Retinal-Specific Category Learning

Luke A. Rosedahl, Miguel P. Eckstein, & F. Gregory Ashby*

University of California, Santa Barbara

Virtually all cognitive theories of category learning (such as prototype theory [1,2,3,4,5] and exemplar theory [6,7,8]) view this important skill as a high-level process that uses abstract representations of objects in the world. Because these representations are removed from visual characteristics of the display, such theories suggest that category learning occurs in higher-level (e.g., association) areas and therefore should be immune to the visual field dependencies that characterize processing of objects mediated by representations in low-level visual areas. This article challenges that view by describing a fully controlled demonstration of visual-field dependence in category learning. Eye-tracking was used to control gaze while participants either learned rule-based categories known to recruit prefrontal-based explicit reasoning, or information-integration categories known to depend on basal-ganglia-mediated procedural learning [9]. Results showed that learning was visual-field dependent with information-integration categories, but we found no evidence of visual-field dependence with rule-based categories. A theoretical interpretation of this difference is offered in terms of the underlying neurobiology. Finally, these results are situated within the broad perceptual-learning literature in an attempt to motivate further research on the similarities and differences between category and perceptual learning.

Categorization is an essential aspect of our daily lives. We make thousands of categorization judgments each day and many can lead to devastating consequences when performed incorrectly (e.g., mistaking a poisonous mushroom as edible). Historically, the categorization literature has overwhelmingly treated category learning as a high-level cognitive process that is largely independent of sensory systems. For example, prototype theory assumes that categorization is a process of activating mental representations of the category prototypes, which are abstract representations of the most typical member in each category [1,2,3,4,5]. Exemplar theory assumes categorization is a process of accessing memory representations of the category exemplars, which are modeled as single points in an abstract psychological space [6,7,8]. The common theme here is that categorization invokes complex comparison processes and abstract stimulus representations that are far removed from the sensory systems used to process the stimuli.

This article reports the results of an experiment that provides evidence against this view – at least in the case of categories thought to be learned procedurally. Specifically, we describe evidence that the learning that occurs in such cases is retinal specific. After the categories are learned, categorization accuracy deteriorates much more if the stimulus is moved to a new untrained location on the participant's retina than if it is moved to the same relative retinal location in the untrained eye, or with any type of movement when the categories can be learned explicitly (e.g., via a rule). These results suggest that procedural category learning is much more dependent on low-level visual representations than rule-based category learning.

There is now abundant evidence that declarative and procedural memory both contribute to category learning [9,10,11,12,13]. Much of this evidence comes from rule-based (RB) and information-integration (II) category-learning tasks. In RB tasks, the categories can be learned via some explicit reasoning process [14]. In the most common applications, only one stimulus dimension is relevant, and the participant's task is to discover this relevant dimension and then to map the different dimensional values to the relevant categories. A variety of evidence suggests that success in RB tasks depends on declarative memory and especially on working memory and executive attention [14,15,16,17]. In II category-learning tasks, accuracy is maximized only if information from two or more incommensurable stimulus components is integrated at some predecisional stage [18,14]. Evidence suggests that success in II tasks depends on procedural memory, which forms stimulus-response associations via striatal-mediated reinforcement learning [19,20,21,22].

Figure I shows the II category-learning task used in the present experiment. Every stimulus was a circular, sine-wave grating that varied across trials on two stimulus dimensions – the width and orientation of the dark and light bars. Each symbol denotes a different stimulus. Note that there are four categories: A, B, C, and D. The stimuli corresponding to each category prototype (i.e., category mean) are also shown. The diagonal lines describe the optimal categorization strategy. Note that these lines divide the 100×100 stimulus space into four regions. Accuracy is maximized if the participant responds with the category label depicted in each of these regions to any stimulus that falls in that region. Note that this strategy is difficult (if not impossible) to describe verbally.

To create the RB category-learning task, we simply rotated the II stimulus space 45° clockwise (and therefore all of the stimulus coordinates were also rotated). This rotation converts the diagonal category bounds to vertical and
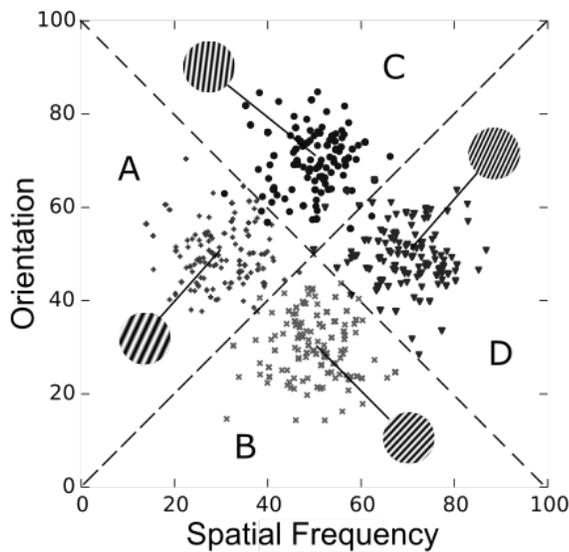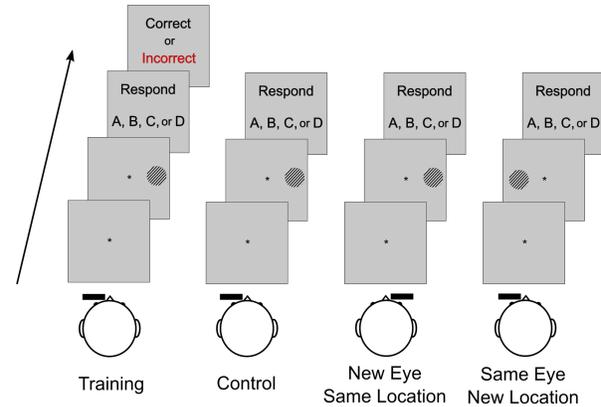
Figure 1. Sample stimuli and four II categories. The dotted lines denote the optimal category decision bounds. The units were chosen so that changes of the same magnitude on each dimension have approximately equal perceptual salience. See the Methods section for transformations that convert units of spatial frequency to cycles per degree, and units of orientation to degrees counterclockwise rotation from horizontal.

