Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

Research

A role for the medial temporal lobes in category learning

Yi-Wen Wang and F. Gregory Ashby

University of California, Santa Barbara, California 93106, USA

Despite much research, the role of the medial temporal lobes (MTL) in category learning is unclear. Two unstructured categorization experiments explored conditions that might recruit MTL category learning and memory systems—namely, whether the stimulus display includes one or two stimuli, and whether category membership depends on configural properties of the stimulus features. The results supported three conclusions. First, in agreement with prior research, learning with single stimulus displays depended on striatal-mediated procedural learning. Second, and most important, learning with pair displays was mediated by MTL declarative memory systems. Third, the use of stimuli in which category membership depends on configural properties of the stimulus features made MTL learning slightly more likely. Overall, the results suggested that the MTL are most likely to mediate learning when the participant must decide which of two configural stimuli belongs to a selected category.

A variety of different neural networks support human learning, which allows us to acquire a rich repertoire of skills and knowledge to survive and prosper. Declarative memory is recruited when learning facts, and procedural memory is recruited when learning new motor skills. Much previous research has focused on identifying specialized brain circuits for specific types of learning. In the case of category learning, the influential work of Knowlton et al. (1996) triggered an avalanche of evidence pointing toward a critical role for the striatum (for reviews, see e.g., Ashby et al. 1998; Ashby and Ennis 2006; Seger and Miller 2010). For example, virtually all studies that have tested patients with striatal dysfunction have reported category-learning deficitis.

In contrast, it has been much more difficult to understand the role of the medial temporal lobes (MTL) in category learning. Some studies have reported impaired category-learning performance in patients with amnesia (Knowlton et al. 1994; Kolodny 1994; Zaki et al. 2003; Hopkins et al. 2004; Barense et al. 2005; Graham et al. 2006; Swainson et al. 2006), but many others have reported intact performance (e.g., Leng and Parkin 1988; Janowsky et al. 1989; Knowlton and Squire 1993; Kolodny 1994; Squire and Knowlton 1995; Filoteo et al. 2001; Bayley et al. 2005). Furthermore, even in tasks that initially seem ideally suited to MTL-mediated learning, such as unstructured categorization, the evidence favors the striatum over the MTL. In unstructured categorization tasks, the stimuli are visually distinct and assigned to each contrasting category randomly, and thus there is no rule- or similarity-based strategy for determining category membership.[1] The only successful strategy in unstructured categorization tasks seems to be some form of paired-associate learning. Even so, when the stimuli are presented one-at-a-time and learning depends on trial-by-trial feedback, evidence suggests that learning is procedural and more basal-ganglia dependent than MTL dependent (Seger and Cincotta 2005; Seger et al. 2010; Lopez-Paniagua and Seger 2011; Crossley et al. 2012).[2] One obvious hypothesis is that the conscious recollection of specific stimuli used in the experiment is MTL dependent, but the trial-and-error learning of arbitrary stimulus-response associations depends more on the basal ganglia.

On the other hand, there are at least four different experimental features that have been reported to recruit MTL learning and memory systems. One is to create categories that include distinct exceptions. In particular, in tasks in which the categories have some coherent structure, but in which they also include a small number of distinct exceptions, evidence suggests that the learning of the exceptions recruits MTL systems (e.g., Davis et al. 2012). A second design feature that seems to make MTL learning and memory more likely is to use observational training, rather than more traditional feedback-based training. For example, Shohamy et al. (2004) reported that patients with Parkinson's disease (PD), who have striatal dysfunction, are impaired in feedback-based learning, but not in observational learning. Two other features that seem to recruit MTL systems are the focus of this article—namely, whether the stimulus display includes one or two stimuli, and whether category membership depends on configural properties of the stimulus features. Our results suggest that the MTL are most likely to

[1]By rule, we mean an instruction that can be described by a Boolean expression that produces a response to every stimulus (e.g., "if the stimulus is large respond A; otherwise respond B").

Corresponding author: fgashby@ucsb.edu

Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.051995.120.

[2]On the other hand, Stark et al. (2018) reported fMRI evidence that pointed to a greater role for the hippocampus than the striatum in unstructured category learning. The strongest evidence for hippocampal involvement came in a condition in which six highly similar stimuli were divided among four categories—so at least two categories contained only a single exemplar. This study differed from traditional unstructured category-learning experiments in at least three ways: the stimuli were more similar than is typical; there were four categories, rather than the more typical two; and some categories included only one exemplar. More research is needed to disambiguate the contributions of these three factors.

© 2020 Wang and Ashby This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first 12 months after the full-issue publication date (see http://learnmem.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

mediate learning when the participant must decide which of two configural stimuli belongs to a selected category.

The nature of the stimulus display

The studies that have most consistently reported learning deficits in MTL-lesioned patients used some form of discrimination task, rather than traditional categorization (e.g., Barense et al. 2005; Graham et al. 2006; Swainson et al. 2006). For example, Barense et al. (2005) displayed two stimuli on each trial—one from each of the contrasting categories—and the participant's task was to identify the stimulus from the "Correct" category. MTL-lesioned patients were impaired. In contrast, the traditional categorization task displays one stimulus on each trial and the participant's task is to assign it to its correct category by pressing the appropriate response key.

Why should it matter whether the stimulus display includes one or two stimuli? One reason is because evidence suggests that striatal-mediated learning is of stimulus-response associations (Ashby et al. 2003; Featherstone and McDonald 2004; Maddox et al. 2004b; Spiering and Ashby 2008; Horvitz 2009; Crossley et al. 2012; Helie and Ashby 2012). In typical category-learning experiments, a single stimulus is presented and the participant's task is to assign it to its correct category via a button press—for example, by pressing a response key labeled "A" or "B." Under these conditions, stimuli in the A category are always associated with an A response, and stimuli in the B category are always associated with the B response. But consider a pair-display task where the display on each trial includes one exemplar from category A and one from B, and the participant's task is to indicate which one is from category A. Now the response keys are labeled "left" and "right" and stimuli from the A category could elicit either response. So learning stimulus-response associations is insufficient for this task. And in fact, Spiering and Ashby (2008) reported evidence that striatal-mediated learning fails under similar conditions. For these reasons, two-stimulus displays might be more likely to recruit MTL systems than single-stimulus displays.

The nature of the stimulus features

Although the MTL might not be ideally suited for reinforcement learning, they have their own unique properties. For example, the MTL are critical for binding or conjoining separate representations in a way that preserves their configural relationship (e.g., Rudy and Sutherland 1989; Sutherland and Rudy 1989; Aggleton et al. 2007; Duncan et al. 2018). In most category-learning experiments, the category exemplars are simple visual objects and categorization depends on feature identity, but not on the configural relationship among the features. Thus, tasks in which category membership depends on configural properties of the stimulus features might also favor the MTL over the striatum.

Sanderson et al. (2006) examined both of these properties in a water-maze study with rats in which the categories identified the location of a submerged platform. In their single-display condition, only one stimulus was displayed on each trial, with category A indicating the platform was to the left and category B to the right. In their pair-display condition, two stimuli were present on every trial, with a category A stimulus above the platform and a category B stimulus above open water.[3] In all conditions, category membership depended on the configural relationship of the stimulus features.

