The Effect for Category Learning on Recognition Memory: A Signal Detection Theory Analysis

Siyuan Yin[1,2,*], Kevin O'Neill[3,*], Timothy F. Brady[4], Felipe De Brigard[1,2,3,5]

[1]Duke Institute for Brain Sciences, Duke University, Durham, NC 27708, USA.
[2]Department of Philosophy, Duke University, Durham, NC 27708, USA.
[3]Center for Cognitive Neuroscience, Duke University, Durham, NC 27708, USA.
[4]Department of Psychology, University of California, San Diego, CA 92093, USA.
[5]Department of Psychology and Neuroscience, Duke University, Durham, NC 27708, USA.

Abstract

Previous studies have shown that category learning affects subsequent recognition memory. However, questions remain as to how category learning affects discriminability during recognition. In this three-stage study, we employed sets of simulated flowers with category- and non-category-inclusion features appearing with equal probabilities. In the learning stage, participants were asked to categorize flowers by identifying the category-inclusion feature. Next, in the studying stage, participants memorized a new set of flowers, a third of which belonged to the learned category. Finally, in the testing stage, participants received a recognition test with old and new flowers, some from the learned category, some from a not-learned category, some from both categories, and some from neither category. We applied hierarchical Bayesian signal detection theory models to recognition performance and found that prior category learning affected both discriminability as well as criterion bias. That is, people that learned the category well, exhibited improved discriminability and a shifted bias toward flowers from the learned relative to the not learned category.

Keywords: category learning; recognition memory; signal detection theory; Bayesian modeling

Introduction

Memory research has shown that prior learning experience affects recognition memory. It is often thought that prior learning is encoded into knowledge structures or *schemas* (Bartlett, 1932). In turn, schemas increase recognition of schema-inconsistent information compared to schema-consistent information, while also increasing false alarms to schema-consistent lures compared to schema-inconsistent lures. Because schema acquisition takes time and learning experiences vary among people, most recognition memory tasks have employed either within-subject designs for pre-acquired schemas (Graesser & Nakamura, 1982) or between-subject designs for individuals with different expertise (Castel et al, 2007). As such, traditional experimental designs do not easily allow manipulation of schema acquisition in a way that enables us to assess their effect on recognition memory performance.

A number of recent studies have unveiled strong connections between schematic and categorical knowledge, leading many researchers to postulate profound similarities in the cognitive processes underlying schematic and categorical learning (Sakamoto & Love, 2004). To contribute to the integration of schematic and categorical learning, and to further explore the effects of prior learning on recognition memory, De Brigard et al. (2017) recently employed a set of computer-generated stimuli (flowers) to explore how learning a novel category affects participants' recognition memory for items from the learned category relative to items from a category they did not learn. However, the studies reported by De Brigard et al. (2017) left several unanswered questions. In particular, the findings could not differentiate between discriminability changes for items from the learned category and a change in response bias because their experiments did not include foils of both learned and not-learned categories, and thus could not provide measures of discriminability and bias for all options. In addition, De Brigard et al.'s (2017) findings did not discriminate between those who learned best and those who learned least during the category-learning phase, potentially obscuring effects on discriminability in recognition memory.

To explore these issues, in the present study we used a modified version of De Brigard et al.'s (2017) paradigm in which flowers from learned and non-learned categories appeared in the learning and study phases with equal probability. Additionally, the current study included lures from both learned and not-learned categories during the recognition test. As such, we were able to implement full hierarchical Bayesian signal detection theory (SDT) models to data from all participants, as well as separate people by the strength of their learning. This modified experimental paradigm, and the SDT models with which the results are analyzed, enables us to further understand the effect of category learning on recognition memory.

Category Learning and Recognition Experiment

Participants

113 individuals participated via Amazon Mechanical Turk (https://www.mturk.com) for monetary compensation. All participants were from the United States and had at least 100 approved hits and overall hit rate $\geq$ 95%. Three participants were excluded because of failure to follow instructions or terminated the experiment in the middle, so data were analyzed with the remaining 110 individuals. All participants

were provided informed consent under a protocol approved by the Duke University IRB.

Materials

Stimulus consisted of MATLAB (2018b)-generated flowers from De Brigard et al. (2017). Flowers varied across five dimensions, with each dimension taking one of three possible values: number of petals (4, 6, or 8), color of petals (blue, green, or yellow), shape of center (circle, triangle, and square), color of center (orange, purple, or turquoise), and number of sepals (1, 2, or 3). Figure 1 illustrates three examples of flowers with different combinations of the features (see further details in De Brigard et al., 2017).