Figure 2. Order of events that occurred in each phase of the New Eye First condition. The New Location First condition was identical except the order of the last two phases was reversed.

horizontal while preserving all category-separation statistics. The rotation also makes the optimal categorization strategy easy to describe verbally. For example, note that after the rotation, the optimal strategy for identifying members of category A is to "Respond A if the bars are wide and the orientation is steep."

Figure II describes the procedures we used. Each participant was first trained on either the RB or II categories with feedback on every trial (no participants learned both category structures). Participants in the RB and II conditions were given identical instructions. They were simply told that their task was to use the feedback during training to learn to assign each disk to its correct category. During the training phase of the experiment, participants wore an eye patch over their left eye. They were instructed to maintain their fixation at the center of the screen throughout the entire categorization trial. The categorization stimulus was always presented 5° right of the fixation point. Eye movements were monitored and any movement after the stimulus appeared aborted the trial. After the training phase, all participants completed 50 more trials that were identical to training except the trial-by-trial feedback was withheld. We call this the control phase because it provides an estimate of the amount of learning that occurred during training.

After the control phase was completed, participants were

randomly assigned to one of two conditions that were identical except for the order of the final two phases. In the New Eye First condition, participants switched the eye patch to their right eye and then completed 50 more trials with the stimulus at the same location as during the control phase. Finally, during the last 50 trials of the experiment, they switched the eye patch back to their left eye and the stimulus was presented 5° left of fixation. So this final phase tested their categorization accuracy on the trained eye, but at an untrained retinal location. The New Location First condition was identical except the order of the last two phases was reversed.

In summary, each participant was trained on either the II or RB category structures and then tested with one of two orders of the test conditions, resulting in a 2×2 factorial design with 2 category structures (RB versus II) crossed with two training orders (New Eye First versus New Location First).

Experiment 1 included 100 participants – 50 in the RB conditions and 50 in the II conditions. All participants received 500 trials of training (with feedback) followed by the three test blocks of 50 (unsupervised) trials each, as described in Figure II. All participants who performed so poorly during the control testing block that we could not reject the null hypothesis that their accuracy was at chance (0.25; i.e., with $\alpha = .01$) were excluded from further analysis. This was necessary because asking whether learning transfers presupposes that there is some learning to transfer. This exclusion criterion eliminated 16 of the 100 participants from further analysis (12 from the RB condition and 4 from the II condition).

The learning curves during the training phase (Phase 1) for the II and RB categories are shown in the top panel of Figure III. Note that by the end of training, accuracy in-

creased to roughly the same level for both category structures (during the control testing block the accuracy difference between the two groups was nonsignificant: mean II accuracy 66%; mean RB accuracy 67%; $t(83) = .20, p = .84, d = .04, 95\%\text{CI}_{\Delta PC} = [-.07, .05]$).

To compare performance when the stimulus moved to the untrained eye versus a new location on the trained eye, we first computed two difference scores (ΔPC) for each participant: 1) New Eye Same Location test block accuracy minus Control test block accuracy, and 2) Same Eye New Location test block accuracy minus Control test block accuracy. Next, to compare these difference scores, we performed a three-factor mixed ANOVA that included two levels of category structure (II and RB), two levels of retinal location (new eye and same eye), and two levels of block order (2nd and 3rd test blocks; Control was always first), with repeated measures on the latter two factors. This analysis showed a main effect of retinal location [$F(1, 80) = 8.15, p = .005, \eta_p^2 = .093, 90\%\text{CI}_{\eta_p^2} = [0.016, .2]$], but not of category structure [$F(1, 80) = .03, p = .86, \eta_p^2 = .003, 90\%\text{CI}_{\eta_p^2} = [0.0, .02]$] or block [$F(1, 80) = 1.9, p = .17, \eta_p^2 = .02, 90\%\text{CI}_{\eta_p^2} = [0.0, .10]$]. However, the ANOVA reported marginally significant interactions for both category structure and retinal location [$F(1, 80) = 3.35, p = .07, \eta_p^2 = .04, 90\%\text{CI}_{\eta_p^2} = [0.0, .13]$] and category structure and block [$F(1, 80) = 2.91, p = .09, \eta_p^2 = .04, 90\%\text{CI}_{\eta_p^2} = [0.0, .12]$], demonstrating the need for further analysis.

To probe these results more closely we ran follow-up two-factor ANOVAs separately on the II and RB difference scores [two retinal locations (new location and new eye) × two blocks (2nd and 3rd), with repeated measures on both factors]. For the II difference scores, the main effect of retinal location was significant [$F(1, 44) = 15.15, p < .001, \eta_p^2 = .26, 90\%\text{CI}_{\eta_p^2} = [0.09, .41]$], but the main effect of block [$F(1, 44) = .04, p = .85, \eta_p^2 = .001, 90\%\text{CI}_{\eta_p^2} = [0.0, .05]$] and the retinal location × block interaction were nonsignificant [$F(1, 44) = .05, p = .83, \eta_p^2 = .001, 90\%\text{CI}_{\eta_p^2} = [0.0, .05]$]. In the case of the RB difference scores, the main effect of block was marginally significant [$F(1, 36) = 3.55, p = .067, \eta_p^2 = .09, 90\%\text{CI}_{\eta_p^2} = [0.0, .25]$], but the main effect of retinal location was not significant [$F(1, 36) = .22, p = .64, \eta_p^2 = .006, 90\%\text{CI}_{\eta_p^2} = [0.0, .10]$], nor was the retinal location × block interaction [$F(1, 36) = .002, p = .96, \eta_p^2 = 0.0, 90\%\text{CI}_{\eta_p^2} = [0.0, .0006]$].