Sanderson et al. (2006) reported that in the pair-display condition, the learning of MTL-lesioned rats was significantly impaired relative to a group that received sham lesions. In contrast, in the single-display condition, the MTL-lesioned and sham-lesioned groups showed no difference in learning. Their inference was that learning in the pair-display condition required processing of the configural relationship of the figure components, which is MTL-dependent. Although they did not perform any basal ganglia lesions, we suspect that learning in the single-display condition was basal-ganglia dependent, given the abundant evidence of basal ganglia involvement in category learning.

The current experiments

Based on these considerations, our Experiment 1 crossed two types of stimulus displays (single vs. pair) with two types of stimuli (configural vs. nonconfigural). The stimuli are shown in Figure 1. The configural stimuli were patterned after the stimuli used by Sanderson et al. (2006). Note that each of these stimuli includes two separate components, and that its mirror image belongs to the contrasting category. Therefore, category membership is not determined by stimulus features, but by their configural relationship. The nonconfigural stimuli are similar to stimuli used in the unstructured categorization experiments of Seger and Cincotta (2005) and Crossley et al. (2012).

Note that each category includes six stimuli, and that there is no logical rule that determines category membership, nor is any similarity-based strategy helpful. Thus, these category are unstructured. Although previous research suggests that feedback-based, unstructured category learning is mediated by procedural learning within the basal ganglia (Seger and Cincotta 2005; Seger et al. 2010; Lopez-Paniagua and Seger 2011; Crossley et al. 2012), it is important to note that all of these previous studies used single stimulus displays and nonconfigural stimuli.

The two stimulus displays are illustrated in Figure 2. The Single conditions replicated the standard category learning display in which a single stimulus is presented and the participant's task is to assign it to category A or B by pressing the appropriate response key. In the Pair conditions, two stimuli were displayed—one from category A and one from category B and the participant's task was to indicate whether the category A stimulus was on the left or right by pressing one key associated with a left response or another key associated with a right response. Note that there is no consistent stimulus-response mapping in the Pair conditions because half

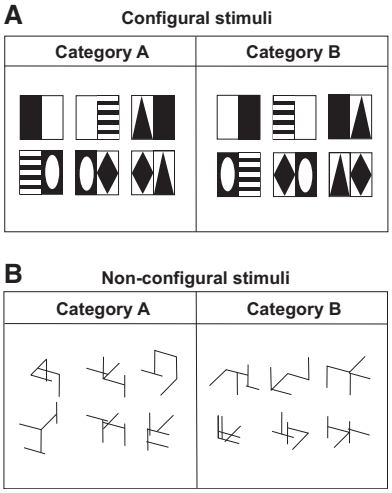

Figure 1. Stimuli used in Experiments 1 and 2.

[3]Sanderson et al. (2006) called the pair-display condition the structural-discrimination condition and the single-display condition the transverse-patterning condition.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

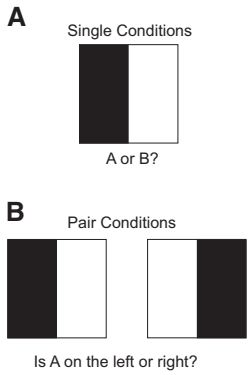

Figure 2. The two different types of stimulus displays used in Experiments 1 and 2.

the time participants pressed one key for category A and half the time they pressed the other key.

Our experiments are purely behavioral, so we have no direct method for assessing MTL involvement. Even so, we have several indirect methods that have been used with success in previous research. First, much previous research suggests that striatal-mediated procedural learning is impaired if the trial-by-trial feedback is delayed for 2.5 sec or more, whereas declarative-memory mediated category learning is unaffected by feedback delays as long as 10 sec (Maddox et al. 2003; Maddox and Ing 2005; Dunn et al. 2012; Crossley and Ashby 2015). Furthermore, within the striatum, the hypothesized cellular basis of the impaired learning caused by feedback delays has been identified (Yagishita et al. 2014). Therefore, Experiment 2 added feedback-delay conditions as a method for identifying striatal-mediated learning.

Feedback delays impair striatal learning more than MTL learning. Ideally, we would also like a method that impairs MTL learning more than striatal learning. However, while some fMRI studies have reported correlated BOLD patterns in MTL during categorization tasks (for review, see Zeithamova et al. 2019), we know of no noncontroversial method for disrupting MTL processing in healthy humans. Instead, we pursued a different goal: disrupt the processing of information that is presumably encoded by MTL. Since there is much evidence that interactions between prefrontal cortex (PFC) and the MTL are critical for visuospatial processing (Petrides 1985; Swainson et al. 2006; Barker and Warburton 2013; Renzi et al. 2013), Experiment 2 used a method known to interfere with PFC-mediated learning—namely, requiring participants to switch quickly to a new task after the categorization feedback is presented. Maddox et al. (2004a) reported that such switching interferes with declarative-memory mediated category learning much more than with procedural category learning—presumably because feedback processing requires attention and effort in declarative systems, but not in procedural systems.

Results

Experiment 1

This experiment used a 2 × 2 factorial design that crossed two levels of stimulus type (Configural vs. Nonconfigural) with two levels of stimulus display (Single vs. Pair). The stimuli are shown in Figure 1 and the display conditions are illustrated in Figure 2. Each condition included several blocks of learning, and a final transfer-test block at the end in which the opposite display type from training was used and no feedback was provided. Participants were excused if they responded correctly on all 48 trials in a block, or if 45 min had elapsed. As a result, some participants completed more blocks than other participants.

Figure 3 shows the accuracy in each condition for all blocks that were completed by at least 10 participants. Note that performance in the Pair condition is consistently better than in the Single condition for both types of stimuli, although this advantage appears to be smaller for the Nonconfigural stimuli than for the Configural stimuli after the fifth block.

To test these conclusions more rigorously, we performed a number of different statistical analyses. Because the dependent variable is dichotomous (i.e., correct vs. incorrect response), our main analysis used logistic regression. First, we tested the full model in which accuracy is a function of trial, display type (Single or Pair), and stimulus type (Configural or Nonconfigural). The results are shown in Table 1.

Note that these results generally reinforced the informal conclusions drawn from a visual inspection of Figure 3. Specifically, the effect of display type was significant [$\chi^2(1) = 9.514$, $P = 0.002$], suggesting that performance was better with pair displays than single displays. A follow-up analysis showed that after block 5, this difference was only significant in the Configural conditions [Configural conditions: $\chi^2(1) = 4.452$, $P = 0.035$; Nonconfigural conditions: $\chi^2(1) = 0.138$, $P = 0.711$]. A reanalysis with all the data (i.e., including blocks with small sample sizes) reached this same conclusion.

Figure 4 shows accuracy during the final transfer-test block in each of the four conditions for all perfect learners (i.e., all participants who achieved 10% correct during their last training block). Recall that the transfer block display was of the opposite type from training. So during transfer, participants in the Pair conditions were tested on Single displays, whereas participants in the Single conditions were tested on Pair displays. For example, note that participants trained on configural stimuli with single stimulus displays were 81% correct during transfer when tested on pair displays.

