Towards a Unified Model Describing Multiple Tasks: Extending the Retrieving Effectively from Memory Model to Categorization

Sinem Aytaç (saytac@syr.edu), **Yu-Wei Chang**, **Michael L. Kalish, & Daniel Corral**

Department of Psychology, Syracuse University
900 South Crouse Ave. Syracuse, NY 13244 USA

Abstract

This study extends the Retrieving Effectively from Memory model, a prominent computational model of episodic memory, to the domain of categorization. Our modeling approach begins with the assumption that same-category items share common features representing defining characteristics of their category, and that they are encoded in the same category list context. We then assumed that category judgments occur based on the comparison of an item's averaged similarity to the exemplars from each category. We use this model to explore how the learning modes of observation and classification might influence category learning and consider several strategies that may emerge during the classification mode. Model simulation results indicate that different strategies which people might adopt during classification can either confer an advantage or pose a disadvantage in category learning. These findings suggest potential avenues for future research, particularly in exploring diverse strategies employed during learning.

Keywords: categorization; observational learning; feedback; classification learning; retrieving effectively from memory

Introduction

Learning about categories is an integral part of our daily lives, often approached through two basic strategies: observation and classification. For instance, consider a child at the zoo. In one scenario, the child observes various animals while their parent provides names for each, representing observational learning. In another scenario, the child actively points out animals, names them, and receives feedback from a parent, which falls under classification learning. In both cases, the child learns to distinguish between furry mammals, slithering reptiles, and soaring birds.

Because of the importance of categorization for cognition, categorization is a highly theorized domain. One viable approach to understanding the fundamental processes involved in categorization is to leverage computational models designed to describe these processes, such as category learning and category judgment. To this end, numerous models have focused on categorization (e.g., Ashby & Townsend, 1986; Corral & Jones, 2012; 2014; Kruschke, 1992; Love et al., 2004; Minda & Smith, 2001; Nosofsky, 1986; Reed, 1972; Smith & Minda, 1998), with some drawing attention to the similarities between categorization and recognition (Nosofsky, 1988; 1991; Nosofsky & Zaki, 1998).

Previous reports have posited that categorization and recognition share a common representational system involving stored exemplars (Nosofsky, 1988; 1991; Nosofsky & Zaki, 1998). However, the decision rules governing categorization and recognition are presumed to differ slightly, such that categorization involves comparing the summed similarities of exemplars from different categories, while recognition decisions rely on familiarity derived from a process in which a probe is compared with all studied items, for instance, exemplars from all categories (Nosofsky, 1988; 1991; Nosofsky & Zaki, 1998). In summary, exemplar models ultimately accounted for both tasks when similarity comparisons to stored exemplars are assumed, albeit with different decision rules (Nosofsky, 1988; 1991; Nosofsky & Zaki, 1998; see also Love et al., 2004, for another example model that can account for both identification and categorization). Building upon this premise, our paper aims to extend a closely related process model in the recognition-memory literature, the Retrieving Effectively from Memory model (REM; Shiffrin & Steyvers, 1997), to explore its applicability in the context of categorization.

REM has demonstrated success and has been effectively extended to account for several phenomena (e.g., Aytaç et al., 2024; Criss et al., 2011; Criss & Shiffrin, 2005; Diller et al., 2001; Kılıç et al., 2017; Malmberg et al., 2004; Malmberg & Shiffrin, 2005; Schooler et al., 2001). REM's success comes from the assumptions including, but not limited to, the *global matching* and *differentiation* accounts. As a global matching model, REM's adaptation to categorization can eventually resemble exemplar models, proposing a category judgment for a test probe based on its similarity to stored exemplars of each category (e.g., Nosofsky, 1986).