Inspired by the perceptual-learning literature[23], we calculated a Transfer Index (TI) score for each condition. The TI measures the amount of learning that transferred from the training location to either the new retinal location or the new eye. We defined the TI as the mean proportion correct above chance at the new location divided by the mean proportion correct above chance during control testing. More specifically, let P_{NE} and P_C denote the mean proportion correct
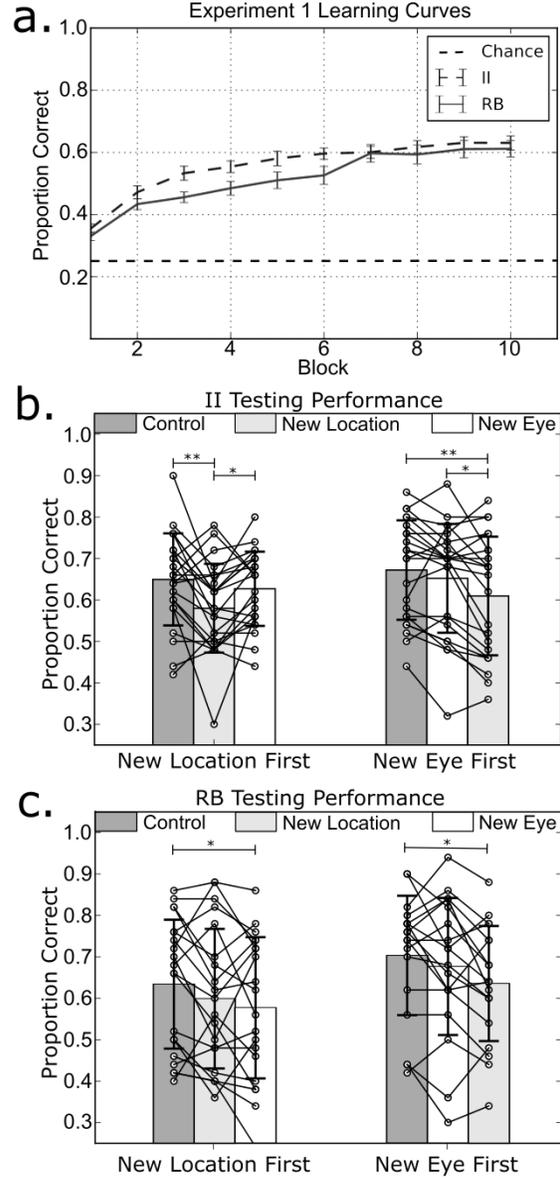

Figure 3. Results of Experiment 1. All error bars show standard error. (a) Learning curves of participants during RB ($n = 38$) and II ($n = 46$) training. Each block included 50 trials. (b & c) Individual participant and mean accuracies during each Experiment 1 test block for the II condition (b) and RB condition (c). Open circles denote individual participant accuracies and the bars denote group mean accuracy. Error bars show standard error. "**" signifies significance after FDR correction (FDR $\leq$ .05; largest significant p-value = .036). "*" signifies uncorrected significance at $\alpha = .05$. All p-values can be found in the text in the two paragraphs that follow Equation 1.

(across participants) during the New Eye phase and the Control phase, respectively. Then the TI for transfer to the new eye equals

$$TI_{NE} = \frac{P_{NE} - .25}{P_C - .25}. \tag{1}$$

The TI for transfer to a new retinal location (i.e., TI_{NL}) was defined analogously. Note that this measure equals 1 if transfer is perfect and 0 in the absence of any transfer. Note also that this statistic differs slightly from traditional perceptual-learning transfer indices. This is because accuracy often begins above chance in most perceptual-learning tasks, whereas in the category-learning tasks used here, accuracy necessarily begins at chance.

The accuracy of participants in each test block is shown in the middle and bottom panels of Figure III. All pairwise differences were assessed for statistical significance using post-hoc paired t-tests corrected for multiple comparisons by holding false discovery rate (FDR) at $\leq .05$ (significance after FDR correction is denoted in Figure III by ** and uncorrected significance at $\alpha = .05$ is denoted by *). Note that in the II conditions, accuracy dropped significantly from control levels with an average TI of .85 when the stimulus moved to a new retinal location on the trained eye (New Location First condition: $t(44) = 3.24, p = .002, d = .64, TI = .86, 95\%CI_{\Delta PC} = [-.12, -.03]$; New Eye First condition: $t(44) = 2.92, p = .006, d = .47, TI = .84, 95\%CI_{\Delta PC} = [-.10, -.03]$), but not when it moved to the untrained eye at the trained location (New Location First condition: $t(44) = 1.05, p = .30, d = .22, TI = .98, 95\%CI_{\Delta PC} = [-.05, 0.0]$; New Eye First condition: $t(44) = .93, p = .36, d = .16, TI = .94, 95\%CI_{\Delta PC} = [-.05, 0.0]$). Additionally, performance was lower in the untrained location than the untrained eye, with an average TI of .89 (New Location First Condition: $t(44) = 2.19, p = .03, d = .48, TI = .88, 95\%CI_{\Delta PC} = [-.08, -.01]$; New Eye First Condition: $t(44) = 1.93, p = .05, d = .31, TI = .90, 95\%CI_{\Delta PC} = [-.07, -.01]$), though this difference is not significant after FDR correction.

In the RB conditions, accuracy decreased monotonically with block. Performance was best in the first test block (i.e., control) and worst in the third test block regardless of retinal location, though none of the differences were statistically significant. In the New Eye First condition the TI from the first to the third block was .86 ($t(36) = 2.50, p = .02, d = .48, TI = .86, 95\%CI_{\Delta PC} = [-.10, -.03]$) while in the New Location First condition it was .84 ($t(36) = 2.22, p = .03, d = .35, TI = .84, 95\%CI_{\Delta PC} = [-.10, -.01]$). The second test block was not significantly different from the control block for either condition with an average TI of .93 (New Location First: $t(36) = 1.36, p = .18, d = .21, TI = .93, 95\%CI_{\Delta PC} = [-.08, .02]$; New Eye First: $t(36) = .99, p = .33, d = .17, TI = .93, 95\%CI_{\Delta PC} = [-.07, .02]$). There was no significant difference between the untrained eye and untrained location for either condition

(New Location First: $t(36) = .86, p = .40, d = .13, TI = .97, 95\%CI_{\Delta PC} = [-.05, .03]$; New Eye First: $t(36) = 1.52, p = .14, d = .27, TI = 1.0, 95\%CI_{\Delta PC} = [-.08, 0.0]$).

