002, Bourne et al., 1999, 2006). In the present study, we develop three models categorization that consisted of three different learning objectives. The results of simulation study indicated that the representational shift observed in Johansen & Palmeri (2002) can be explained by selective attention, selective exemplar memorization, or mere random chance. Although three models exhibited very similar classification generalization patterns, their acquired internal representations (via different learning objective functions) were rather different. Since there is no empirical data to evaluate the models, we could not infer their descriptive validity. However, the results of simulation can be helpful in designing future empirical studies.

Acknowledgment

This work was in part supported by the Japan Society for the Promotion of Science KAKENHI (Grant No. 20700235) and the Support Center for Advanced Telecommunications Technology Research (SCAT).

References

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

Bourne, L.E., Jr., Healy, A., Kole, J. A. & Graham, S. M. (2006). Strategy shifts in classification skill acquisition: Does memory retrieval dominate rule use?. *Memory & Cognition, 34*, 903-913.

Bourne, L. E., Jr., Healy, A. F., Parker, J. T., & Rickard, T. C. (1999). The strategic basis of performance in binary classification tasks: Strategy choices and strategy transitions. *Journal of Memory & Language, 41*, 223-252.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15*, 55-71.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology, 45*, 482-553.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

Matsuka, T. & Corter, J. E. (2008). Observed attention allocation processes in category learning. *Quarterly Journal of Experimental Psychology, 61*, 1067-1097.

Matsuka, T., Sakamoto, Y., Chouchourelou, A., & Nickerson, J. V. (2008). Toward a descriptive cognitive model of human learning. *Neurocomputing, 71*, 2446-2455.

Medin, D. L., & Schaffer, M. M. (1978). Context Theory of Classification Learning. *Psychological Review 85*, 207-238.

Newell, A., & Rosenbloom, P. S., (1981). Mechanism of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skill and the their acquisition* (pp 1-55). Hillsdale, NJ: Lawrence Erlbaum Associates.

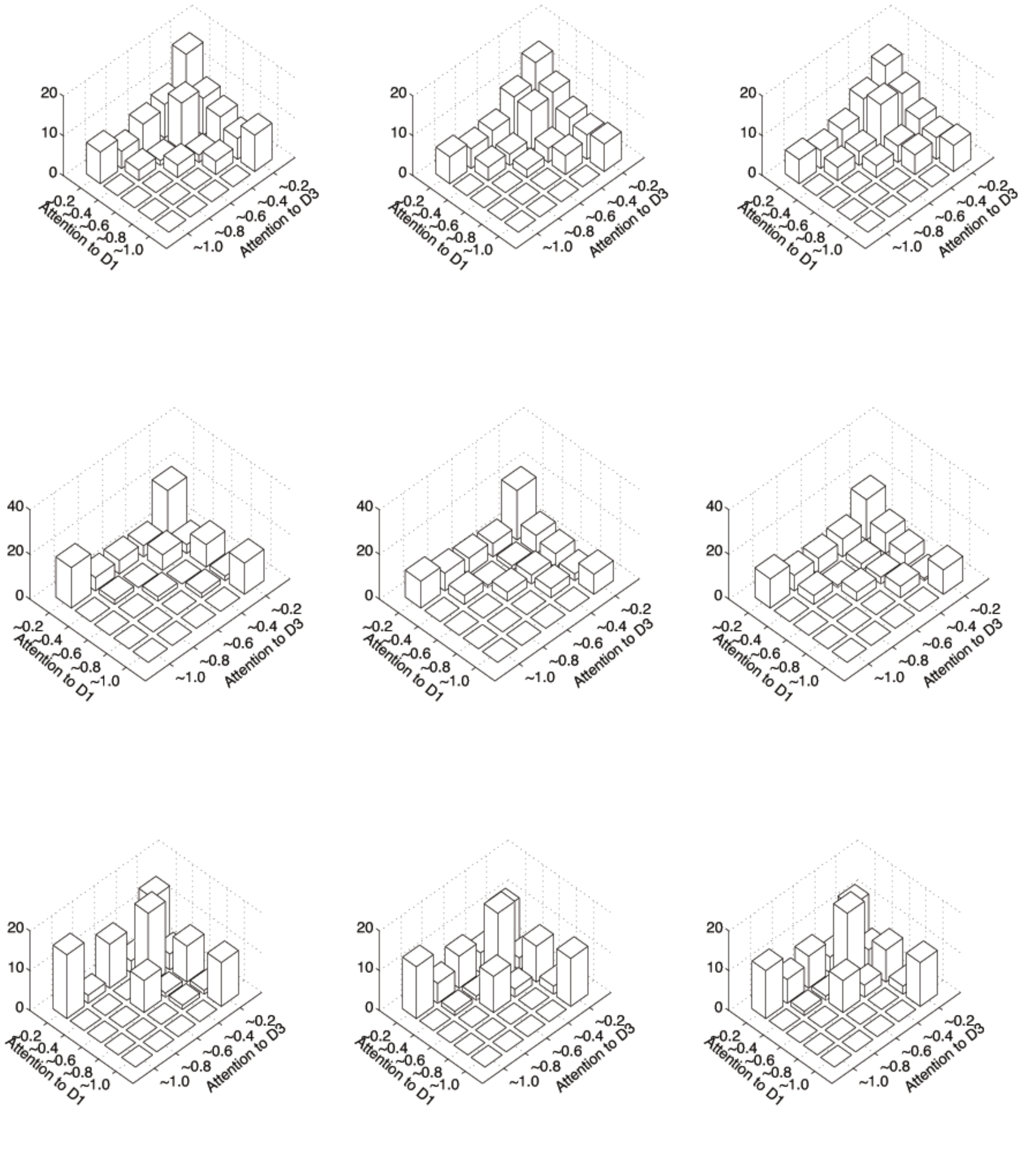

Figure 2: Learned relative selective attention weights for standard CaALCOVE (top row), attention-penalizing CaALCOVE (middle row), and memorization-penalizing CaALCOVE (bottom row). Left column shows learned attention weights at after 2nd training block, middle for 16th block, and right column for the last block.