Note that transfer accuracy was good in all conditions (above 80%). The main effect of stimulus type was significant [mean proportion correct was 0.92 for nonconfigural stimuli, and 0.86 for configural stimuli; $\chi^2(1) = 6.838$, $P < 0.001$], suggesting that transfer performance was better for the nonconfigural stimuli than for the configural stimuli. In contrast, the main effect of display type was not significant [Display type: $\chi^2(1) = 0.213$, $P = 0.644$]. The interaction effect was also significant [$\chi^2(1) = 153.223$, $P < 0.001$]. Both simple main effects were significant [Configural stimuli: $\chi^2(1) = 47.079$, $P < 0.001$; Nonconfigural stimuli: $\chi^2(1) = 106.36$,

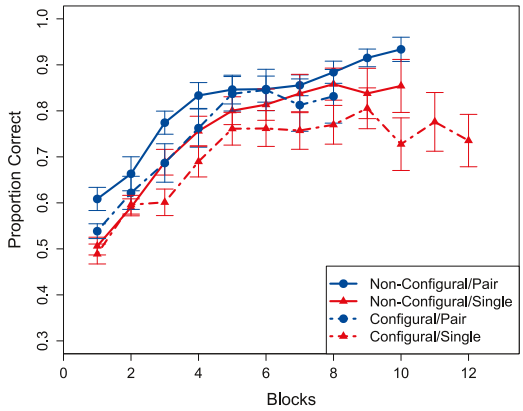

Figure 3. Experiment 1 learning curves for each condition from all blocks that included 10 or more participants. Error bars represent standard errors. See Figure 9 in the Appendix for the sample size of each block.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

Table 1. Logistic regression results from Experiment 1

Effect	χ^2	*df*	*P*
Trial	6730.523	1	<0.001
stimulus type	2.746	1	0.097
display type	9.514	1	0.002
display × stimulus	0.021	1	0.885
display × trial	74.147	1	<0.001
stimulus × trial	24.861	1	<0.001
display × stimulus × trial	67.539	1	<0.001

Display type includes Single and Pair conditions. Stimulus type includes Configural and Nonconfigural conditions.

$P<0.001$], indicating that transfer accuracy was significantly higher with pair-display training than with single-display training for the configural stimuli, which is consistent with the hypothesis that pair-display learning is MTL dependent and thus is flexible. However, this advantage was not seen in the Nonconfigural condition, suggesting that pair-display training without configural stimuli may not be as strongly MTL dependent.

Interestingly, we also found evidence that participants might have learned different strategies during pair- and single-display training. During pair-display training, participants always had to indicate which of the two stimuli belonged to the "Good" category. Then, during the transfer test, they had to indicate whether the single displayed stimulus belonged to the "Good" or "Bad" category. To assess how their pair-display training affected their single-display categorization, we examined the mean response times (RTs) of all perfect learners during the transfer test. Results showed that participants responded "Good" more quickly than "Bad" [$F_{(1,44)}=15.504$, $P<0.001$]. In contrast, for participants who trained with single displays, mean RTs to "Good" and "Bad" stimuli did not significantly differ during the last training block [$F_{(1,43)}=3.482$, $P=0.069$]. Therefore, pair-display training seemed to cause participants to learn a qualitatively different categorization strategy than single-display training. One possibility is that having to identify which of the two stimuli was "Good" encouraged a comparison process that was absent during single-display training and biased participants more toward "Good" stimuli than "Bad" stimuli.

Experiment 2

Experiment 1 showed that learning is faster with pair display training than with single display training, that performance was better with the nonconfigural stimuli than with the configural stimuli—regardless of display type, and also that the knowledge that is acquired with either type of training generalizes to the opposite display type. Experiment 2 focuses on the question of how the configural-stimulus categories are learned under single- and pair-display conditions.

Experiment 2 was also a 2 × 2 factorial design, with two levels of stimulus display (Single vs. Pair), and two levels of learning disruption (Feedback Delay vs. Task Switching). The stimuli were the same as the Configural conditions of Experiment 1 (i.e., see Fig. 1). All other methods were the same as in Experiment 1. For data analysis purposes, we included results of the Configural conditions from Experiment 1 as control conditions in which there was no learning disruption. Hereafter, we refer to these as the None conditions (because the level of learning disruption was none).

Figures 5 and 6 show the learning curves separately for the Feedback-Delay (Fig. 5) and Task-Switching conditions (Fig. 6), along with the learning curves from the None conditions as controls. Only data from all blocks that were completed by at least 10 participants are included. First, note from Figure 5 that the feedback delay appeared to have no effect on learning with pair displays, but to impair performance substantially with single displays. Second, note from Figure 6 that the apparent effect of switching quickly to the memory-scanning task was to impair learning in the Pair condition much more than in the Single condition.

We followed the same statistical approach as in Experiment 1. First, we tested the full model in which accuracy is a function of trial, display type (Single or Pair), and disruption type (None, Feedback Delay, or Task Switching). The results are shown in Table 2.

Note that there was a main effect of display type, suggesting that, as in Experiment 1, performance was significantly better in the Pair display conditions than in the Single conditions. In addition, the main effect of disruption type was also significant, which suggests that feedback delay and/or quickly switching to a second task impaired performance in at least some display conditions.

To investigate these effects more closely, we performed several follow-up logistic regression analyses. First, we separately examined the data from the Single/None and Single/Delay conditions [i.e., via a two-factor analysis—trial × disruption (None vs. Feedback Delay)]. The results showed that performance was significantly worse in the Feedback Delay condition [$\chi^2(1)=5.727$, $P=0.017$]. Next, we repeated this analysis for the Pair/None and Pair/Delay conditions. In this case, the effect of feedback delay was not significant [$\chi^2(1)=0.266$, $P=0.606$]. The hypothesis that the feedback delay impaired learning more in the Single display condition is further supported by the fact that the number of participants who failed to achieve 60% correct by block 8 (the fewest blocks completed by any nonlearner) was significantly higher in the Single/Delay condition than in the Pair/Delay condition (12 of 27 participants in the Single/Delay condition; three of 25 participants in the Pair/Delay condition; $Z=2.58$, $P=0.0049$).

We then performed similar analyses to investigate the effects of quickly switching to the second (memory-scanning) task. In the Single display conditions, Task-Switching performance was not significantly different from None performance [$\chi^2(1)=1.144$, $P=0.285$], but with Pair displays, performance was worse in the Task-Switching than in the None condition [$\chi^2(1)=3.999$, $P=0.046$]. The number of nonlearners was not significantly different in these two conditions (6 of 21 participants in the Pair/Task-Switching condition; 7 of 22 participants in the Single/Task-Switching condition; $Z=0.232$, $P=0.408$). Critically, performance in the Pair/Task-Switching condition was no better than performance in either of the Single conditions.