The hypothesis that II learning includes a retinal specific component predicts that the observed accuracy drop in the II conditions when the stimulus moved to a new retinal location should be a general effect that holds for all categories and responses. To test this prediction, and to rule out category- or response-specific accounts of the accuracy drop (e.g., that the drop is caused by some sort of Simon-like effect[24]), we fit models derived from General Recognition Theory (GRT) to the 4×4 confusion matrices from each test block. GRT is a multidimensional generalization of signal detection theory[25,26,27]. It not only estimates decision bounds, but in addition it estimates category representations (under the assumption that these have multivariate normal distributions). The results showed that in the II conditions, all category representations became considerably noisier when the stimulus moved to a new retinal location, but not when it moved to a new eye, whereas in the RB conditions, the effects were still general (i.e., not category or response specific), but the increase in noise was similar for both experimental conditions. Thus, this analysis suggests that the performance changes that occurred when the stimulus moved locations were general, and not category or response specific.

The results showed significant retinal specificity for II category learning, but not for RB learning. Instead, the RB participants showed a block effect in which accuracy decreased in each successive test block regardless of retinal location. This is likely due to unique differences between RB and II learning. Considerable evidence suggests that in RB tasks, participants learn abstract category labels (e.g., A, B, C, and D) and that they use working memory to remember which response keys are associated with each label. All trials during the test blocks were unsupervised, so recovery from any failure to recall the correct category label-response key association was difficult or impossible. In contrast, in II tasks, participants learn response positions, not category labels, and therefore there is no need to learn category label-response key associations[28,15,29,30,31,16]. In fact, several RB participants complained about forgetting which response key was associated with each category during the transition between testing blocks but no II participants voiced this complaint.

Although we found no evidence of retinal-specific learning in the RB conditions of Experiment 1, it is possible that the reduction in accuracy across test blocks masked a small retinal-specific effect. Experiment 2 provides an alternative test of this hypothesis. Although Experiment 2 was performed before Experiment 1, we chose to present them in the reverse order since Experiment 1 has the conceptually simpler design.

As discussed in the Introduction, RB and II categorization tasks were chosen for this research because considerable

evidence suggests they are learned in qualitatively different ways – that is, RB learning depends primarily on declarative memory, whereas II learning depends on procedural memory. The hypothesis that motivated the present research is that category learning mediated by procedural systems is more likely to be retinal specific than category learning mediated by declarative systems. Therefore, the strongest test of this prediction requires experimental conditions that minimize any possible contribution of procedural learning in the RB task. For example, recent evidence suggests that procedural learning occurs in the background while participants are engaging in explicit rule learning[32].

For these reasons, Experiment 2 had a modified experimental paradigm that was selected to minimize the contribution of procedural learning. In particular, in the RB conditions of Experiment 2, feedback was delayed by 2.5 sec on every trial during training, and a noise mask filled the delay interval to disrupt visual imagery. Previous research has shown that these manipulations abolish almost all procedural learning but have no effect on RB learning[33]. This is because II learning is thought to depend on dopamine-mediated synaptic plasticity within the striatum, and the evidence is good that for this type of plasticity to succeed, the dopamine must arrive at the critical synapses within a few seconds of synaptic activity[34]. The mask is necessary to prevent participants from maintaining synaptic activity during the delay period via mental imagery.[32,35,36]. In addition, because procedural learning develops slowly, Experiment 2 included 200 fewer RB trials (for a total of 300 training trials for RB vs. 500 training trials for II). All other Experiment 2 procedures were identical to those used in Experiment 1.

As in Experiment 1, we excluded from further analysis all participants who performed so poorly that we could not reject the null hypothesis that their control test block accuracy was at chance (0.25; i.e., with $\alpha = .01$). This eliminated 21 of the 100 participants (12 from the RB condition and 9 from the II condition).

The learning curves during the training phase (Phase 1) for the II and RB categories are shown in the top panel of Figure IV. As in Experiment 1, note that training accuracy increased to roughly the same level for both category structures (RB and II accuracies during the control testing block were not significantly different: mean II accuracy 69%; mean RB accuracy 71%; $t(78) = .70, p = .49, d = .16, 95\%\text{CI}_{\Delta PC} = [-.08, .04]$).

As in Experiment 1, we again compared performance when the stimulus moved to the untrained eye versus a new location on the trained eye via a three-factor mixed ANOVA on the difference scores (i..e, test block 2 or 3 accuracy minus control block accuracy). The ANOVA included two levels of category structure (II and RB), two levels of retinal location (New Eye and Same Eye), and two levels of block number (2nd and 3rd test blocks), with repeated mea-

sures on the latter two factors. The results showed a significant main effect of retinal location [$F(1, 75) = 17.8, p < .001, \eta_p^2 = .19, 90\%\text{CI}_{\eta_p^2} = [.07, .31]$] but no significant effect of category structure [$F(1, 75) = .44, p = .51, \eta_p^2 = .006, 90\%\text{CI}_{\eta_p^2} = [0.0, .06]$] or block [$F(1, 75) = 1.7, p = .20, \eta_p^2 = .02, 90\%\text{CI}_{\eta_p^2} = [0.0, .10]$]. The only significant interaction was between category structure and retinal location [$F(1, 75) = 14.7, p < .001, \eta_p^2 = .16, 90\%\text{CI}_{\eta_p^2} = [.05, .29]$].

To probe these results more closely we again ran followup two-factor ANOVAs separately on the II and RB difference scores [retinal location (new eye and same eye) × block (2nd and 3rd) with repeated measures on both factors]. In the II conditions, there was a significant main effect of retinal location [$F(1, 39) = 32.5, p < .001, \eta_p^2 = .45, 90\%\text{CI}_{\eta_p^2} = [.25, .57]$], but the main effect of block [$F(1, 39) = 1.0, p = .32, \eta_p^2 = .03, 90\%\text{CI}_{\eta_p^2} = [0.0, .15]$] and the retinal location × block interaction [$F(1, 39) = .03, p = .87, \eta_p^2 = .001, 90\%\text{CI}_{\eta_p^2} = [0.0, .02]$] were both nonsignificant. In the RB conditions there was no significant effect of block [$F(1, 36) = .61, p = .44, \eta_p^2 = .02, 90\%\text{CI}_{\eta_p^2} = [0.0, .13]$], retinal location [$F(1, 36) = .03, p = .88, \eta_p^2 = .001, 90\%\text{CI}_{\eta_p^2} = [0.0, .05]$], or retinal location × block interaction [$F(1, 36) = .25, p = .62, \eta_p^2 = .01, 90\%\text{CI}_{\eta_p^2} = [0.0, .10]$].