Procedure

We closely followed the procedure from the fourth experiment in De Brigard et al. (2017), with some modifications (see below). The experiment had three phases: learning, study, and test. At the beginning of each phase participants read the instructions for 90s.

Figure 1. Examples of MATALB-generated flowers. From left to right: 4 blue petals- orange circle center -1 sepal; 6 green petals- purple triangle center-2 sepals; and 8 yellow petals- blue square center-3 sepals. See more in De Brigard et al., 2017.

In the learning phase, participants were told they would see a flower on the screen and will have to determine whether or not it belonged to the species *avlonia*. Participants were told that avlonias differed from other flowers in one simple way (e.g., only avlonias have four petals), and their task was to find out what the simple way was. At the beginning of the learning phase, participants were informed of all five possible dimensions—number of petals, color of petals, etc.—across which flowers may vary and saw two example flowers for illustration. They then made binary choices "yes" or "no" on each trial to categorize each flower by pressing "y" or "n", respectively, and there were 54 trials in total. Immediately after their responses, feedback with the word "Correct" or "Incorrect" was displayed. Participants were ensured that they could guess at the beginning but eventually they would find out the simple way that made a flower an avlonia. Each participant was assigned to a category-inclusion feature consisting of one possible value from one of the five dimensions. Additionally, participants were also assigned a "Not-learned" category, defined by a value of a different dimension, of which participants were never informed or given feedback. Both of these assignments were counterbalanced across participants. In all phases of the experiment, all values of all stimulus features did not differ in their statistical proporties, such that flowers having the learned feature (i.e., that were avlonias) appeared on one-third of the trials, while the other two-thirds of the trials included flowers displaying the other two values of the Learned category-inclusion feature. Likewise, one-third of the trials presented flowers in the Not-learned category, while the other two-thirds of the trials included flowers with the other two values of the Not-learned category-inclusion feature. Importantly, the category-inclusion features for the Learned and Not-learned categories were independent, such that one-ninth of all flowers were in both the Learned and the Not-learned categories (Both condition), two-ninths of all flowers were in the Learned category but not the Not-Learned category (Learned condition), two-ninths of all flowers were in the Not-learned category but not the Learned category (Not-learned condition), and four-ninths of all flowers were in neither the Learned nor the Not-learned category (Neither condition). Table 1 summarizes the distribution of values for the Learned and Not-Learned category-inclusion features.

In the study phase, participants were asked to memorize 18 flowers. Each flower was shown alone for 5s followed by a 1s blank. Of the 18 flowers, four were in the Learned category but not the Not-Learned category (Learned), four were in the Not-learned category but not the Learned category (Not-Learned), two were in both categories (Both), and eight were members of neither category (Neither). To incentivize memorization, participants were told that they would receive an extra bonus for remembering above 85% of the stimuli. None of these 18 flowers were presented during the learning phase (Table 1).

Finally, in the testing phase, participants were told that they would see 54 flowers, one on each trial, and that their task was to remember whether or not stimuli were shown before in the study phase by pressing "yes" or "no". Of the 54 flowers, 18 were *old*—i.e. were presented in the studying phase—while the remaining 36 were *new*. Of these new flowers, four were from the Learned category only, four were from the Not-learned category only, two were from Both, and eight were from Neither. Of note, these new flowers were not shown during the study phase. All flowers were presented randomly and each trial was self-paced.

In sum, there were four types of trials in these three phases. Table 1 illustrates some possible combinations of Learned and Not-learned features. For each subject, one-third of trials included the learned category inclusion feature, which was chosen randomly from the three possible values from one of the five dimensions. Orthogonally, one-third of the trials included a not-learned category inclusion feature, i.e., the value of a dimension that could define a category of which participants were not aware of. This not-learned category inclusion feature was chosen randomly from the values belonging to the remaining four dimensions different from the dimension with the learned category inclusion feature. Membership in the Learned and Not-learned categories was independent of one another.