Category learning, as illustrated in the opening example, is commonly studied using one of two tasks: classification, the most prevalent method, and observation, its alternative. The observational learning mode requires learning an item with an associated category label, which is comparable to the basic learning process in REM. In contrast, the classification learning mode inherently offers two study opportunities: 1) after the initial category judgment, and 2) after receiving feedback. The REM model appears to be an appropriate framework to represent the nature of this learning mode, as well, given its core assumption of differentiation. Differentiation would suggest that additional study leads to updating of memory traces, resulting in a more complete and accurate representation of items in memory (e.g., Aytaç et al., 2024; Kılıç et al., 2017; 2021; Shiffrin & Steyvers, 1997). Therefore, our paper will also explore the feasibility of REM's adaptation to the different learning modes, including observation and classification, by preserving its core assumptions.

Previous studies have revealed inconsistent findings regarding different types of learning modes (e.g., observation

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

versus classification) and their advantages. For example, while Ashby et al. (2002) demonstrated an advantage of classification learning over observational learning, other studies did not find a significant difference between the two in terms of category-judgment accuracy after single-item training (Corral & Carpenter, 2024; Levering & Kurtz, 2015; Patterson & Kurtz, 2020, Experiment 1). Furthermore, Patterson and Kurtz (2020) found an overall advantage of observational learning over classification learning when training included pair comparisons—pairs from the same category or different categories. Thus, which approach facilitates optimal learning is under debate in the categorization literature and awaits an answer, for which process models like REM can provide insights.

The aims of the present work are twofold: 1) to investigate the feasibility of extending the REM model to the domain of category learning, and 2) to apply this model to the learning modes of observation and categorization. Following a review of the REM model, we will propose its extension to a categorization task, with a specific emphasis on how category learning unfolds under observation versus classification. Subsequently, we will conduct several simulations and report the model's predictions, which will be further discussed in terms of their implications for optimal category learning.

Retrieving Effectively from Memory

REM describes episodic memory through memory traces embodying our life experiences. These traces are represented as vectors, with each element corresponding to a unique feature. Take, for instance, going to a zoo. In this case, each animal is stored as an individual vector, representing its characteristics (e.g., the parrot as a bird, with vibrant and colorful feathers, and its ability to mimic sounds). In conventional REM implementations, it is assumed that each vector consists of twenty feature values, and each feature value v is a positive integer randomly sampled from a geometric distribution with a parameter g:

$$P(v) = (1\text{-}g)^{v\text{-}1}\, g, \quad \text{with } v = 1, 2, \ldots, \infty \qquad \text{(Eq.1)}$$

When a memory trace is formed, each feature is assumed to be stored with the probability parameter u. Elements of the vector corresponding to non-stored features are assigned a value of zero. The accuracy of this storage is governed by the probability c. In cases where this accuracy is not maintained, which is anticipated to occur with the complementary probability of 1-c, a random value is stored instead (this random value is sampled from the same geometric distribution). Returning to our earlier illustration, the feature representing the class of the parrot may either not be stored in memory at all or be incorrectly stored as, for example, "mammal". Overall, this storage process is expected to create memory traces comprising correct, incorrect, and absent features, reflecting the inherent incompleteness and proneness to errors in memory.

In REM, the retrieval process relies on a global matching process, wherein a probe item is compared with the existing memory traces. This comparison process assesses the extent to which the features in the probe align with or differ from those present in each trace among a total of N traces. For a probe item j:

$$\lambda_{(i,j)} = (1\text{-c})^{nq_{(i,j)}} \prod_{v=1}^{\infty} \left[\frac{c+(1-c)g(1-g)^{v-1}}{g(1-g)^{v-1}} \right]^{nm(v,i,j)} \qquad \text{(Eq.2)}$$

with v indexing feature values in an ith episodic trace. The number of non-zero features that mismatch is nq, whereas the number of non-zero features that match is nm. Features that do not contain information (i.e., features with a value of zero) are not considered. That is, while incorrectly stored features decrease the similarity of the trace to its probe, absent features do not impact this similarity. These matches are averaged across traces, which yields an odds ratio Φ:

$$\Phi_j = \frac{1}{N} \sum_{i=1}^{N} \lambda_{(i,j)} \qquad \text{(Eq.3)}$$

This ratio captures the relative support that the probe item was previously encountered. When the odds ratio is higher than the decision criterion (typically set to 1), the item is endorsed as "old". Conversely, if the odds ratio is lower than the criterion, the probe item is not endorsed (i.e., it is judged to be "new").