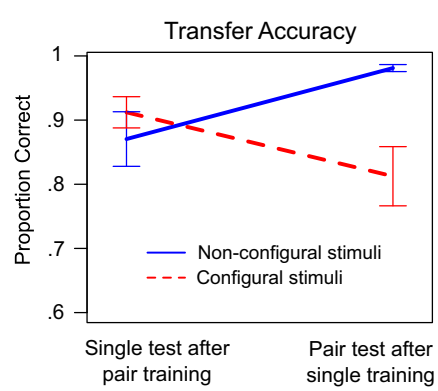

Figure 4. Experiment 1 accuracy during the final transfer-test block for perfect learners (i.e., all participants who achieved 100% correct during the last training block). Error bars represent standard errors. See Figure 9 in the Appendix for the sample size of each group.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

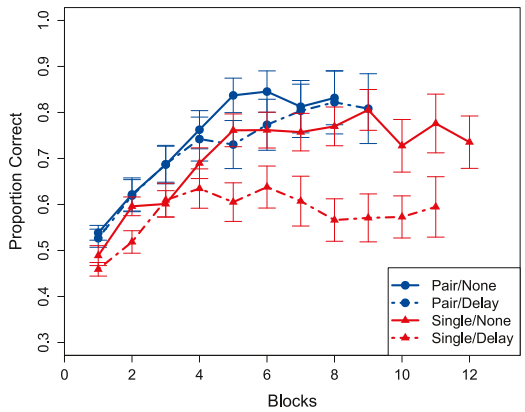

Figure 5. Experiment 2 learning curves for the None and Feedback-Delay conditions from all blocks that included 10 or more participants. Error bars represent standard errors. See Figure 9 in the Appendix for the sample size of each block.

We also replicated all these analyses without excluding any data—that is, we included data from all blocks, even those for which the sample size was less than 10. The results were qualitatively identical. In particular, Pair performance was still significantly better than Single performance, feedback delay still impaired Single performance but not Pair performance, and task switching still impaired Pair but not Single performance.

One possible concern is that our results may have been biased by the unequal sample sizes across blocks that resulted from our strategy to increase participant motivation by terminating their training after any block in which they achieved perfect accuracy (see Fig. 9 in the Appendix for the sample sizes of each block in every condition of both experiments). To investigate this possibility, we repeated all analyses on the data from the first 8 blocks for all participants who completed at least 8 blocks of training. This was the largest number of blocks that included at least 10 participants in all conditions. The learning curves are shown in Figures 7 (Feedback-Delay conditions) and 8 (Task-Switching conditions). We replicated all the above analyses on these reduced data sets and the results were again qualitatively identical. The same results were also obtained when the analyses were done on data from the first 6, 7, or 9 blocks of training, rather than the first 8 (see Fig. 10 in the Appendix).

Note that with each disruption type, some participants were able to learn perfectly within 7 blocks. Of course, this does not mean that they were not affected by the manipulation. For example, a Pair-Display participant who responded correctly on every trial of block 7 under Task-Switching conditions might have responded correctly on every trial of block 5 under None conditions. Our experiment was not designed to address this question. The important point is that our results show that, at the group level, task switching impairs pair-display category learning.

Discussion

We reported the results from two experiments that used unstructured categories. The experiments were designed to investigate the effects on learning of two different design features. One feature was whether the stimulus display included one or two stimuli, and the second was whether or not category membership depended on configural properties of the stimulus features.

Our results supported prior conclusions that with single stimulus displays, unstructured category learning depends on procedural learning that previous research suggests is mediated in the striatum, regardless of whether category membership depended on configural properties of the stimuli. The best evidence for this came in Experiment 2, which showed that a feedback delay of 2.5 sec highly impaired single-display learning, but had no effect on learning with pair displays. Prior research has shown that feedback delays of this type impair learning in tasks that are known to be mediated by procedural learning, but have little or no effect on learning in tasks that depend on declarative memory systems (Maddox et al. 2003; Maddox and Ing 2005; Dunn et al. 2012; Crossley and Ashby 2015).

Our most important finding, however, was that a switch to pair displays seemed to cause a switch to learning strategies that are supported by declarative memory systems, especially when category membership depended on configural properties of the stimulus features. First, although the categories were identical in the Single and Pair conditions, learning was significantly better with the pair displays, and this advantage appeared to be even greater with configural stimuli. Second, the same feedback delay that impaired single-display learning had no effect on pair-display learning. Third, a manipulation that reduced the time and attention available for feedback processing (i.e., quickly switching to a second task) impaired pair-display learning but not single-display learning. Maddox et al. (2004a) reported that this same manipulation impaired learning in a declarative-memory mediated task, but not in a task that recruited procedural learning.

Although more research is needed, the most likely scenario is that learning with pair displays depended on MTL memory systems. First, the absence of a feedback-delay effect rules out procedural learning. Second, the unstructured nature of the categories means that there is no simple explicit rule that determines category membership and therefore learning could not have depended on rule discovery. Third, Experiment 2 closely followed the design of Sanderson et al. (2006), who reported that MTL lesions impaired learning with pair displays, but not with single displays.

Of course, humans have a larger repertoire of executive skills and a higher working-memory capacity than rats, and therefore, our design cannot fully exclude the possibility that learning in the Pair conditions was mediated primarily by working memory. For example, participants may have verbalized an association of the category label with the stimulus configuration (in the case of the configural stimuli) or with some unique set of stimulus features (in the case of the nonconfigural stimuli), and then retained and rehearsed this association in working memory during the task. If so, then the MTL would not be critical for humans to learn these

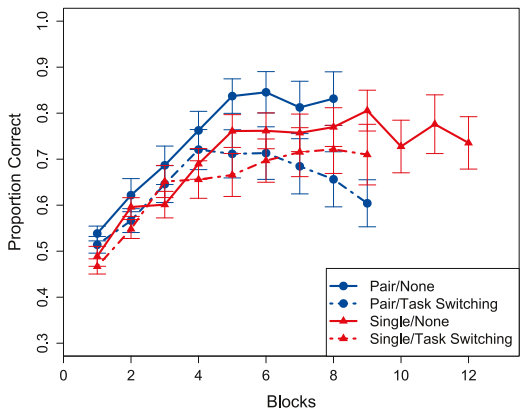

Figure 6. Experiment 2 learning curves for the None and Task-Switching conditions from all blocks that included 10 or more participants. Error bars represent standard errors. See Figure 9 in the Appendix for the sample size of each block.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

Table 2. Logistic regression results from Experiment 2

Effect	χ^2	*df*	*P*
Trial	8675.294	1	<0.001
display type	14.279	1	<0.001
disruption type	6.215	2	0.045
display × disruption	3.900	2	0.142
disruption × trial	803.831	2	<0.001
display × trial	1236.004	1	<0.001
display × trial × disruption	285.849	2	<0.001

Display type includes Single and Pair. Disruption type includes None, Feedback Delay, and Task Switching.

categories. Even so, there are a number of reasons this hypothesis seems unlikely. First, Figure 3 shows that even in the easiest condition (i.e., Pair display/Nonconfigural stimuli) average accuracy was still increasing after 4 blocks of training. The duration of learning therefore exceeded the duration of working memory maintenance by several orders of magnitude. Second, it has been reported that MTL-lesioned patients are impaired in spatial associative learning with just six pairs of stimuli (Petrides 1985). Third, Barense et al. (2005) reported that "feature ambiguity" predicts learning difficulty for MTL-lesioned patients better than the number of items; specifically, they demonstrated that when each stimulus feature is associated with both categories, exactly as with our configural stimuli, MTL-lesioned patients were impaired in discrimination learning with as few as four stimulus pairs. For these reasons, we believe the most likely hypothesis is that success in our pair-display conditions recruited MTL learning and memory systems.

Intuition seems to suggest that unstructured categories must require explicit memorization for learning to be successful, and therefore that MTL systems should be recruited, regardless of the conditions. Episodic memory surely mediates the participant's conscious recollection of specific stimuli that were shown during the experimental session, but with single-stimulus displays, episodic memory might not be ideally suited for using the trial-by-trial feedback to associate an arbitrary categorization response with each stimulus. Instead, such feedback-based trial-and-error learning is a hallmark of striatal-mediated procedural learning.