Table 1: Examples of possible combinations of Learned and Not-learned feature. Each row indicates one possible combination for one participant. A_1, A_2 and A_3 indicate three

possible values (denoted by 1, 2, and 3) of one randomly selected dimension out of five dimensions (denoted by A, B, C, D, and E; here we use only A and B for illustration purpose) -- number of petals, color of petals, shape of center, color of center, and number of sepals. B_1, B_2 and B_3 indicate three possible values of another randomly selected dimension out of the remaining four dimensions. Both condition has learned category inclusion feature and not-learned category inclusion feature features, and Neither condition does not have learned category inclusion feature or not-learned category inclusion feature features. The number of trials shown in the table is for learning and testing phases only. The number of trials for each feature during the study phase is 2 (not shown).

Learned feature	Not-learned feature	Number of trials	Probabilities
A_1	B_1	6	1/9
A_1	B_2	6	1/9
A_1	B_3	6	1/9
A_2	B_1	6	1/9
A_2	B_2	6	1/9
A_2	B_3	6	1/9
A_3	B_1	6	1/9
A_3	B_2	6	1/9
A_3	B_3	6	1/9

Results

Learning. We measured the learning performance by calculating the percentage of correct responses in the learning phase (Figure 2). We found participants were, in general, able to detect the single feature that categorized avlonias. The overall accuracy rates for both stimuli during the last twenty trials were 82.3%. Note that because we do not inform participants of the feature in advance, they necessarily begin at 50% accuracy at the beginning of the learning phase.

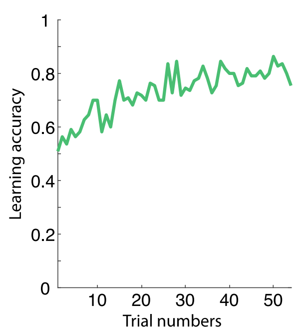

Figure 2. Learning performance during learning phase.

Memory Accuracy. We analyzed hit and false alarm (FA) rates separately for flowers of each type (Figure 3). To examine the learning effects for the four conditions (i.e., flowers that belong to the Learned category, Not-learned category, Both categories, Neither category), we implemented a two-way Bayesian repeated measures ANOVA. People exhibited increased hit rates for stimuli containing learned features included in Learned (M_{Hit} = 0.65, SD_{Hit} = 0.28) and Both (M_{Hit} = 0.70, SD_{Hit} = 0.35) conditions during the testing phase, but not toward stimuli not including those features in Not-learned (M_{Hit} = 0.56, SD_{Hit} = 0.28) and Neither (M_{Hit} = 0.58, SD_{Hit} = 0.22) conditions (Figure 2 and Table 2). We followed up with Bayesian paired samples t-tests which showed evidence supporting that hit rates in the Learned condition were higher than those in the Not-learned (BF_{10} = 2.15) and Neither conditions (BF_{10} = 1.70), but not in the Both condition (BF_{10} = 0.25) (See the scale of evidence in Jeffreys, 1998). Similarly, there was evidence indicating that hit rates for the Both condition were higher than those in the Not-learned (BF_{10} = 135.38) and Neither condition (BF_{10} = 30.56). Hit rates in the Not-learned condition were not different from the Neither condition (BF_{10} = 0.14). Also, there was weak evidence for FA rates in the Learned condition (M_{FA} = 0.59, SD_{FA} = 0.24) being higher than for the Not-learnsed (M_{FA} = 0.53, SD_{FA} = 0.23; BF_{10} = 0.43) and Neither condition (M_{FA} = 0.53, SD_{FA} = 0.19; BF_{10} = 0.61). We found no evidence for differences in other pairs of conditions (Both condition: M_{FA} = 0.57, SD_{FA} = 0.30).

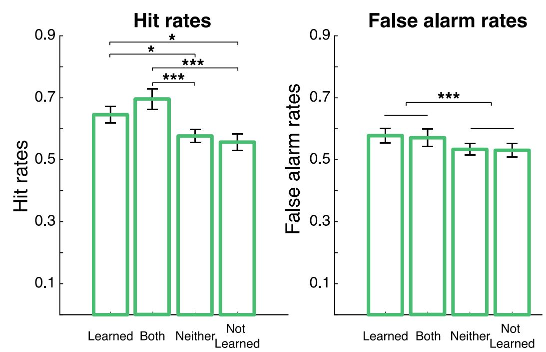

Figure 3. Hit and false alarm rates during testing phase. Left panel: hit rates. Right panel: false alarm rates. Stimulus with the category-inclusion value appeared in the Learned and Both conditions, and not in the Neither and Not-Learned conditions. * $BF_{10} > 1$, *** $BF_{10} > 10$.