The mean and individual participant accuracies in each test block are shown in the bottom two panels of Figure IV. As in Experiment 1, all pairwise differences were assessed for significance using two-sample paired t-tests. FDR corrected significance (at FDR $\leq$ .05) is denoted by ** and uncorrected significance (at $\alpha = .05$) is denoted by *.

As in Experiment 1, accuracy dropped significantly in the II condition when the stimulus moved to the untrained location, with an average TI of .72 (New Location First: $t(39) = 4.48, p = .00006, d = .99, TI = .69, 95\%\text{CI}_{\Delta PC} = [-.18, -.07]$; New Eye First: $t(39) = 3.79, p = .0005, d = .96, TI = .73, 95\%\text{CI}_{\Delta PC} = [-.16, -.07]$), but not when the stimulus was shown in the same location on the untrained eye (New Location First: $t(39) = .70, p = .49, d = .20, TI = 1.0, 95\%\text{CI}_{\Delta PC} = [-.07, .03]$; New Eye First: $t(39) = 1.34, p = .19, d = .45, TI = .94, 95\%\text{CI}_{\Delta PC} = [-.09, .01]$). Performance was also lower in the untrained location than the untrained eye, with an average TI of .78 (New Location First Condition: $t(39) = 3.78, p = .0005, d = .96, TI = .73, 95\%\text{CI}_{\Delta PC} = [-.15, -.06]$; New Eye First Condition: $t(39) = 2.45, p = .02, d = .75, TI = .82, 95\%\text{CI}_{\Delta PC} = [-.12, -.03]$), though after FDR correction, the performance difference was only significant in the New Location First Condition.

In contrast to Experiment 1, performance in the the RB condition did not decrease significantly across blocks. The accuracy decrease when the stimulus moved to the untrained eye was significant before FDR correction in the New Eye
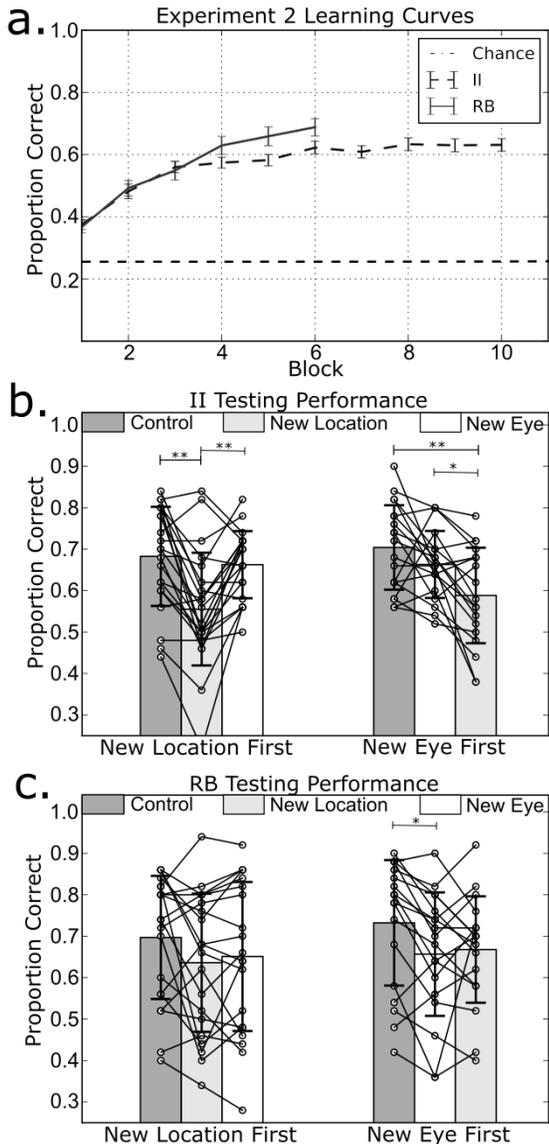
Figure 4. Results of Experiment 2. All error bars show standard error. (a) Learning curves of participants during RB ($n = 38$) and II ($n = 41$) training. Each block included 50 trials. (b & c) Individual participant and mean accuracies during each Experiment 2 test block for the II condition (b) and RB condition (c). Open circles denote individual participant accuracies and the bars denote group mean accuracy. Error bars show standard error. "**" signifies significance after FDR correction (FDR $\leq$ .05; largest significant p-value = .002). "*" signifies uncorrected significance at α = .05. All p-values can be found in the text following this figure.

First condition ($t(36) = 2.24, p = .03, d = .50, TI = .85, 95\%\text{CI}_{\Delta PC} = [-.12, -.04]$), but not in the New Location First condition)($t(36) = 1.44, p = .16, d = .28, TI = .88, 95\%\text{CI}_{\Delta PC} = [-.09, 0.0]$). When the stimulus moved to the untrained location, the accuracy decrease was nonsignificant in both conditions (New Location First: $t(36) = 1.91, p = .06, d = .39, TI = .89, 95\%\text{CI}_{\Delta PC} = [-.12, 0.0]$; New Eye First: $t(36) = 1.91, p = .06, d = .46, TI = .91, 95\%\text{CI}_{\Delta PC} = [-.13, .03]$). In addition, there was no significant difference between performance for the untrained location and the untrained eye (New Location First: $t(36) = .47, p = .64, d = .09, TI = 1.10, 95\%\text{CI}_{\Delta PC} = [-.05, .02]$; New Eye First: $t(36) = .33, p = .74, d = .08, TI = 1.0, 95\%\text{CI}_{\Delta PC} = [-.08, .06]$).