If procedural learning is advantageous, then why isn't it used in the pair-display conditions? Procedural learning requires a consistent association between a stimulus and a response. For example, performance is impaired if the location of the response keys is switched after learning is complete in tasks that recruit procedural learning, but not in declarative-memory mediated tasks (Ashby et al. 2003; Maddox et al. 2004b; Spiering and Ashby 2008; Crossley et al. 2012). With pair displays, no such consistent mapping is available, and so procedural learning is defeated. Evidence supporting this hypothesis comes from Helie and Ashby (2012), who reported results from a same-different categorization task that used a pair display. On each trial, two stimuli were presented and the participant's task was to indicate whether the stimuli were from the same or different categories. There was no evidence of any learning with categories that were learned procedurally under single-display conditions, whereas learning was good with categories that recruited declarative strategies in single-display conditions.

We also found other evidence that learning with pair displays was qualitatively different than learning with single displays. First, accuracy was consistently higher with pair displays. This could be because pair displays provide more information than single displays. For example, a well-known result from signal detection theory is that d' in pair-display tasks is $\sqrt{2}$ larger than in single-display tasks if an optimal decision strategy is used. But the optimal strategy with pair displays is different than with single displays—for example, it requires comparing the two stimuli. Therefore, signal detection theory predicts that the higher accuracy we observed with pair displays is because participants used different strategies with single and pair displays. Second, after training with pair displays, participants responded to the "Good" stimuli more quickly than the Bad stimuli during the single-display transfer-test block. In contrast, this difference was not evident in participants who trained with single displays. Therefore, pair-display training may have encouraged participants to compare and contrast the two stimuli, which might then have led to better declarative memory for "Good" stimuli.

Our results also suggest that the use of stimuli in which category membership depends on configural properties of the stimulus features (i.e., see Fig. 1) had at most a small effect on the nature of learning. First, the evidence strongly suggested that learning was mediated by procedural systems when the display included a single stimulus, even when category membership depended on configural properties of the stimulus. Second, the advantage of pair displays over single displays persisted with the nonconfigural stimuli, in which the configural relations among stimulus features were less critical. Even so, more research is needed to settle this question since so many previous studies have implicated the MTL in encoding configural properties of a stimulus (e.g., Rudy and Sutherland 1989; Sutherland and Rudy 1989; Aggleton et al. 2007; Duncan et al. 2018).

It is also important to note that our results are inconsistent with the leading neurobiologically detailed theory of human category learning—called COVIS (Ashby et al. 1998; Ashby and Valentin 2017). Briefly, COVIS assumes that humans learn categories in at least two qualitatively different ways. An executive attentional system uses working memory to learn explicit rules, whereas a procedural system uses dopamine-mediated reinforcement learning when perceptual similarity determines category membership and the optimal strategy is difficult or impossible to describe verbally and also to learn the arbitrary stimulus-response associations required with unstructured categories. The rule-learning system is mediated by a broad neural network that includes the prefrontal cortex, the anterior cingulate, and the MTL, whereas the procedural-learning system depends primarily on the basal

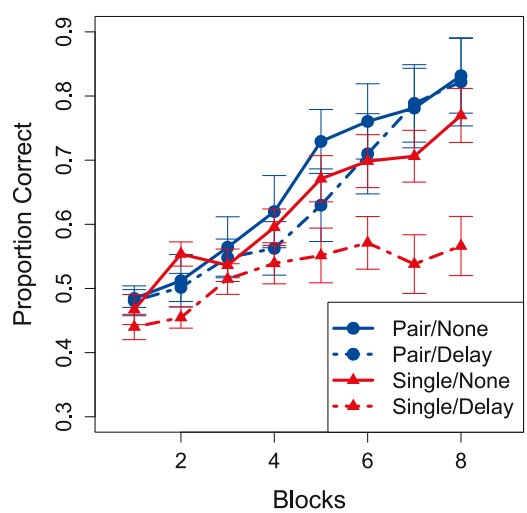

Figure 7. Experiment 2 learning curves for the None and Feedback-Delay conditions for the first 8 blocks. Participants who were excused before block 8 were excluded so that each condition includes equal sample sizes in every block. Error bars represent standard errors.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

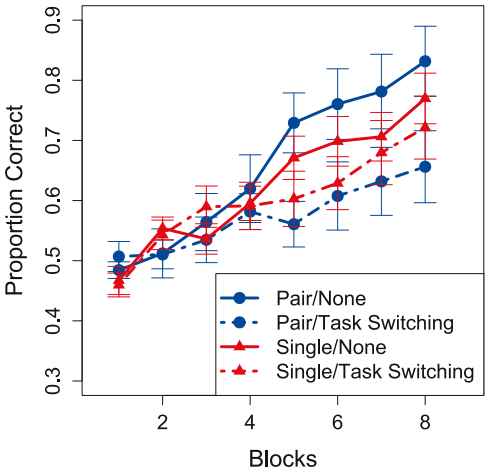

Figure 8. Experiment 2 learning curves for the None and Task-Switching conditions for the first 8 blocks. Participants who were excused before block 8 were excluded so that each condition includes equal sample sizes in every block. Error bars represent standard errors.

ganglia, and especially the striatum. COVIS assigns a role to the MTL, but only in consolidating memories of which categorization rules were investigated during the course of learning. A complete theory of category learning will need to generalize this component of the COVIS model to account for the results described here.

Historically, the cognitive neuroscience of category learning literature has focused heavily on the critical roles played by the basal ganglia. Meanwhile, the contributions of the MTL have been more controversial, and unambiguous evidence of MTL involvement has been limited to certain restricted special cases (e.g., the learning of exceptions). Our results suggest the interesting hypothesis that this bias in the literature may be due largely to the predominance of single-stimulus displays in category-learning experiments. Pair displays may naturally recruit MTL-mediated declarative learning, rather than basal ganglia-mediated procedural learning. Furthermore, pair displays may also promote different categorization strategies than single displays. For example, categorization with pair displays may be more of a comparative process than with single displays. We believe these hypotheses should be the targets of future research.

Materials and Methods

Experiment 1

Participants

One hundred and two undergraduates from UC Santa Barbara participated as partial fulfillment of a course requirement. They were each randomly assigned to one of the four experimental conditions: 26 in Pair/Configural, 28 in Single/Configural, 24 in Single/Nonconfigural, 24 in Pair/Nonconfigural.