Table 2: Bayesian repeated measures ANOVA

Rates	Best models	BF_{Model}	BF_{10}
Hit	Learned	11.02	6.93×10^3
False Alarm	Learned	13.93	32.35

BF: Bayes Factor.

We followed up with Bayesian paired samples t-tests which showed evidence supporting that hit rates in the Learned condition were higher than those in the Not-learned (BF_{10} = 2.15) and Neither conditions (BF_{10} = 1.70), but not in the Both condition (BF_{10} = 0.25) (See the scale of evidence in Jeffreys, 1998). Similarly, there was evidence indicating that hit rates for the Both condition were higher than those in the Not-learned (BF_{10} = 135.38) and Neither condition (BF_{10} = 30.56). Hit rates in the Not-learned condition were not different from the Neither condition (BF_{10} = 0.14). Also, there was weak evidence for FA rates in the Learned condition (M_{FA} = 0.59, SD_{FA} = 0.24) being higher than for the

Not-learned (M_{FA} = 0.53, SD_{FA} = 0.23; BF_{10} = 0.43) and Neither condition (M_{FA} = 0.53, SD_{FA} = 0.19; BF_{10} = 0.61). We found no evidence for differences in other pairs of conditions (Both condition: M_{FA} = 0.57, SD_{FA} = 0.30).

To explore the effect of category learning separately on response bias and discriminability (e.g., d'), we conducted a hierarchical Bayesian parameter estimation analysis within a SDT framework. To that end, we fit the accuracy data from three groups, i.e., (1) *all* participants (n = 110), (2) *experts*, i.e., participants whose accuracy of the last twenty learning trials was greater than or equal to 80% (n = 66), and (3) *non-experts*, i.e., participants whose accuracy of the last twenty learning trials was less than 80% (n = 44), to a SDT model in which the parameters were estimated using a hierarchical Bayesian approach (Lee, 2008). As such, two parameters of discriminability were estimated: (1) the sensitivity, d', that is measured by the distance between the signal and noise distributions indicating the discriminability of the signal trials from the noise trials; and (2) the criterion or bias, c, that is measured by the distance between the actual criterion used for responding and the unbiased criterion (i.e., $d'/2$).

The hierarchical model of SDT is able to partially pool individual parameters by taking into account group-level distributions, thus yielding more reliable estimates than non-hierarchical, full individual difference models. In this model, individual parameters are drawn from group-level (normal) distributions with estimated means and standard deviations. The model assumes that the estimated means quantify discriminability and criterion-bias for each of the four conditions, and precision quantifies the similarity among individual behavior.

In this implementation, our SDT model has four parameters per condition, reflecting properties of the average subject and how the subjects vary: mean discriminability μ_d, precision of discriminability τ_d, mean criterion μ_c, and precision of criterion τ_c. The prior on the mean discriminability was set to be very wide so as to be uninformative over the range of reasonable d' values (i.e., 0-4), with only a slight pull toward 0, consistent with previous research. Specifically, individual d_i was drawn from a normal distribution with mean and precision $\mu^d \sim N(0, 0.001)$ and $\tau^d \sim Gamma(0.001, 0.001)$, respectively. Individual c_i was then drawn from the normal distribution with two group-level parameters $\mu^c \sim N(0, 0.001)$ and $\tau^c \sim Gamma(0.001, 0.001)$. We implemented the hierarchical SDT model in JAGS, a sampler that utilizes a version of the BUGS programming language (Version 3.3.0) called from MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States). Posterior distributions were approximated by 3 Monte Carlo Markov Chain methods with 5000 samples from each chain, after a burn-in of 1000 samples. Convergence of chains was evaluated with the $\hat{R}$ statistic.

We first estimated the mean sensitivity and mean criterion-bias for each condition by calculating the posterior distributions of hit and FA rates for all participants--group (1). We found that in the Learned condition, this was skewed toward 1 for both hit and false alarm rates, significantly above the other three conditions (Figure 4A), indicating the people had both more hits and more false alarms in this condition. Furthermore, for participants from group (2, expert-learners), hit and FA rates in both Learned and Both conditions were skewed toward 1, significantly above than those under Not-learned and Neither conditions (Figure 4B), whereas for participants from group (3, non-expert-learners) there were no differences (Figure 4C), suggesting the main effect was driven by the expert-learners.