As in Experiment 1, we fit GRT models to the category-response confusion matrices from all test blocks. The results were qualitatively identical to Experiment 1 – that is, all changes in accuracy that occurred when the stimulus moved were general effects and not category or response specific.

The absence of a consistent block effect in the Experiment 2 RB conditions suggests that the RB block effect that occurred in Experiment 1 probably did not mask any significant retinal-specific learning. Furthermore, the similarity of the Experiment 1 and 2 RB results suggests that procedural learning probably had little or no effect on performance in the RB conditions. The most important conclusion of Experiment 2, however, is that it showed the same major results as Experiment 1.

The results of the two experiments together provide strong support for the hypothesis that categorization is retinal specific when the learning is procedural, but not when the learning is explicit. This is important because it runs counter to prevailing historical views of category learning as a high-level cognitive process (e.g., as in prototype[5] and exemplar theories[8]) that depends only on abstract mental representations that are far removed from the basic-level visual processing needed to perceive the stimulus. Additionally, our finding that II but not RB learning includes a retinal-specific component adds to the list of 25+ empirical dissociations between RB and II categorization that have been previously reported[37].

Why should II category learning be retinal specific, but not RB learning? One possible account of this difference is rooted in the COVIS theory of category learning[14,37,38]. COVIS proposes that success in RB tasks is mediated by the learning of explicit rules, which depends principally on prefrontal cortex (PFC; and to a lesser extent on other regions including anterior cingulate, the head of the caudate nucleus, and the hippocampus). PFC receives visual input from inferotemporal cortex and other high-level visual areas that contain neurons with large receptive fields covering both visual hemispheres, so COVIS predicts that when the stimulus moves to an untrained retinal location, the same high-level

visual neurons are excited and RB categorization accuracy remains high.

In contrast, COVIS assumes that success in II tasks depends on procedural learning that is primarily mediated by synaptic plasticity at cortical-striatal synapses within the dorsal striatum. The dorsal striatum receives direct visual input from all areas of visual cortex except V1[39]. Previous work[40] has reported that V2 was the only visual cortical area showing significant task-related activity in an fMRI study of II category learning that used the same sine-wave gratings that were used here. V2 neurons are characterized by small receptive fields that do not cross the fovea[41]. Therefore, according to this account, moving the stimulus to an untrained location on the retina excites different V2 neurons that project onto medium spiny neurons in the dorsal striatum via untrained synapses, causing impaired categorization performance. Moving the stimulus to the untrained eye, on the other hand, still excites the same V2 neurons because most V2 neurons are binocular[42], so the learning transfers to the untrained eye.

In other words, COVIS predicts that, although the RB and II tasks include the same stimuli, the category (or motor) learning that occurs in those two tasks depends primarily on input from different visual areas – more specifically, that learning in RB tasks uses visual representations from higher-level visual areas than learning in II tasks.

The COVIS account of the differences between the II and RB results raises the interesting question of whether the retinal specificity we found is a general property of II learning or whether it is restricted to II tasks that use simple, primitive stimuli like the sine-wave gratings used here. For example, what would happen in an II task that used human faces as stimuli? Presumably, categorization responses would be based on perceptual representations from the fusiform face area of inferotemporal cortex (which projects to the body of the caudate nucleus). If so, then COVIS predicts that II learning should be retinal invariant since inferotemporal cortical neurons exhibit large receptive fields.

Traditional findings of retinal-specificity have been limited to perceptual learning. Perceptual learning is often studied by examining the ability to differentiate between stimulus attributes such as contrast[43], orientation[44], and stimulus offset[45], or to recognize a target in a target-distractor paradigm[46]. Unlike categorization tasks where observers classify many different stimuli into categories, perceptual learning tasks typically include only two or few stimuli that need to be classified or identified. Learning in such studies is often found to be retinal specific[46,47], sometimes even eye specific[48], and can occur with or without feedback[48,49]. These findings overlap partially with ours. We found that II learning was retinal specific, but not eye specific, and many previous studies have reported that II learning requires immediate feedback[35,36]. In contrast, our results suggest that RB learning is neither retinal nor eye specific, and previous studies have found that at least some RB learning is possible in the absence of any feedback[50].

The overlapping but distinct profiles of category and perceptual learning suggest that there might be some common underlying mechanisms at work, especially in the case of II category learning. This hypothesis suggests many interesting new avenues of research. For example, several perceptual-learning studies have reported that retinal specificity during orientation discrimination can be eliminated via simultaneous training on an irrelevant task in the non-trained peripheral location[51,52]. An important future study is to see whether this manipulation is equally effective in II category learning. Similarly, the perceptual-learning literature has recently been exploring the hypothesis that perceptual learning can be accounted for by reinforcement-learning or rule-based mechanisms that are similar to those thought to mediate II and RB category learning[53,54,52]. This work suggests that an exciting theoretical challenge might be to account for retinal-specific II learning and traditional perceptual-learning phenomena with a similar underlying mechanistic model.

Overall, our findings demonstrate the importance of considering visual field location during categorization training and highlight another experimental factor to consider when designing virtual reality training paradigms for search and categorization tasks. Additional work on the role of categorization during visual search will likely cast further light onto this subject and potentially suggest new training methods to remove or minimize the impact of visual field dependence on the important category judgments we make each day. Finally, these findings close the gap between category learning and perceptual learning – demonstrating several similarities and differences that can be leveraged to better explore the neural basis of both phenomena and providing motivation for both fields to learn from each another.

Methods

Participants

Two-hundred students (100 for Experiment 1 and 100 for Experiment 2) at the University of California, Santa Barbara participated in a one-hour experiment in exchange for course credit. Experiment 2 was performed before Experiment 1, but is presented second in the manuscript for better readability. All participants had normal (20/20) or corrected-to-normal vision using contact lenses. Fifty participants performed the II task and 50 performed the RB task in each experiment. All relevant ethical regulations were followed and the study protocol was approved by the Human Subjects Committee at UCSB. Informed consent was obtained from all participants, and every participant was allowed to quit the experiment at any time for any reason and still receive credit.