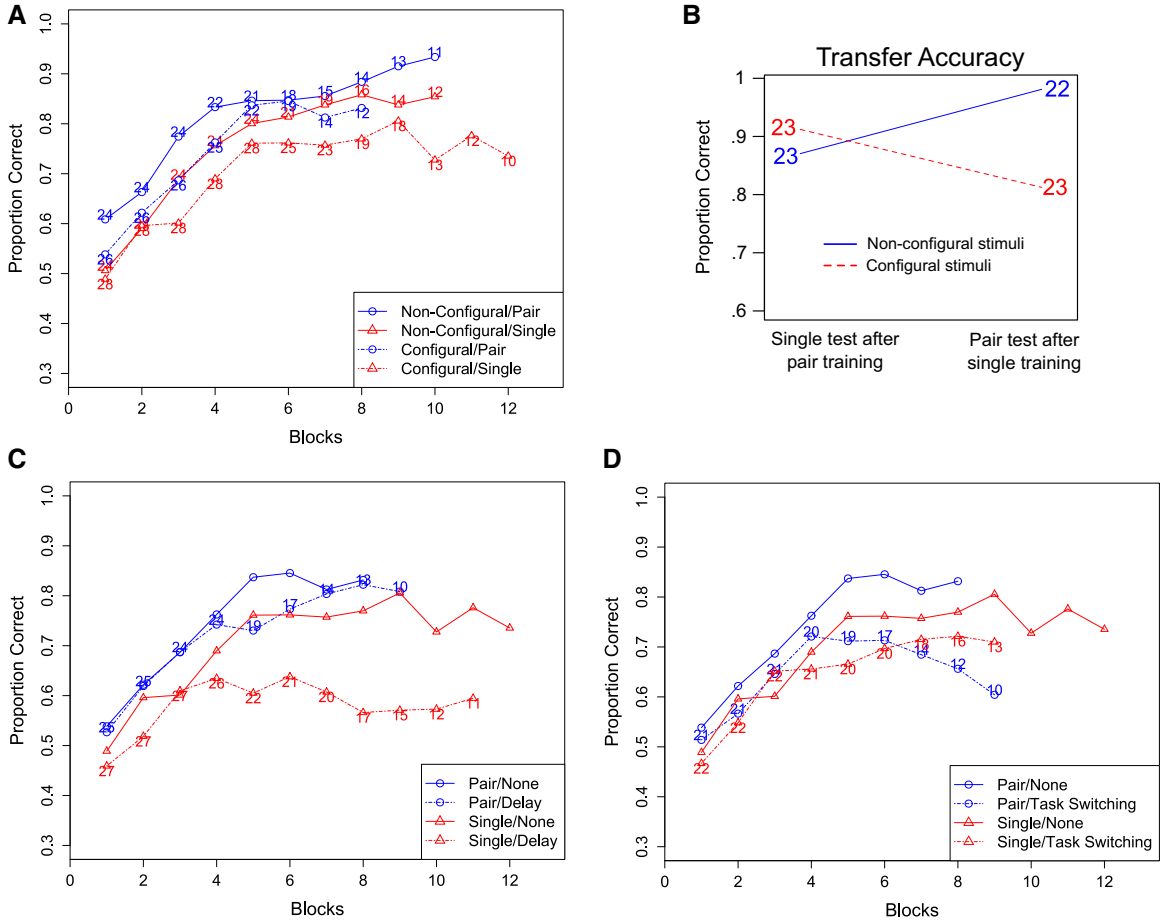

Figure 9. Sample sizes in each block and condition of Experiments 1 and 2.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

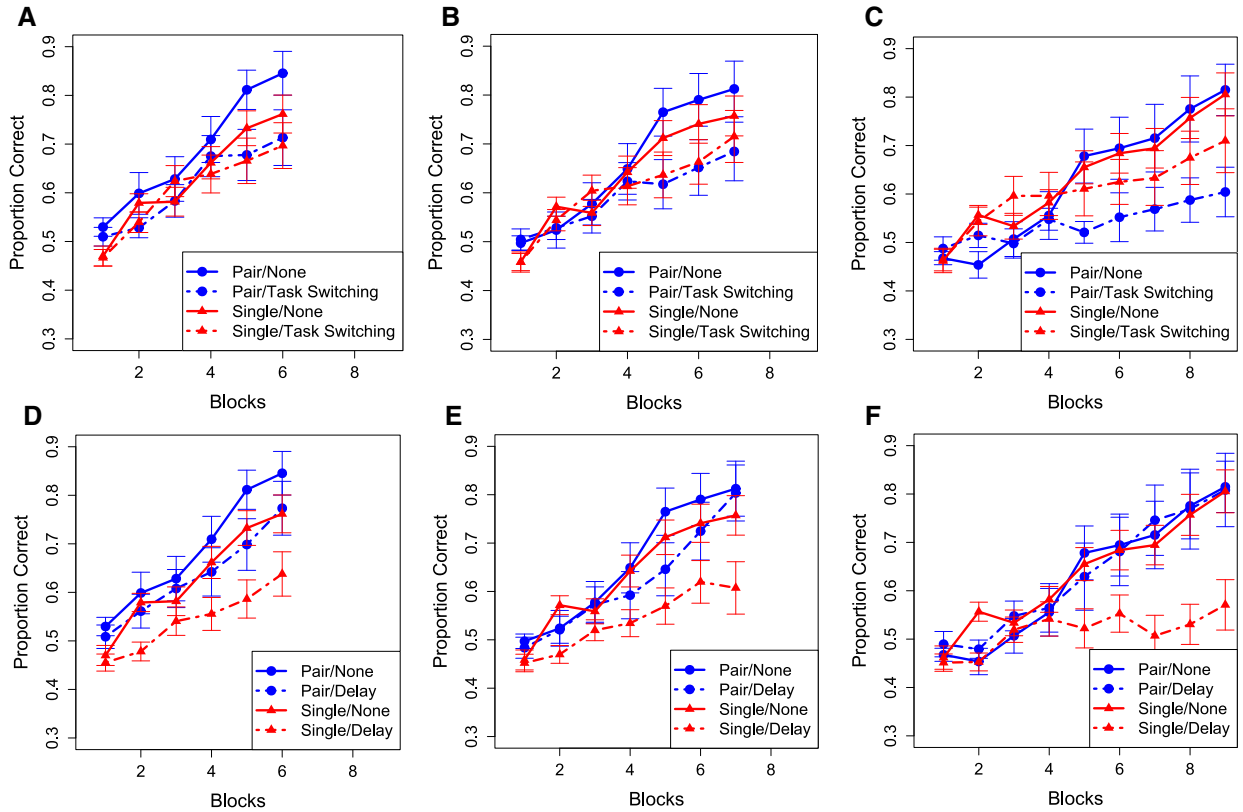

Figure 10. Learning curves from Experiment 2 for the Task-Switching (*A*–*C*) and Delay (*D*–*F*) conditions, along with the Experiment 1 curves for comparison (None condition). Each panel includes equal sample sizes in every block. Panels A and D include data from all participants who completed at least six blocks, panels B and E include data from all participants who completed at least seven blocks, and panels C and F include data from all participants who completed at least nine blocks. Error bars represent standard errors. The sample sizes for each panel are shown in Figure 9.

Stimuli and apparatus

The stimuli are shown in Figure 1. Half of the configural stimuli were from Sanderson et al. (2006), and half were created based on the same principles. The 12 stimuli formed six mirror-image pairs. Each figure included 2 elements, and each element was as likely to be on the left as on the right in both categories. Eight of the nonconfigural stimuli were identical to those used by Crossley et al. (2012) and Seger and Cincotta (2005), and four were created based on the principles described by Seger and Cincotta (2005). The stimuli were presented on a gray background using the Psychophysics Toolbox (Brainard 1997) running in MATLAB. The order in which the stimuli were presented was randomized across participants and blocks. Responses were made with keys labeled on a keyboard (d key labeled as "A," and k key as "B").

Procedures

The experiment was a 2 × 2 fully crossed factorial design, with two levels of stimulus display (single vs. pair), and two levels of stimulus type (configural vs. nonconfigural). Each condition included several blocks of learning. There was a final transfer-test block at the end. Each block included 48 trials.