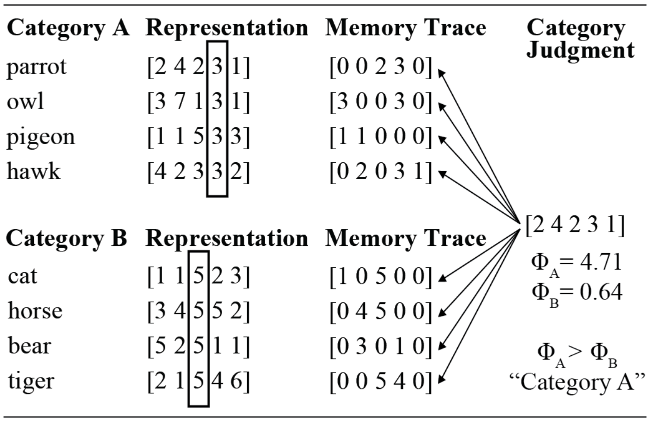
In the following section, we will describe the REM model extended to categorization; in other words, the model retrieving effectively from memory, categorization (REMC).

REMC – A Model of Categorization

For REM to describe categorization, we first need to establish how categories are represented and stored in memory. We begin by assuming that unique characteristics of categories are represented by shared features common to all items belonging to that category. Take, for example, parrots, owls, and pigeons, which all fall under the category of *birds* and share features such as wings and feathers. In the proposed model, each exemplar of the bird category (e.g., parrot, owl, pigeon) is represented as an individual vector that shares common categorical features (e.g., having wings and feathers) with all other exemplars. In this case, these common features are assigned the same integer value, standing in the same position for all exemplars (see Schooler et al., 2001, for the implementation of a similar idea to manipulate similarity between items). Figure 1 illustrates how the model represents items from the same category through a shared feature.

The model's operation during the study phase is partially dependent on the task at hand, such as whether it involves observational or classification learning. Therefore, we will first begin by describing the learning process during the observational learning task and then move on to the classification learning task.

In the case of *observational learning*, where the item's category label is provided during its presentation, a memory trace representing the presented item is stored in the provided category list. This storage process is imperfect and incomplete, as discussed in the earlier section, due to the probabilities of u and c that are applied. The task results in the creation of multiple lists (e.g., one list for each category), each consisting of item-memory traces of category-specific exemplars (see Figure 1). [1] It is important to note that in the observational learning task, items are prevented from being stored in incorrect category lists, marking a notable distinction from the classification learning task.



Category A	Representation	Memory Trace	Category Judgment
parrot	[2 4 2 3 1]	[0 0 2 3 0]	
owl	[3 7 1 3 1]	[3 0 0 3 0]	
pigeon	[1 1 5 3 3]	[1 1 0 0 0]	
hawk	[4 2 3 3 2]	[0 2 0 3 1]	[2 4 2 3 1]
Category B	**Representation**	**Memory Trace**	
cat	[1 1 5 2 3]	[1 0 5 0 0]	Φ_A= 4.71 Φ_B= 0.64
horse	[3 4 5 5 2]	[0 4 5 0 0]	
bear	[5 2 5 1 1]	[0 3 0 1 0]	$\Phi_A > \Phi_B$
tiger	[2 1 5 4 6]	[0 0 5 4 0]	“Category A”

Figure 1: The figure illustrates a toy example with two categories (e.g., birds and mammals) and their exemplars. Items in the same category list share a common feature representing the characteristics of that category. Studying through either categorization task results in the creation of multiple lists that consist of memory traces representing category-specific exemplars. During the test, a probe (in this example, an old item *parrot*) is compared to the memory traces from each category list separately. Whichever category list provides the highest similarity is then selected by the model.