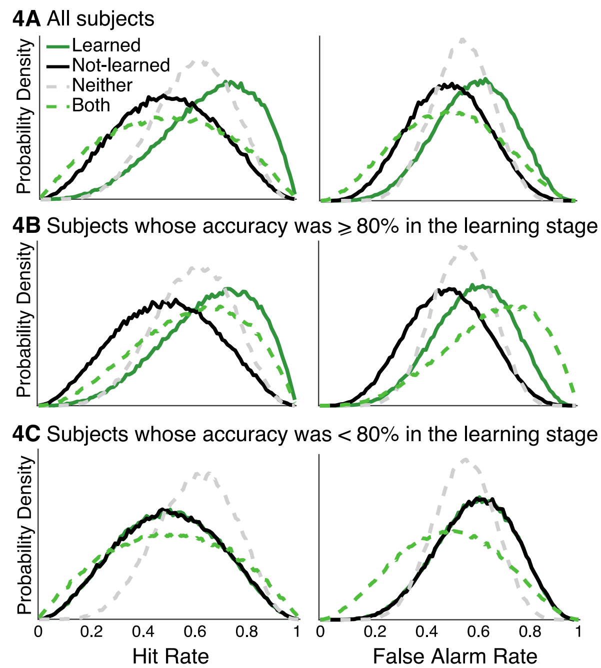

Figure 4. Posterior distribution of hit and FA rates for each of the four conditions.

To further investigate differences in discriminability/bias we performed a two-way Bayesian repeated measures ANOVA on estimated individual sensitivity and criterion-bias measures from each group. For individual sensitivity/d' of all subjects (1), we found main effects for Learned and Not-learned categories as well as their interaction, while for criterion-biases, we only found a main effect for the Learned category. For group (2, expert-learners), we found a significant main effect for the Learned category and a significant interaction between the Learned and Not-learned categories for both sensitivity and criterion-bias measures. For individual sensitivity of group (3, non-expert-learners), we found main effects for Learned and Not-learned categories as well as their interaction, while for criterion-biases we did not observe main effects of categories or their interaction (Table 3). These results indicate that participants who clearly excelled at learning the category during the learning stage—which we here operationalize as those participants whose accuracy for the last twenty trials was above 80%—were more sensitive to (i.e., increased discriminability/d') other flowers in this category and also and tended to say 'old' more often for these in general (i.e., Learned and Both conditions) compared to flowers not in the category (i.e., Not-learned and Neither conditions).

Follow-up Bayesian paired sample t-tests on sensitivity and criterion-bias for participants from group (1)—i.e. all participants—showed decisive evidence supporting that the sensitivity d_i for the Learned condition ($d_{Learned} = 0.18 \pm 0.05$) was higher than for the other three conditions ($d_{Not-learned} = 0.07 \pm 0.11$, $d_{Neither} = 0.12 \pm 0.03$, $d_{Both} = 0.38 \pm 0.18$), while the sensitivity d_i for Both was higher than the Not-learned and Neither conditions. As for the criterion-bias c_i , the evidence was also decisive supporting that the bias c_i for the Learned condition ($c_{Learned} = -0.32 \pm 0.38$) was lower than for the other three conditions ($c_{Not-learned} = -0.12 \pm 0.23$, $c_{Neither} = -0.15 \pm 0.34$, $c_{Both} = -0.39 \pm 0.28$), while the criterion-bias c_i for Both was lower than the Not-learned and Neither conditions. No strong evidence supported any differences between Not-learned and Neither conditions for both sensitivity d_i and bias c_i. (Table 4A).

For participants from group 2, this analysis revealed strong evidence that support differences in sensitivity d_i ($d_{Learned} = 0.290 \pm 0.048$, $d_{Not-learned} = 0.119 \pm 0.107$, $d_{Neither} = 0.126 \pm 0.084$, $d_{Both} = 0.339 \pm 0.083$) in almost all pairwise contrasts except Not-learned versus Neither. The same trend was also found in bias c_i ($c_{Learned} = -0.360 \pm 0.359$, $c_{Not-learned} = -0.073 \pm 0.262$, $c_{Neither} = -0.161 \pm 0.365$, and $c_{Both} = -0.543 \pm 0.191$). These results suggest that participants who mastered the learned features well in the learning stage were overall more sensitive to flowers with those features.

