

ColorSep

Avi Chanales¹

¹New York University (NYU)

March 15, 2019

Abstract

Intro

Our memories are not veridical snapshots of the past, rather, our memories are stored adaptively in order to best guide future behavior. One example in which experiences are stored adaptively in memory is when two experiences are very similar and easily confusable. Research has shown that in these instances, the hippocampus serves to distort the representations of the similar experiences by separating them into distinct neural patterns in order to minimize the interference between them (**REFS**). Recently, a number of studies (**Chanales 2017, Favila 2016, Hulbert 2014, Schlichting 2015, Kim 2017**) have provided evidence that over extended learning the hippocampus can actually repel the representations of similar experiences apart to the point where two similar events are coded more distinctly than two unrelated events. In this way the hippocampal code exaggerates the differences between overlapping events. Although this hippocampal repulsion has been shown to benefit future learning by reducing the confusability of the corresponding memories (**Favila 2016**), it not known if this repulsion that occurs at the neural level also occurs at the featural level of the corresponding memories. One intriguing possibility is that competition between similar levels can also repel the feature values of those memories apart such that those memories will be remembered as being more different from each other than they actually were.

The theoretical and empirical research conducted on the repulsion of hippocampal patterns suggest some critical factors underlying the emergence of this effect. First, the divergence of hippocampal activity patterns is a direct response to competition between the overlap in the events' neural patterns. The hippocampus resolves this competition by distancing the competing representations apart from one another. Therefore, the extent of repulsion should scale with the amount of competition between the similar events at the start of learning. Second, the divergence of the hippocampal patterns occurs over learning. As similar events continue to compete, the hippocampus will continue to distance their representations apart. Thus, a complete repulsion of the patterns will only be observed after repeated experience with the similar events.

Here, across multiple behavioral studies, we tested the idea that repeated encoding of highly similar stimuli would yield a similar repulsion of feature values in memory. We designed a novel paradigm in which participants performed an associative learning task that required them to discriminate between pairs of objects that were identical except for their color values. Critically, we adjusted the similarity of the object pairs by varying the difference in hue angle between the paired objects. After learning we probed participants memory for the color of each object using a continuous scale (REF). Assessing memory for the objects in this manner allowed us to measure whether color estimates were biased towards or away from the competing object's color.

This design allowed to test several predictions of the repulsion account. First, if repulsion also occurs at the featural level of memories, we should observe systematic distortions in color memory such that color estimates for each object would be biased away from the competing object’s color. Second, if the repulsion of the feature values of memory is a response to competition, then maximal repulsion in color memory should occur for objects with highly similar color values. Lastly, we predicted that these repulsion biases should be adaptive to memory performance by decreasing the confusability of between the similar objects. Thus the the degree of repulsion between objects should be associated with decreased interference between them during a memory test.

Results

Similarity between items leads to a repulsion of their feature values in memory

In Experiment 1 participants completed an associative learning task in which they studied 36 pairs of colored objects and faces. To create competition between the pairs, the 36 objects each contained a competing object that was identical to its competitor aside from its color value (Figure 2.1A). To modulate the level of competition, we varied the difference in hue angle of the color between the competing objects across three conditions of color similarity: high similarity (24° difference), moderate similarity (48° difference), and low similarity (72° difference; Figure 1A). The object-face pairs were learned over 8 rounds of study and test. During a study round each colored object was presented with its associated face. Importantly, the two competing objects were never presented simultaneously, thus discrimination between them was entirely memory based. After each study round participants completed two tests: (1) a color memory test during which participants reported the color of each object using a color wheel, and (2) an object-face associative memory test during which subjects selected the face that corresponded to each object. On each test trial, the subject was presented with an object and four face choices: the face paired with object (‘target’), the face paired with the competing object (‘competitor’), and two faces paired with different objects (‘non-competitor’). Immediately after all learning rounds participants completed an additional color memory test (post-test) and then returned, after a 24h delay, for a final color memory test (Day 2 post-test) (Figure 2.1B).