We now turn our focus to *classification learning*, where participants are initially presented with an item and asked to make a category judgment during the study phase. After making a classification judgment, participants receive feedback indicating the item’s correct category before proceeding to the next item. In order to model the classification learning task, we assume that the initial decision about the item’s category in the study is made by comparing the item with the existing exemplars studied thus far from each category list. The model then returns the category with the highest similarity as an answer. However, what changes occur in memory after this decision might be modeled in a few different ways, which we will detail below.

One approach, and perhaps the most intuitive, to modeling the classification task is to have two storage attempts: a) after the initial category judgment and b) after the feedback. In this case, the item is first stored in the category list that corresponds to the classification judgment. Then, upon receiving feedback, if the category judgment was correct, the just-stored memory trace is updated. This updating occurs by replacing the trace’s empty features with integers through a probabilistic process, involving encoding parameters u and c (therefore, some features might remain unchanged). However, the eventual outcome is expected to be a more complete and accurate representation of the item in memory (also referred to as differentiation; see Aytaç et al., 2024; Kılıç et al., 2017; 2021; Shiffrin & Steyvers, 1997). In contrast, if the judgment was incorrect, the item is additionally stored in the correct category list that aligns with the corresponding feedback.

Critically, however, despite differentiation, this approach is expected to negatively impact performance on subsequent category judgments, as incorrect judgments lead to the erroneous storage of items in the wrong category list. We will further elaborate on this point later in the model simulation section.

An alternative approach involves modeling a single storage attempt, which occurs only *after* receiving feedback. In this case, the initial judgment about an item’s category may not influence memory at all or may impact the learning process solely after feedback has been provided. For instance, the learning strength of an item’s category could vary based on whether the judgment is consistent or inconsistent with subsequent feedback.

This approach has indeed been examined in a recent article by Kılıç et al. (2021), where the effects of learning during the test versus the study on item-recognition memory were simultaneously evaluated. Kılıç et al. (2021) found that including feedback during testing reduced interference from the tested items (i.e., reduced output interference) while increasing interference from studied items (i.e., increased list-length effect).

Interpreting these findings within the REM framework, Kılıç et al. (2021) suggested that feedback might decrease the tendency for erroneous modifications to memory, such as incorrect updating of memory traces after a false judgment. Their model-fitting results suggest that after feedback is provided, people store the item as a new memory trace, rather than making modifications to existing knowledge (i.e., memory traces) that are based on a false “old” judgment.[2] Therefore, this second approach, which we propose here to model a classification learning task, aligns with and extends recent findings.

We explore both of these approaches in the following sections, where we report model simulations and then predictions.

[1] The lists representing categories can alternatively be appended as vectors to item vectors representing their exemplars. Then, during category judgments, the model can go through a two-step process (e.g., REM.4; Shiffrin & Steyvers, 1997, pp. 155-156), where in the first step exemplars are activated based on the similarity of their category vectors to reinstated category information. However, for simplicity, here we assume that this activation is perfect, meaning that all studied exemplars contribute to the decision process.

[2] Note that Kılıç et al. (2021) do not rule out the possibility of erroneous modifications to memory after initial judgments about test items, prior to receiving corrective feedback. They acknowledge that incorrect updating might nevertheless occur with initial judgments, alongside the storage of new traces following feedback. They highlight the importance of further research to distinguish between these two possibilities.

Model Simulations

In this set of simulations, we first created two different category list items from two non-overlapping prototypes that contained twenty features each (l = 20). Items in the same category shared *one* common feature with each other, which represents the main characteristic of that category (l_s = 1). The shared common feature among items in the same category resulted in a 5% overlap among these items. The features representing the properties of items, as well as the characteristics of categories, were randomly sampled from a geometric distribution with the parameter g = 0.35, as described in Eq. 1.

Study Phase

We simulated the study phase, which mirrors the learning phase where participants study exemplars from each category through either mere observation or classification followed by feedback. To simulate the study phase, we included 50 items from each category. We set the storage parameter u to 0.2 and the storage-accuracy parameter c to 0.7 (see Table 1). In the case of the *observational learning* task, this simulation generated two category lists that contain incomplete and error-prone memory traces of their exemplars.