Stimuli and Categories

The stimuli and categories were the same for both experiments. The stimuli were circular sine-wave gratings presented on 21-inch monitors (1280 × 1024 resolution). All stimuli had the same size, shape and contrast, and differed only in bar width (as measured by cycles per degree of visual angle or cpd) and bar orientation (measured in degrees counterclockwise rotation from horizontal). The stimuli from the II conditions were generated first. Each category was defined by a bivariate normal distribution with means for Bar Width and Orientation as follows: Category A (30, 50), B (50, 30), C (50, 70), and D (70, 50). The variance on both stimulus dimensions was set to 50 in all four distributions and all covariances were set to 0. Thus, the distributions differed only in their means.

The stimuli defining the A, B, C, and D categories were generated as follows: 1) 300 random samples were drawn from each of the bivariate normal distributions that defined the categories; 2) the samples were linearly transformed so that the sample statistics (means, variances, covariances) exactly matched the population parameters that defined the distributions; 3) each resulting sample value, denoted by the ordered pair (x_1, x_2), was used to generate a stimulus with bar width equal to $x_1^* = \frac{x_1}{30} + 0.25$ cpd and bar orientation equal to $x_2^* = \frac{\pi}{200}x_2 + \frac{\pi}{9}$ degrees counterclockwise rotation from horizontal.

The RB categories were created by rotating the II stimuli 45 degrees clockwise about the point (50,50).

Eye Tracking

Eye tracking (at 250 HZ) was performed using an Eyelink I eye tracker by SR Research Ltd. Participants used a chin rest to minimize movement error, and at the beginning of each session 9-point calibration was performed. Additional recalibration was performed before each test block and during the training as needed (i.e. when participants were reported as breaking fixation but reported that they were fixating).

Procedures

At the start of the experiment, all participants were told that they would be shown disk-like images, each of which belonged to one of four categories (A, B, C, or D), and that their task was to use the feedback during training to learn to assign each disk to its correct category. The instructions were identical for RB and II participants. All conditions included four phases, which are described in Figure II in the case of the New Eye First conditions. The New Location First conditions were identical except the ordering of the final two phases was reversed. In Phase 1, participants wore an eye patch over their left eye and were trained with trial-by-trial feedback. Training was split into blocks of 50 trials with a brief (less than 20 seconds) rest period after each block where participants could rest without removing their head from the chin-rest. Each rest period ended after the participant pressed a key or waited 20 seconds.

For Experiment 1, all participants received 500 training trials (10 training blocks) followed by 150 testing trials (3 testing blocks). For Experiment 2, participants in the II condition received 500 training trials while participants in the RB condition received 300 training trials.

On every trial the stimulus was presented for 150ms at a location 5° right of fixation. If the participant's eyes moved more than 1° away from the center fixation cross or a saccade was detected (using saccade velocity thresholds of 35 deg/sec and 9000 deg/s^2 for velocity and acceleration respectively) immediately before or during stimulus presentation, the trial was aborted. This resulted in on average fewer than 50 aborted trials per participant.

After the participant responded, feedback was provided in the form of a green correct or red incorrect label displayed in the center of the screen for 1 second. In Experiment 1, the next stimulus was shown automatically after the feedback disappeared and as soon as the participant re-fixated their gaze at the center of the screen. In Experiment 2, the participant initiated stimulus presentation by pressing the space bar anytime they wanted after the feedback disappears, so long as they were fixating the center of the screen.

Phases 2 – 4 each included 50 trials of categorization with no trial-by-trial feedback. Phase 2 was a control that was identical to training except for the omission of feedback. In the New Eye First conditions, the stimulus conditions during Phase 3 were identical to Phases 1 and 2. However, prior to the first Phase 3 trial, the participant moved the eye patch to cover the right eye. So Phase 3 tested whether the categorization knowledge acquired during training was specific to the right eye. Prior to the start of Phase 4, the participant moved the eye patch back to the left eye. This phase was identical to the Control (Phase 2) except the stimulus was presented 5° left of fixation. So this phase tested whether category knowledge transferred to an untrained retinal location in the trained eye. The New Location First conditions were identical except the ordering of Phases 3 and 4 was reversed.

Code Availability

All code used in this study is available from the corresponding author upon request.

Data Availability

All data from this study are available from the corresponding author upon request.

Conflict of Interests

The authors declare no conflict of interest

Contributions

L.R. and F.G.A conceived and designed the experiment with input from M.E. L.R. managed data collection and analyzed the data. L.R., M.E., and F.G.A. wrote the paper. All authors approved the final draft of the manuscript.

Corresponding Author

All correspondence should be directed to F. Gregory Ashby.

References

Homa, D., Sterling, S. & Trepel, L. Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory* **7**, 418–439 (1981).

Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. *Journal of experimental psychology* **77**, 353–363 (1968).

Reed, S. K. Pattern recognition and categorization. *Cognitive psychology* **3**, 382–407 (1972).

Rosch, E. H. Natural categories. *Cognitive psychology* **4**, 328–350 (1973).

Smith, J. D. & Minda, J. P. Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **24**, 1411–1430 (1998).

Estes, W. K. Array models for category learning. *Cognitive Psychology* **18**, 500–549 (1986).

Medin, D. L. & Schaffer, M. M. Context theory of classification learning. *Psychological Review* **85**, 207–238 (1978).

Nosofsky, R. M. Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General* **115**, 39–57 (1986).

Ashby, F. G. & Maddox, W. T. Human category learning. *Annual Review of Psychology* **56**, 149–178 (2005).

Ashby, F. G. & Maddox, W. T. Human category learning 2.0. *Annals of the New York Academy of Sciences* **1224**, 147–161 (2010).

Poldrack, R. A. *et al.* Interactive memory systems in the human brain. *Nature* **414**, 546–550 (2001).

Poldrack, R. A. & Packard, M. G. Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* **41**, 245–251 (2003).

Squire, L. R. Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory* **82**, 171–177 (2004).

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. & Waldron, E. M. A neuropsychological theory of multiple systems in category learning. *Psychological Review* **105**, 442–481 (1998).

Maddox, W. T., Ashby, F. G., Ing, A. D. & Pickering, A. D. Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition* **32**, 582–591 (2004).

Waldron, E. M. & Ashby, F. G. The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review* **8**, 168–176 (2001).