The experiment lasted no more than 50 min depending on the participant's response speed and the number of blocks required for learning. Participants were told that the stimuli could be separated into two categories "Good" and "Bad," and that their task was to learn to assign each stimulus to its correct category. To motivate participants, they were told that the learning stage would end as soon as they achieved 100% correct in a block, which therefore would allow them to leave earlier. Otherwise, the learning stage would terminate after 45 min.

On each trial of the learning stage, the stimuli were displayed on the screen until the participant pressed a key. The Single and Pair conditions are illustrated in Figure 2. In the Single conditions, only one figure was displayed in the middle of the screen on each trial, and participants had to assign the stimulus to the "Good" or "Bad" category by pressing the "A" key for "Good" and the "B" key for "Bad." Participants were instructed to use the left index finger for the "A" key, and the right index finger for the "B" key, and to respond within 5 sec. In the Pair conditions, one stimulus from each category was displayed on the screen, one on the left and the other on the right, with their position randomized on each trial. Participants were told that only one of the two figures was "Good," and they had to identify the "Good" figure by pressing the "A" key if it was on the left and the "B" key if it was on the right. In the configural conditions, the two stimuli were always mirror images.

The participant's response was followed by a 500 msec blank screen, and then a line of text was displayed for 500 msec as feedback. If the response time was more than 5 sec, they would see "Too Slow" in white; otherwise, they would see "Correct" in green or "Incorrect" in red, corresponding to the accuracy of their response. The interval between trials was 1 sec for all conditions.

The final block of 48 trials was a transfer-test block in which no feedback was provided. Participants in the Single conditions received Pair displays during this block, whereas participants in the Pair conditions received Single displays.

Experiment 2

Participants

Ninety five undergraduates from UC Santa Barbara participated as partial fulfillment of a course requirement. They were each

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

randomly assigned to one of the four experimental conditions: 25 in Pair/Feedback Delay, 21 in Pair/Task Switching, 27 in Single/Feedback Delay, and 22 in Single/Task Switching.

Stimuli and apparatus

The stimuli and apparatus were identical to those used in the Configural conditions of Experiment 1.

Conditions and procedures

The experiment was a 2 × 2 factorial design, with two levels of stimulus display (Single vs. Pair) crossed with two levels of learning disruption (Feedback Delay vs. Task Switching). Each condition included several blocks of learning. Unlike Experiment 1, there was no final transfer-test block. Each block included 48 trials.

The procedures were identical to Experiment 1, except that in the Feedback-Delay conditions, the response was followed by a 2.5 sec noise mask, and then by the same 500 msec text feedback as in the Control conditions; and in the Task-Switching conditions, the response was followed by a 500 msec blank screen, the 500 msec feedback, and then immediately by the Sternberg (1966) memory-scanning task. On each memory-scanning trial, four random digits were shown on the screen for 1 sec, followed by a blank screen for 1 sec, then a random digit, and then the probe "Was it in the list? (A) Yes, (B) No". The "Yes" and "No" responses were randomly paired with the "A" and "B" keys on each trial. There was no feedback in the memory-scanning task. To ensure that participants paid enough attention to the memory-scanning task, a warning "Please pay more attention to the memory task" would appear on the screen every 10 trials when memory-scanning accuracy dropped below 90%. The interval between trials was 1 sec for all conditions.

Acknowledgments

This research was supported by grant 2R01MH063760 from the National Institutes of Health.

Appendix

This appendix provides more detailed results of Experiments 1 and 2. Figure 9 shows the sample sizes for each block and condition of both experiments. Figure 10 shows learning curves for all participants who completed at least six blocks of training (panels A and D), seven blocks of training (panels B and E), or nine blocks of training (panels C and F).

References

Aggleton JP, Sanderson DJ, Pearce JM. 2007. Structural learning and the hippocampus. *Hippocampus* **17:** 723–734. doi:10.1002/hipo.20323

Ashby FG, Ennis JM. 2006. The role of the basal ganglia in category learning. *Psychol Learn Motiv* **46:** 1–36. doi:10.1016/S0079-7421(06)46001-1

Ashby FG, Valentin VV. 2017. Multiple systems of perceptual category learning: theory and cognitive tests. In *Handbook of categorization in cognitive science*, 2nd edn (ed. Cohen H, Lefebvre C), pp. 157–188. Elsevier, New York.

Ashby FG, Alfonso-Reese LA, Turken A, Waldron EM. 1998. A neuropsychological theory of multiple systems in category learning. *Psychol Rev* **105:** 442–481. doi:10.1037/0033-295X.105.3.442

Ashby FG, Ell SW, Waldron EM. 2003. Procedural learning in perceptual categorization. *Mem Cogn* **31:** 1114–1125. doi:10.3758/BF03196132

Barense MD, Bussey TJ, Lee AC, Rogers TT, Davies RR, Saksida LM, Murray EA, Graham KS. 2005. Functional specialization in the human medial temporal lobe. *J Neurosci* **25:** 10239–10246. doi:10.1523/JNEUROSCI.2704-05.2005

Barker GRI, Warburton EC. 2013. Object-in-place associative recognition memory depends on glutamate receptor neurotransmission within two defined hippocampal-cortical circuits: a critical role for ampa and nmda receptors in the hippocampus, perirhinal, and prefrontal cortices. *Cereb Cortex* **25:** 472–481. doi:10.1093/cercor/bht245

Bayley PJ, Frascino JC, Squire LR. 2005. Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature* **436:** 550–553. doi:10.1038/nature03857

Brainard DH. 1997. The psychophysics toolbox. *Spat Vis* **10:** 433–436. doi:10.1163/156856897X00357

Crossley MJ, Ashby FG. 2015. Procedural learning during declarative control. *J Exp Psychol Learn Mem Cogn* **41:** 1388–1403. doi:10.1037/a0038853

Crossley MJ, Madsen NR, Ashby FG. 2012. Procedural learning of unstructured categories. *Psychon Bull Rev* **19:** 1202–1209. doi:10.3758/s13423-012-0312-0

Davis T, Love BC, Preston AR. 2012. Learning the exception to the rule: model-based fMRI reveals specialized representations for surprising category members. *Cereb Cortex* **22:** 260–273. doi:10.1093/cercor/bhr036

Duncan K, Doll BB, Daw ND, Shohamy D. 2018. More than the sum of its parts: a role for the hippocampus in configural reinforcement learning. *Neuron* **98:** 645–657. doi:10.1016/j.neuron.2018.03.042

Dunn JC, Newell BR, Kalish ML. 2012. The effect of feedback delay and feedback type on perceptual category learning: the limits of multiple systems. *J Exp Psychol Learn Mem Cogn* **38:** 840–859. doi:10.1037/a0027867

Featherstone R, McDonald R. 2004. Dorsal striatum and stimulus-response learning: lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a stimulus-response-based instrumental discrimination task, while sparing conditioned place preference learning. *Neuroscience* **124:** 23–31. doi:10.1016/j.neuroscience.2003.10.038

Filoteo JV, Maddox WT, Davis JD. 2001. Quantitative modeling of category learning in amnesic patients. *J Int Neuropsychol Soc* **7:** 1–19. doi:10.1017/S1355617701711010