Table 1: Parameter values used in the model simulations.

Parameter	Value	Description
l	20	Number of features in a vector (i.e., vector length)
l_s	1	Number of shared features representing the characteristic(s) of a category
g	0.35	Feature frequency
c	0.7	Probability of correctly copying a feature
u	0.2	Probability of storing a feature
u_{cj}	0.1, 0.2	Probability of storing a feature after category judgment
u_{fb}	0.2, 0.4	Probability of storing a feature after feedback

Note. The parameters, including l, g, and c, are set to their conventional values, while others such as u are determined by the researchers based on prior modeling work.

Transitioning to the *classification learning* task, each item presentation initiated the category judgment process, as the task necessitates making a category judgment before revealing the item's category. Specifically, each item presentation commenced an evaluation of the item's category (i.e., whether it belongs to category A or B) by comparing its features with those of the items encountered thus far (see Eqs. 2–3).

Regarding the first approach discussed earlier, involving multiple storage attempts, we explored different strategies. First, we started by applying a constant probability value for storing a feature after the category judgment (u_{cj}= 0.2) and after feedback (u_{fb}= 0.2). Specifically, features of an item were stored with the probability u_{cj}= 0.2 (and correctly stored with the probability c = 0.7) in a memory list according to the first category judgment. Then, after receiving feedback, if the judgment was correct in the first place, the already-stored trace was updated. However, if the judgment was *not* correct, the trace was additionally stored in the correct category list. The latter process of updating/additional storage occurred based on the u_{fb}= 0.2 (and c = 0.7).

We next decided to set a lower storage parameter (u_{cj}= 0.1) for the initial category judgment in the study while keeping the u_{fb} parameter the same as before for the feedback portion of the task. This approach was expected to mirror the hesitancy of relying on one's own judgment and reduce the impact caused by false category judgments.

Additionally, to completely eliminate the effects of false category judgments in classification learning, we explored the model involving a single storage attempt, which occurs only after feedback has been provided (for a precursor implementation in item-recognition memory, see Kılıç et al., 2021). We once again started by setting the parameter u_{fb} to 0.2. This way, the model is expected to produce comparable results as in the observational learning task because this approach is identical to observational learning from the modeling perspective.

Finally, we decided to incorporate the model's category judgment into the learning process, which would impact learning positively when correct judgments are made. To do so, we set the parameter u_{fb} to 0.4 when feedback aligned with the preceding category judgment.

Test Phase

We conducted simulations of the test phase, which reflects the phase during which participants make category judgments on previously studied items as well as new items. In a comparable manner, our simulations included a test list comprising all previously studied items alongside an equal number of new items. Specifically, the test list contained 50 old items, allowing a direct evaluation of memory for those items, and 50 new items, facilitating an assessment of the generalization and transfer of that memory to novel cases. For each item presented in the test, the model compared the features in the probe with those in the item traces from each category *separately* (see Eq. 2). Subsequently, these matches were averaged across traces separately for each category, resulting in two distinct odds ratios (see Eq. 3). The odds ratios from each category were then compared, and the category with the highest odds ratio was returned as the answer by the model, indicating the item's category (see Figure 1).

Overall, we simulated the model five times, once for observational learning and once for each classification approach. We considered one thousand synthetic participants per simulation and reported the model's predictions averaged across said participants below.

Model Predictions

The model overall makes reasonable predictions in terms of accuracy for category judgments on both studied and new items, achieving accuracy levels above chance, with higher accuracy predicted for old items compared to new items (e.g., Patterson & Kurtz, 2020). The question that remains pertains to the learning modes and the differences between them, which we will now examine in detail.

In the case of observational learning, the model demonstrates favorable predictions, as shown in Figure 2A. The model's accuracy in category judgments notably exceeds chance for both types of items, with predictions of 86% accuracy for items studied during the observational learning task and 69% accuracy for new items that were not studied but that share a common feature with the corresponding category exemplars (i.e., transfer items).