Zeithamova, D. & Maddox, W. T. Dual-task interference in perceptual category learning. *Memory & Cognition* **34**, 387–398 (2006).

Ashby, F. G. & Gott, R. E. Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**, 33–53 (1988).

Ashby, F. G. & Ennis, J. M. The role of the basal ganglia in category learning. *Psychology of Learning and Motivation* **46**, 1–36 (2006).

Filoteo, J. V., Maddox, W. T., Salmon, D. P. & Song, D. D. Information-integration category learning in patients with striatal dysfunction. *Neuropsychology* **19**, 212–222 (2005).

Knowlton, B. J., Mangels, J. A. & Squire, L. R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).

Nomura, E. *et al.* Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex* **17**, 37–43 (2007).

Jeter, P. E., Dosher, B. A., Petrov, A. & Lu, Z.-L. Task precision at transfer determines specificity of perceptual learning. *Journal of vision* **9**, 1–1 (2009).

Simon, J. R., Sly, P. E. & Vilapakkam, S. Effect of compatibility of sr mapping on reactions toward the stimulus source. *Acta psychologica* **47**, 63–81 (1981).

Ashby, F. G. & Townsend, J. T. Varieties of perceptual independence. *Psychological Review* **93**, 154–179 (1986).

Ashby, F. G. & Soto, F. A. Multidimensional signal detection theory. In Busemeyer, J. R., Townsend, J. T., Wang, A. & Eidels, A. (eds.) *The oxford handbook of computational and mathematical psychology*, 13–34 (New York: Oxford University Press, 2015).

Soto, F. A., Zheng, E., Fonseca, J. & Ashby, F. G. Testing separability and independence of perceptual dimensions with general recognition theory: A tutorial and new r package (grtools). *Frontiers in psychology* **8**, 696 (2017).

Ashby, F. G., Ell, S. W. & Waldron, E. M. Procedural learning in perceptual categorization. *Memory & Cognition* **31**, 1114–1125 (2003).

Maddox, W. T., Bohil, C. J. & Ing, A. D. Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review* **11**, 945–952 (2004).

Maddox, W. T., Glass, B. D., O'Brien, J. B., Filoteo, J. V. & Ashby, F. G. Category label and response location shifts in category learning. *Psychological Research* **74**, 219–236 (2010).

Spiering, B. J. & Ashby, F. G. Response processes in information–integration category learning. *Neurobiology of Learning and Memory* **90**, 330–338 (2008).

Crossley, M. J. & Ashby, F. G. Procedural learning during declarative control. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41**, 1388–1403 (2015).

Maddox, W. T., Ashby, F. G. & Bohil, C. J. Delayed feedback effects on rule-based and information-integration category learning. *Journal of experimental psychology: learning, memory, and cognition* **29**, 650 (2003).

Yagishita, S. *et al.* A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* **345**, 1616–1620 (2014).

Maddox, W. T., Ashby, F. G. & Bohil, C. J. Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29**, 650–662 (2003).

Maddox, W. T. & Ing, A. D. Delayed feedback disrupts the procedural-learning system but not the hypothesis testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **31**, 100–107 (2005).

Ashby, F. G. & Valentin, V. V. Multiple systems of perceptual category learning: Theory and cognitive tests. In Cohen, H. & Lefebvre, C. (eds.) *Handbook of categorization in cognitive science, Second Edition*, 157–188 (New York: Elsevier, 2017).

Cantwell, G., Crossley, M. J. & Ashby, F. G. Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin & Review* **22**, 1598–1613 (2015).

Heimer, L. *The human brain and spinal cord: functional neuroanatomy and dissection guide* (Springer Science & Business Media, 2012).

Waldschmidt, J. G. & Ashby, F. G. Cortical and striatal contributions to automaticity in information-integration categorization. *Neuroimage* **56**, 1791–1802 (2011).

Gattass, R., Gross, C. & Sandell, J. Visual topography of v2 in the macaque. *Journal of Comparative Neurology* **201**, 519–539 (1981).

Hubel, D. H., Wiesel, T. N., Yeagle, E. M., Lafer-Sousa, R. & Conway, B. R. Binocular stereoscopy in visual areas v-2, v-3, and v-3a of the macaque monkey. *Cerebral Cortex* **25**, 959–971 (2013).

Yu, C., Klein, S. A. & Levi, D. M. Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision* **4**, 4–4 (2004).

Jeter, P. E., Dosher, B. A., Liu, S.-H. & Lu, Z.-L. Specificity of perceptual learning increases with increased training. *Vision research* **50**, 1928–1940 (2010).

Hung, S.-C. & Seitz, A. R. Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *Journal of Neuroscience* **34**, 8423–8431 (2014).

Dill, M. & Fahle, M. The role of visual field position in pattern–discrimination learning. *Proceedings of the Royal Society of London B: Biological Sciences* **264**, 1031–1036 (1997).

Nazir, T. A. & O'Regan, J. K. Some results on translation invariance in the human visual system. *Spatial vision* **5**, 81–100 (1990).

Fahle, M., Edelman, S. & Poggio, T. Fast perceptual learning in hyperacuity. *Vision research* **35**, 3003–3013 (1995).

Petrov, A. A., Dosher, B. A. & Lu, Z.-L. Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision research* **46**, 3177–3197 (2006).

Ashby, F. G., Queller, S. & Berretty, P. M. On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics* **61**, 1178–1199 (1999).

Zhang, T., Xiao, L.-Q., Klein, S. A., Levi, D. M. & Yu, C. Decoupling location specificity from perceptual learning of orientation discrimination. *Vision research* **50**, 368–374 (2010).

Zhang, J.-Y. *et al.* Rule-based learning explains visual perceptual learning and its specificity and transfer. *Journal of Neuroscience* **30**, 12323–12328 (2010).

Kahnt, T., Grueschow, M., Speck, O. & Haynes, J.-D. Perceptual learning and decision-making in human medial frontal cortex. *Neuron* **70**, 549–559 (2011).

Kumano, H. & Uka, T. Neuronal mechanisms of visual perceptual learning. *Behavioural brain research* **249**, 75–80 (2013).