For participants from group 3, the sensitivity d_i of Learned condition ($d_{Learned} = 0.069 \pm 0.077$) was higher than those of Not-learned condition ($d_{Unlearned} = 0.003 \pm 0.080$) and lower than those of Both condition ($d_{Both} = 0.391 \pm 0.413$), but not different from those of Neither condition ($d_{Neither} = 0.096 \pm 0.019$). The sensitivity d_i of Both condition were higher than those of the other conditions, and the d_i of Not-learned condition were lower than those of Neither condition. As for biases c_i, the Bayesian paired t test did not show strong evidence supporting any differences between pairs of conditions ($c_{Learned} = -0.264 \pm 0.415$, $c_{Unlearned} = -0.183 \pm 0.155$, $c_{Neither} = -0.135 \pm 0.294$, and $c_{Both} = -0.132 \pm 0.176$), except moderate evidence suggesting differences in c_i between Learned and Neither conditions as well as Learned and Both conditions.

Figure 5 illustrates the joint posterior distributions of discriminability and bias for each condition. The main panel shows 15000 samples form the joint posterior of the mean μ^d and μ^c . The side panels show the marginal distribution for each of the group-level means. For all subjects, the group-level sensitivity d_i differed the most between Both and Not-learned conditions, and the group-level biases c_i were negative in Both and Learned conditions. That is, participants exhibited better sensitivity toward flowers with learned features and a tendency to overrespond "yes" in the recognition memory tasks (Figure 5A).

Table 3: Bayesian repeated measures ANOVA

Datasets	SDT parameters	Best models	BF_{Model}	BF_{10}
All subjects	d_i	Learned+Not-learned+Learned×Not-learned	5.39×10^{26}	1.43×10^{73}
	c_i	Learned	12.04	4.17×10^{13}
Experts	d_i	Learned+Not-learned+Learned× Not-learned	25.97	1.27×10^{51}
	c_i	Learned+Not-learned+Learned× Not-learned	884.46	1.27×10^{19}
Non-Experts	d_i	Learned+Not-learned+Learned× Not-learned	1.40×10^{8}	1.19×10^{13}
	c_i	Null model	4.26	1.00

d_i and c_i are individual sensitivity and biases estimated by the hierarchical Bayesian parameter estimation.

Table 4A: Bayesian paired sample t test for sensitivity and bias with all subjects. Numbers shown in the table indicate Bayes Factors.

Category comparison	Sensitivity d_i	Bias c_i
Learned vs. Not-learned	4.19×10^{12}	6.04×10^{4}
Learned vs. Neither	1.73×10^{23}	9.03×10^{2}
Learned vs. Both	8.77×10^{16}	0.773
Not-learned vs. Neither	3.24×10^{2}	0.183
Not-learned vs. Both	1.53×10^{27}	3.87×10^{10}
Neither vs. Both	8.72×10^{25}	2.03×10^{6}

Table 4B: Bayesian paired sample t test for sensitivity and bias with only subjects whose accuracy of the last twenty learning trials was above or equal to 80% (i.e., experts). Numbers shown in the table indicate Bayes Factors.

Category comparison	Sensitivity d_i	Bias c_i
Learned vs. Not-learned	2.37×10^{14}	1.75×10^{5}
Learned vs. Neither	7.29×10^{22}	1.48×10^{2}
Learned vs. Both	3.06×10^{2}	1.04×10^{3}
Not-learned vs. Neither	0.154	0.810
Not-learned vs. Both	8.56×10^{19}	3.25×10^{18}
Neither vs. Both	3.48×10^{19}	3.85×10^{8}

Table 4C: Bayesian paired sample t test for sensitivity and bias with only subjects whose accuracy of the last twenty learning trials was less than 80% (i.e., non-experts). Numbers shown in the table indicate Bayes Factors.