Graham KS, Scahill VL, Hornberger M, Barense MD, Lee AC, Bussey TJ, Saksida LM. 2006. Abnormal categorization and perceptual learning in patients with hippocampal damage. *J Neurosci* **26:** 7547–7554. doi:10.1523/JNEUROSCI.1535-06.2006

Helie S, Ashby FG. 2012. Learning and transfer of category knowledge in an indirect categorization task. *Psychol Res* **76:** 292–303. doi:10.1007/s00426-011-0348-1

Hopkins RO, Myers EC, Shohamy D, Grossman S, Gluck M. 2004. Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia* **42:** 524–535. doi:10.1016/j.neuropsychologia.2003.09.005

Horvitz JC. 2009. Stimulus-response and response–outcome learning mechanisms in the striatum. *Behav Brain Res* **199:** 129–140. doi:10.1016/j.bbr.2008.12.014

Janowsky JS, Shimamura AP, Kritchevsky M, Squire LR. 1989. Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behav Neurosci* **103:** 548–560. doi:10.1037/0735-7044.103.3.548

Knowlton BJ, Squire LR. 1993. The learning of natural categories: parallel memory systems for item memory and category-level knowledge. *Science* **262:** 1747–1749. doi:10.1126/science.8259522

Knowlton BJ, Squire LR, Gluck MA. 1994. Probabilistic classification learning in amnesia. *Learn Mem* **1:** 106–120.

Knowlton BJ, Mangels JA, Squire LR. 1996. A neostriatal habit learning system in humans. *Science* **273:** 1399–1402. doi:10.1126/science.273.5280.1399

Kolodny JA. 1994. Memory processes in classification learning: an investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychol Sci* **5:** 164–169. doi:10.1111/j.1467-9280.1994.tb00654.x

Leng NR, Parkin AJ. 1988. Double dissociation of frontal dysfunction in organic amnesia. *Br J Clin Psychol* **27:** 359–362. doi:10.1111/j.2044-8260.1988.tb00800.x

Lopez-Paniagua D, Seger CA. 2011. Interactions within and between corticostriatal loops during component processes of category learning. *J Cogn Neurosci* **23:** 3068–3083. doi:10.1162/jocn_a_00008

Maddox WT, Ing AD. 2005. Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *J Exp Psychol Learn Mem Cogn* **31:** 100–107. doi:10.1037/0278-7393.31.1.100

Maddox WT, Ashby FG, Bohil CJ. 2003. Delayed feedback effects on rule-based and information-integration category learning. *J Exp Psychol Learn Mem Cogn* **29:** 650–662. doi:10.1037/0278-7393.29.4.650

Maddox WT, Ashby FG, Ing AD, Pickering AD. 2004a. Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Mem Cogn* **32:** 582–591. doi:10.3758/BF03195849

Maddox WT, Bohil CJ, Ing AD. 2004b. Evidence for a procedural-learning-based system in perceptual category learning. *Psychon Bull Rev* **11:** 945–952. doi:10.3758/BF03196726

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

Petrides M. 1985. Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* **23:** 601–614. doi:10.1016/0028-3932(85)90062-4

Renzi C, Schiavi S, Carbon C-C, Vecchi T, Silvanto J, Cattaneo Z. 2013. Processing of featural and configural aspects of faces is lateralized in dorsolateral prefrontal cortex: a TMS study. *Neuroimage* **74:** 45–51. doi:10.1016/j.neuroimage.2013.02.015

Rudy JW, Sutherland RJ. 1989. The hippocampal formation is necessary for rats to learn and remember configural discriminations. *Behav Brain Res* **34:** 97–109. doi:10.1016/S0166-4328(89)80093-2

Sanderson DJ, Pearce JM, Kyd RJ, Aggleton JP. 2006. The importance of the rat hippocampus for learning the structure of visual arrays. *Eur J Neurosci* **24:** 1781–1788. doi:10.1111/j.1460-9568.2006.05035.x

Seger CA, Cincotta CM. 2005. The roles of the caudate nucleus in human classification learning. *J Neurosci* **25:** 2941–2951. doi:10.1523/JNEUROSCI.3401-04.2005

Seger CA, Miller EK. 2010. Category learning in the brain. *Annu Rev Neurosci* **33:** 203–219. doi:10.1146/annurev.neuro.051508.135546

Seger CA, Peterson EJ, Cincotta CM, Lopez-Paniagua D, Anderson CW. 2010. Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and granger causality modeling. *Neuroimage* **50:** 644–656. doi:10.1016/j.neuroimage.2009.11.083

Shohamy D, Myers C, Grossman S, Sage J, Gluck M, Poldrack R. 2004. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain* **127:** 851–859. doi:10.1093/brain/awh100

Spiering BJ, Ashby FG. 2008. Response processes in information-integration category learning. *Neurobiol Learn Mem* **90:** 330–338. doi:10.1016/j.nlm.2008.04.015

Squire LR, Knowlton BJ. 1995. Learning about categories in the absence of memory. *Proc Natl Acad Sci* **92:** 12470–12474. doi:10.1073/pnas.92.26.12470

Stark SM, Frithsen A, Mattfeld AT, Stark CE. 2018. Modulation of associative learning in the hippocampal-striatal circuit based on item-set similarity. *Cortex* **109:** 60–73. doi:10.1016/j.cortex.2018.08.035

Sternberg S. 1966. High-speed scanning in human memory. *Science* **153:** 652–654. doi:10.1126/science.153.3736.652

Sutherland RJ, Rudy JW. 1989. Configural association theory: the role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology* **17:** 129–144.

Swainson R, SenGupta D, Shetty T, Watkins L, Summers B, Sahakian B, Polkey CE, Robbins T. 2006. Impaired dimensional selection but intact use of reward feedback during visual discrimination learning in parkinson's disease. *Neuropsychologia* **44:** 1290–1304. doi:10.1016/j.neuropsychologia.2006.01.028

Yagishita S, Hayashi-Takagi A, Ellis-Davies GC, Urakubo H, Ishii S, Kasai H. 2014. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* **345:** 1616–1620. doi:10.1126/science.1255514

Zaki SR, Nosofsky RM, Jessup NM, Unversagt FW. 2003. Categorization and recognition performance of a memory-impaired group: evidence for single-system models. *J Int Neuropsychol Soc* **34:** 394–406. doi:10.1017/S1355617703930050

Zeithamova D, Mack ML, Braunlich K, Davis T, Seger CA, van Kesteren MT, Wutz A. 2019. Brain mechanisms of concept learning. *J Neurosci* **39:** 8259–8266. doi:10.1523/JNEUROSCI.1166-19.2019

Received May 21, 2020; accepted in revised form June 19, 2020.

Downloaded from learnmem.cshlp.org on September 15, 2020 - Published by Cold Spring Harbor Laboratory Press

LEARNING & MEMORY

A role for the medial temporal lobes in category learning

Yi-Wen Wang and F. Gregory Ashby

Learn. Mem. 2020, **27:**
Access the most recent version at doi:10.1101/lm.051995.120

References This article cites 47 articles, 10 of which can be accessed free at:
http://learnmem.cshlp.org/content/27/10/441.full.html#ref-list-1

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first 12 months after the full-issue publication date (see http://learnmem.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

© 2020 Wang and Ashby; Published by Cold Spring Harbor Laboratory Press