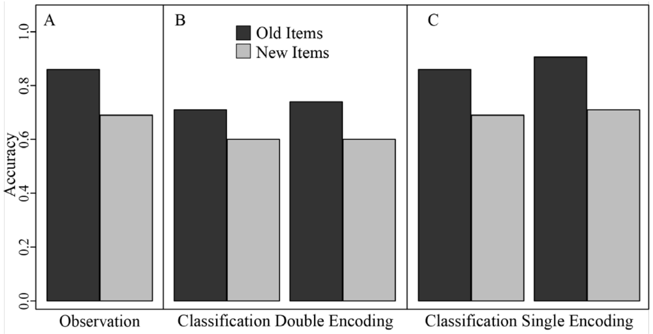


Figure 2: The figure illustrates the predictions of the proposed model (REMC) on category-judgment accuracy as a function of learning mode (observation vs. classification). Classification learning is explored in the model with two main approaches: a) one that leads to the encoding of items and their associated category after both category judgment and feedback (i.e., classification with double encoding), and b) another that leads to the encoding of items only after feedback (i.e., classification with single encoding). Dark gray bars represent the model's predictions regarding the accuracy for studied items while light gray bars represent these predictions for new items.

Looking at the model's predictions for the classification learning task involving multiple storage attempts with a constant probability, the model predicts above-chance level accuracy for category judgments (e.g., 71% accuracy for old items and 60% accuracy for new items; see the left-hand side of Figure 2B). However, as noted previously, the model expects lower performance on the test after learning via classification compared to observational learning because incorrect category judgments committed in the classification learning task result in erroneous encoding of these items in the wrong category lists.

Thus, when we adjusted the storage parameter for the initial judgment to a lower value, this change indeed increased the category-judgment accuracy for old items to 74% while not affecting the accuracy for new items (i.e., 60%; see the right-hand side of Figure 2B). Again, this improvement in accuracy occurs because incorrect category judgments during the study have a diminished impact with a lower storage parameter.

Moreover, after completely eliminating the adverse effects of false category judgments in classification learning through a single storage attempt post-feedback, the model yielded comparable results to those observed in the observational learning mode (see the left-hand side of Figure 2C and Figure 2A). Further incorporating the model's category judgment into the learning process produced 90% and 71% accuracy for old and new items, respectively (see the right-hand side of Figure 2C), surpassing the predicted performance in the observational learning mode. This prediction is important because it completely reverses the expected outcome for different learning modes.

In the next two sections, we will explore if these outcomes change when varying shared features within- and between-category items.

Varying Shared Features Within Category Items

For completeness, we also ran simulations with same-category items sharing more than one common feature. Specifically, within-category items overlapped by more than 5%. It is important to note that shared features manipulated here refer to the diagnostic features assumed to define a category (e.g., feathers and wings), not to irrelevant features that may randomly match across items (e.g., color).

As shown in Figure 3, the increase in within-category similarity improved the accuracy of categorization for both studied items and transfer items after learning with either task.

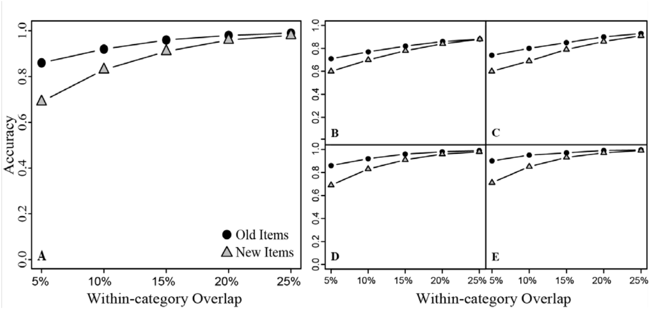


Figure 3: The figure illustrates the predictions of the proposed model on category-judgment accuracy as a function of the number of shared diagnostic features in items from the same category. Panel A represents the predictions for observational learning, Panels B and C correspond to classification learning with double encoding (u_{cj}= 0.2 and 0.1, respectively), and Panels D and E correspond to classification learning with single encoding (u_{fd}= 0.2 and 0.4, respectively).