Category comparison	Sensitivity d_i	Bias c_i
Learned vs. Not-learned	1.15×10^2	0.411
Learned vs. Neither	1.35	1.63
Learned vs. Both	2.64×10^3	1.24
Not-learned vs. Neither	4.26×10^6	0.30
Not-learned vs. Both	4.92×10^4	0.41
Neither vs. Both	9.94×10^2	0.16

For subjects whose accuracy in the last twenty learning trials was greater than or equal to 80%, the difference in the group-level sensitivity d_i between Learned and Both conditions was less but the difference between Learned and Not-learned or Neither were greater. The group-level biases c_i in Learned and Both conditions were more negative than those in Not-learned and Neither conditions (Figure 5B). The results suggested that participants who learned category-relevant features well had better discriminability and stronger biases toward flowers with learned features. For subjects whose accuracy in the last twenty learning trials was less than 80%, the group-level sensitivity d_i differed the most between Both and Not-learned conditions, whereas the group-level biases c_i became closer to each other across conditions (Figure 5C). The results indicated that participants who did not learn the category-relevant features well had worse discriminability and little biases toward flowers with learned features.

Discussion

In this study we measured the extent to which learning novel categories influences recognition memory, and we focused on sensitivity and biases estimated in Bayesian SDT modeling. First, we corroborated previous findings that people exhibited biases toward stimuli within a learned category compared to stimuli not in the category, even when the relevant features are equally sampled during learning and study (De Brigard et al., 2017). That is, hit rates of stimuli with learned features (i.e., Learned and Both trials) were higher than stimuli with other values for that feature (i.e., Not-learned and Neither trials) (Figure 2). False alarm rates showed the same pattern. Going beyond this, we first fit full Bayesian SDT models and compared two measures of discriminability—sensitivity and criterion-bias—in four conditions. We observed that experts exhibited greater sensitivity and more negative criterion-bias than non-experts. We found greater discriminability for Learned and Both conditions than Not-learned and Neither conditions, which suggested people formed better memories of studied flowers with learned features. It is also clear that there was a response bias for Learned and Both conditions (Figure 4), indicating a tendency to overrespond "yes" (i.e., the flower was shown in the study stage) for these, in addition to the actual improved memory sensitivity. These results suggest that category learning affected recognition memory, improving discriminability as well as affecting response bias.

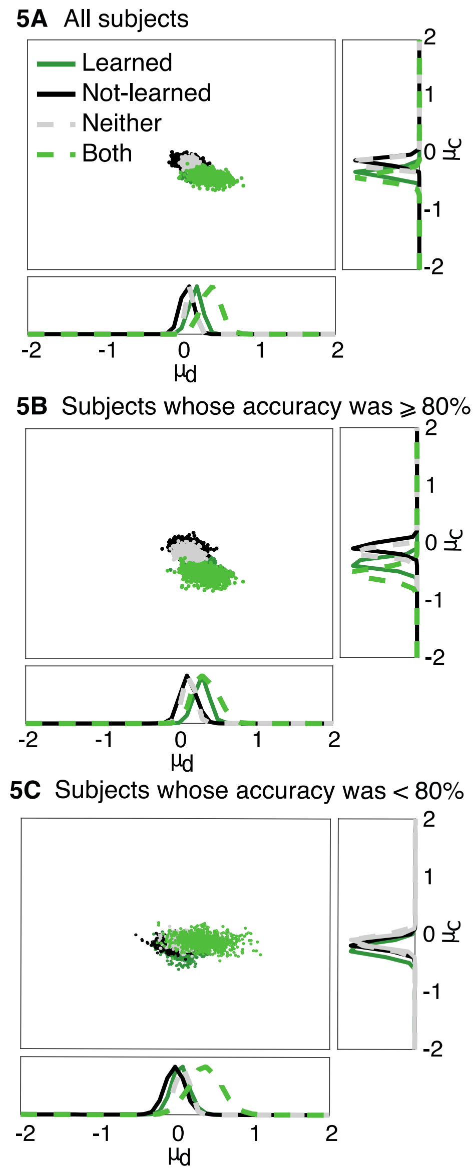

Figure 5: the joint distribution of mean discriminability d and mean bias c. The side panels show the corresponding marginal distribution. μ_d and μ_c are the group-level means of discriminability and criterion.

In the current study we employed a yes/no learning strategy to create new categories for novel stimuli and ask how they influence subsequent recognition memory. Our findings of the influence of category information on recognition memory are consistent with findings that show the influence of existing categories (Bae et al., 2015; Persaud & Hemmer, 2016) as well as newly learned episodic information about a category (Brady et al., 2018) on continuous recall measures. This suggests more insight into the influence of newly learned categories on memory looking at the effect of novel category learning on recognition memory employing continuous measures. In addition, future studies may investigate whether different learning strategies may elicit the same biases. For example, supervised (i.e., with explicit guidance on category-inclusion criteria) and unsupervised (i.e., without explicit guidance), or active (i.e., trying to learn category-inclusion criteria with instant feedback) and passive (i.e., merely observing stimuli and their corresponding categories) learning processes may largely change biases toward stimuli with learned features.