We first turned to the associative memory test data to test that subjects successfully learned to discriminate between the similar object colors. Subjects learned to choose the target face over the learning rounds ($F(1,22) = 435.4$, $p < 0.0000001$; Figure 2.2A) such that by the last round of learning they reliably chose the target faces over the competitor faces in all similarity conditions ($ps < 0.0000001$). However, confirming that our similarity modulation influenced the discrimination difficulty between the colors, target hit rates across learning varied by similarity condition ($F(2,44) = 13.04$, $p = 0.00003$). Average target selection rates were lower in the high similarity condition compared to both the moderate similarity ($t(22) = -3.98$, $p = 0.0006$) and low similarity conditions ($t(22) = -4.65$, $p = 0.0001$). Target selection rates did not differ between the moderate and low similarity condition ($t(22) = -0.72$, $p = 0.48$). This confirmed that increasing the similarity between colors increased the competition (interference) between them, particularly for the high similarity pairs. Turning toward the color memory test, color error for the objects steadily decreased over learning ($F(1,22) = 166.2$, $p < 0.0000001$; Figure 2.2B) such that by the last round of learning there was no difference in color error between similarity condition ($F(2,44) = 1.22$, $p = 0.31$). Thus participants were able to learn each object’s associated color across all conditions.

Next we assessed if and how color memory for the objects was distorted over learning. Specifically, we

were interested in whether there was a bias in color memory for each object relative to its competing color. We quantified this color bias by calculating the percentage of trials in which the color estimate fell away from the competing color. Any deviations in the measure from 50% would reflect a bias in color memory with values greater than 50% reflecting a repulsion away from the competing color and values below 50% reflecting an attraction towards the competing color. We calculated this bias measure separately for each round and condition to see how the biases in color memory changes with experience (Figure 2C). There were significant changes in the color bias measures over learning across all the conditions ($F(1,22) = 20.34$, $p = 0.0002$). At the start of learning color estimates tended to be biased towards the competing color while towards the end of learning they were biased away from the competing color. Thus, there were dynamic shifts in the internal color representations of the objects over learning. However, to better quantify the bias in these representations and how they differed by similarity condition we turned to the post-test which was designed to have more trials and therefore more statistical power to detect condition differences.

First we assessed the absolute color error for the objects in the post-test. Color error in the post-test varied across similarity conditions ($F(2,44) = 4.03$, $p = 0.025$). Color error was higher for low similarity objects compared to color error in the moderate similarity ($t(22) = 2.44$, $p = 0.023$) and high similarity ($t(22) = 2.13$, $p = 0.044$) conditions. Color error did not differ between the moderate and high similarity conditions ($t(22) = 0.18$, $p = 0.86$). However, we were most interested in whether there was a bias in color errors relative to the competing color. Critically, color bias highly varied by similarity condition ($F(2,44) = 10.11$, $p = 0.0002$). While there was a strong repulsion bias away from the competing color in the high similarity condition ($t(22) = 4.31$, $p = 0.0003$), there was no bias in either the moderate similarity ($t(22) = -0.67$, $p = 0.51$) or low similarity ($t(22) = -0.03$, $p = 0.97$) conditions. Thus participants systematically exaggerated the color of the object away from the competing color when competition between the colors was highest. This repulsion effect persisted after a 24 hour delay. During the Day 2 post-test, color bias still varied by similarity condition ($F(2,42) = 9.82$, $p = 0.0003$) with a selective repulsion bias in the high similarity condition ($t(22) = 4.31$, $p = 0.0002$; moderate and low similarity $ps > 0.51$).

%%%%%%%%OLD%%%%%%%%%

Next we turned to the critical post test to assess if and how color memory for the objects was distorted over learning. After completing all 8 rounds of study, color error in the post-test marginally varied across similarity conditions ($p = 0.073$). Color error was lower for moderate similarity objects compared to low similarity ($p = 0.037$) and high similarity color error did not differ from either moderate ($p = 0.62$) or low similarity ($p = 0.13$) color error. However, we were most interested in whether there was a bias in color errors relative to the competing color. We quantified the bias in color estimates two ways. First, we signed the distance of each error depending on whether the color estimate was positioned *towards* the competing color (positive error) or *away from* the competing color (negative error). Thus, any deviation in this measure from zero would indicate a bias in color memory relative to the competing color with positive values reflecting an attraction towards and negative values reflecting a repulsion away from the competing color. Second, we calculated the percentage of trials in which the estimate fell away from the competing color. Similarly, any deviations in the measure from 50% would reflect a bias in color memory. While signing the error allowed for an estimate of the effect size (in degrees) of the distortion of the competing colors in color space, the percent error estimate was more robust to outliers in color memory as this measure was less influenced by extreme values in color error. Critically, both measures of color bias scores highly varied by similarity condition (percent away: $p = 0.0002$, signed error: $p = 0.0004$). While there was a strong color bias away from the competing color in the high similarity condition (percent away: signed error: $p = 0.002$), there was no bias in either the moderate (percent away: $p = 0.33$, signed error: $p = 0.22$) or low similarity (percent away: $p = 0.92$, signed error: $p = 0.20$) conditions. Thus participants systematically exaggerated the color of the object away from the competing color when competition between the colors was highest. This effect persisted after a 24 hour delay. Color bias still varied by similarity condition (STAT) with a selective repulsion bias in the high similarity condition on the Day2 post-test.