Varying Shared Features Between Categories

Next, we varied the number of overlapping features in items from different categories. To be clear, these overlapping features are diagnostic for defining each category but are not sufficient to differentiate between the two presented categories. For example, being warm-blooded is a characteristic of birds that distinguishes them from other animal classes, such as reptiles, but this feature is not enough to differentiate birds from mammals, as mammals are also warm-blooded.

As shown in Figure 4, category-judgment accuracy is expected to decrease for both old and new items as the similarity between the categories increases. For the new items, specifically, accuracy drops to chance level, as there is no way to distinguish the two categories with 100% overlap in diagnostic features.

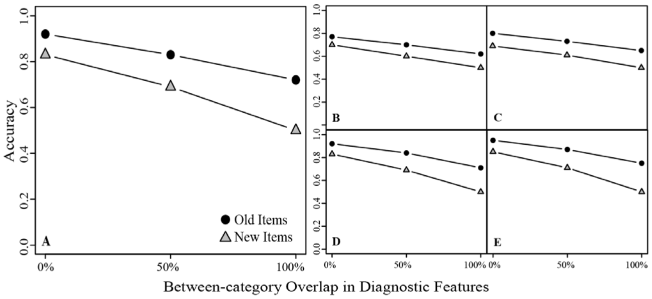


Figure 4: The figure illustrates the predictions of the proposed model on category-judgment accuracy as a function of the number of shared features in items from different categories. Panel A represents the predictions for observational learning, Panels B and C represent classification learning with double encoding (u_{cj}= 0.2 and 0.1, respectively), and Panels D and E represent classification learning with single encoding (u_{fd}= 0.2 and 0.4, respectively).

Regarding the different learning modes, the model predicts the sharpest decrease in accuracy for new items that follow the observational learning task and the classification task with a single encoding attempt. This prediction holds upon scrutinizing the decline in accuracy rates; that is, the observation and classification tasks with single encoding exhibit a slightly higher proportional deterioration in category judgment accuracy for new items as between-category similarity increases.

Discussion

The present work extends one of the most prominent computational accounts of human memory, the REM model (Shiffrin & Steyvers, 1997), to categorization. We began by assuming that items from the same category share common features that represent characteristics of the category and that they are encoded in the same list context representing categorical information. We then assumed that category judgments occur based on the comparison of an item's averaged similarity to the exemplars from each category. This proposed model accounts for both memory (e.g., item recognition performance) and category learning with the same system, which represents its major strength.

In this study, we additionally explored how different types of learning modes, namely observation and classification, might affect categorization and considered a few different strategies that may take place during classification learning. These different strategies for classification noticeably affected the model's predictions regarding category-judgment accuracy at test. For example, the model predicted a disadvantage of classification learning over observation learning when classification provides two encoding attempts, including after initial category judgment and after feedback. However, relying solely on feedback for learning in the former task produced comparable results to observational learning. Furthermore, the model predicted a benefit of the classification mode on category learning when correct category judgments in the study contribute to learning.

Classification leads to poorer category learning because of erroneous initial judgments, which results in the incorrect storage of exemplars in the wrong category. In cases like this, where there is a mismatch between expectation and feedback, observation seems to aid learning. Conversely, when the initial judgment is consistent with feedback, classification appears to be more beneficial for category learning. Overall, the effectiveness of these different modes may hinge on the individual's existing knowledge and beliefs.

While the model's predictions provide valuable insights into the inconsistencies observed in the literature (e.g., Ashby et al., 2002; Corral & Carpenter, 2024; Levering & Kurtz, 2015; Patterson & Kurtz, 2020), certain contrasting findings in the categorization literature may have also resulted from using different category manipulations, such as feature-based or relational categories (Ashby et al., 2002; Levering & Kurtz, 2015; Patterson & Kurtz, 2020) or the adoption of discriminative or generative approaches during learning (Hsu & Griffiths, 2010; Levering & Kurtz, 2015). Future modeling work can focus on adopting a discriminative approach during the classification learning mode by separately defining the encoding success of each feature—specifically assigning a unique storage probability for diagnostic versus non-diagnostic features.