It is worth noting that, in the current study, we used a somewhat arbitrary threshold to classify expert and non-expert learners. Future studies may apply Bayesian analyses to explore individual differences in learning and compare estimates of individual learning rates to individual recognition memory effects. This could provide a better characterization of the data rather than a binary division.

Previous studies have mainly focused on category learning and memory during the course of an experiment, but how these categories are acquired is also critical in this processing. In this study, we used a set of well controlled stimuli – computer simulated flowers – so that we can manipulate the degree of exposure of different features and reveal how the learning process affects recognition memory. In future work, it would be useful to adopt more naturalistic stimuli to examine the mechanisms of category learning in real world settings and how this varies as a function of context and with different age groups.

We applied SDT models to measure the effect of category learning on recognition memory. This effect may also be related to different learning procedures: for example, explicit reasoning and the nature and timing of feedback, which may or may not be directly associated with the learned feature only (Ashby & Maddox, 2005). Other categorization models such as the generalized context model (GCM, Nosofsky, 1986), the general recognition theory (GRT, Ashby & Townsend, 1986), or the deterministic exemplar model (DEM, Ashby & Maddox, 1993) will be worth exploring to make more refined quantitative accounts of the influence of category learning on recognition memory.

Finally, it is important to note that in the current study, participants were given binary choices in the testing stage (old/new). While this allowed us to apply signal detection models to probe the effect of category learning on recognition memory, to do so we needed to assume an equal variance signal detection model. Adopting a confidence scale and ROC analysis based on confidence rating data would provide a refined gauge of discriminability in the recognition memory task and allow us to measure the memory signal accurately, even in the case of unequal variance (as is common in recognition memory experiments). This would allow us to be more certain we had separately measured response bias and discriminability and address the nature of the memory signal more clearly (e.g., address whether unequal variance signal detection model, or a hybrid threshold and signal detection model is more applicable; Wixted, 2007). Broadly, however, our results show that participants discriminate more toward stimuli with learned features than those with not-learned features. These results contribute to our understanding of how prior category learning influences recognition memory.

Acknowledgments

This study is supported by a grant from the Office of Naval Research (N00014-17-1-2603) to FDB.

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of Perceptual Independence. *Psych. Rev., 93(2)*, 154-179.

Ashby, F. G., & Maddox, W. T. (1993). Relations between Prototype, Examplar, and Decision Bound Models of Categorization. *J. Math. Psych., 37*, 372-400.

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Ann. Rev. Psych., 56*, 149-178.

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. Journal of Experimental Psychology: General, 144(4), 744-763.

Brady, T. F., Schacter, D.L., and Alvarez, G.A. (2018). The adaptive nature of false memories is revealed by gist-based distortion of true memories. PsyArXiv Preprint. https://doi.org/10.31234/osf.io/zeg95

Castel, A. D., McCabe, D. P., Roediger, H. L., III, & Heitman, J. L. (2007). The dark side of expertise: Domain specific memory errors. *Psych. Sci.*, 18, 3–5.

De Brigard, F., Brady, T. F., Ruzic, L., & Schacter, D. L. (2017). Tracking the emergence of memories: A category-learning paradigm to explore schema-driven recognition. *Mem Cogn*, *45*, 105-120.

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 59–109). New York: Academic Press.

Jeffreys, H. (1998). The Theory of Probability (3rd ed.). Oxford, England.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psych. Bull. Rev., 15(1),* 1-15.

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *J. Exp. Psych.: General, 115(1),* 39-57.

Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology, 88*, 1-21.

Sakamoto, Y., & Love, B. C. (2010). Learning and retention through predictive inference and classification. *J. Exp. Psych.: Applied*, 16, 361–377.

Wixted, J.T.. (2007) Dual-Process Theory and Signal-Detection Theory of Recognition Memory. *Psych. Rev., 114 (1)*, 152-76.

Yonelinas, A., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S. & King L. (1996) Signal-Detection, Threshold, and Dual-Process Models of Recognition Memory: ROCs and Conscious Recollection. *Cons. Cog., 5*, 418-441.