Our simulation results suggest that people might adopt different strategies, such as relying on their own judgment versus partially or completely ignoring their own judgment and relying on feedback. These varying strategies are expected to yield contrasting findings regarding which learning mode benefits category learning more. More importantly, individuals may adopt these strategies based on their metacognitive awareness of their prior knowledge. Future work should aim to test these predictions that distinguish different strategies in classification learning and determine empirical conditions under which observation provides superior learning to classification, and vice versa.

References

Ashby F. G., Maddox, W. T., & Bohil C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory and Cognition, 30*(5), 666-677. https://doi.org/10.3758/BF03196423

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154-179. https://doi.org/10.1037/0033-295X.93.2.154

Aytaç, S., Kılıç, A., Criss, A. H., & Kellen, D. (2024). Retrieving effectively from source memory: Evidence for differentiation and local matching processes. *Cognitive Psychology, 149,* 101617. https://doi.org/10.1016/j.cogpsych.2023.101617

Corral, D., & Carpenter, S. K. (2024). *Acquiring complex concepts through testing and explanatory feedback.* Under revision.

Corral, D., & Jones, M. (2012). Learning of relational categories as a function of higher-order structure. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).

Corral, D., & Jones, M. (2014). The effects of relational structure on analogical learning. *Cognition*, *132*(3), 280-300. https://doi.org/10.1016/j.cognition.2014.04.007

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. J*ournal of Memory and Language, 64(*4), 316-326. https://doi.org/10.1016/j.jml.2011.02.003

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1199-1212. https://doi.org/10.1037/0278-7393.31.6.1199

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 414-435. https://doi.org/10.1037/0278-7393.27.2.414

Hsu, A., & Griffiths, T. (2010). Effects of generative and discriminative learning on use of category variability. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

Kılıç, A., Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2017). Models that allow us to perceive the world more accurately also allow us to remember past events more accurately via differentiation. *Cognitive Psychology, 92*, 65-86. https://doi.org/10.1016/j.cogpsych.2016.11.005

Kılıç, A., Fontaine, J., Malmberg, K. J., & Criss, A. H. (2021). The moderating role of feedback on forgetting in item recognition. *Computational Brain and Behavior, 4,* 178-190. https://doi.org/10.1007/s42113-020-00090-y

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22-44. https://doi.org/10.1037/0033-295X.99.1.22

Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory and Cognition, 43*(2), 266-282. https://doi.org/10.3758/s13421-014-0458-2

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*(2), 309-332. https://doi.org/10.1037/0033-295X.111.2.309

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 319-331. https://doi.org/10.1037/0278-7393.30.2.319

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 322-336. https://doi.org/10.1037/0278-7393.31.2.322

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775-799. https://doi.org/10.1037/0278-7393.27.3.775

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 700-708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 3-27. https://doi.org/10.1037/0096-1523.17.1.3

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*(4), 247-255. https://doi.org/10.1111/1467-9280.00051

Patterson, J. D., & Kurtz, K. J. (2020). Comparison-based learning of relational categories (you'll never guess). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 851-871. https://doi.org/10.1037/xlm0000758

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*(3), 382-407. https://doi.org/10.1016/0010-0285(72)90014-X

Schooler, L. J., Shiffrin, R. M., & Raaijmakers, J. G. (2001). A Bayesian model for implicit effects in perceptual identification. *Psychological Review*, *108*(1), 257-272. https://doi.org/10.1037/0033-295X.108.1.257

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM–retrieving effectively from memory. *Psychonomic Bulletin and Review, 4*(2), 145-166. https://doi.org/10.3758/BF03209391

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1411-1436. https://doi.org/10.1037/0278-7393.24.6.1411