%%%%%%%%%

Next we tested the extent to which this repulsion bias in color estimates was adaptive to memory performance. Specifically we asked if the degree of exaggeration between the competing colors in the high similarity condition was related to improved discrimination performance on the association test. To test this we correlated each subject’s repulsion bias during the post test with their average target selection rate in the last 3 rounds of the associative memory test (Figure 2.2E). In support of an adaptive benefit of the repulsion, we found a positive correlation between the repulsion bias in the post-test and performance at discriminating between the objects during the association test ($r = 0.62$, $p = 0.002$). Thus subjects who showed greater repulsion of the similar colors in memory demonstrated less interference between those corresponding objects. We next wanted to rule out the possibility that repulsion of the colors didn’t improve participants discrimination performance over learning but rather the subjects that initially perceived the competing colors as being more different due to perceptual noise were the ones who exhibited better discrimination performance. If difference in perception explain variability in discrimination performance then we would expect the repulsion bias to predict discrimination performance at the start of learning. However, we found no correlation between the repulsion bias and discrimination performance in the association test averaged across the first three rounds of learning ($r = 0.23$, $p = 0.29$).

further confirming that the learned distortion in colorspace explained the improvement in discrimination performance.

The limits of repulsion

In Experiment 1 we found maximal repulsion in the condition with the greatest similarity between colors (24°). However, the degree of repulsion might not monotonically increase as similarity between the colors increases. If similarity between the colors becomes too great there might be a point where repulsion dissipates. To test whether there is a limit to repulsion we ran another experiment in which we introduced an ultra similarity condition in which the colors of the object pairs were only separated by 6° (Figure 3A). In addition to the ultra similarity condition we kept the high and moderate similarity conditions as comparison points. This also gave us another opportunity to replicate the repulsion bias in the high similarity condition found in Experiment 1.

In the associative memory test subjects learned to choose the target face over the learning rounds ($F(1,37) = 326.9$, $p < 0.0000001$; Figure 2.3A) such that by the end of learning they reliably chose the target faces over the competitor faces in all similarity conditions ($ps < 0.0000001$). However, as in Experiment 1, discrimination difficulty between the colors varied by similarity condition ($F(2,74) = 129.9$, $p < 0.00000001$). Average target selection rates were lower in the ultra similarity condition compared to both the high similarity ($t(37) = -11.39$, $p < 0.00000001$) and moderate similarity conditions ($t(37) = -16.26$, $p < 0.00000001$). And average target selection rate was lower in the high similarity condition compared to the moderate similarity condition ($t(37) = -4.34$, $p = 0.0001$). Thus, increasing the similarity between colors increased the discrimination difficulty between them. Turning toward the color memory test, color error for the objects steadily decreased over learning across all conditions ($F(1,37) = 186.5$, $p < 0.0000001$; Figure 2B) suggesting participants were able to learn each object’s associated color.

COLOR REPULSION PARAGRAPH

local maximum at 24,

24 degrees seem to be relative sweet spot

Feature repulsion depends on task demands

Next we asked whether the repulsion of similar color values depends on the task demand to separate the objects or if it is a byproduct of any associative learning. In the previous experiments, the associative memory task required participants to distinguish between the two similar colored objects. Here, we ran an experiment that was identical to Experiment 1, except we changed the task during learning so that optimal performance required integrating across the overlapping face-object pairs (Figure 2.4A). Specifically, we introduced an inference test in which one of the learned faces appeared and participants had to select the face (from a set of 4 faces) that was paired with the same object. With this subtle task change, discriminating the color of the objects was no longer relevant to performance on the task. Rather, optimal performance required linking between the two faces in memory via their shared association with the common object (irrespective of its color). We then tested how this change in task demand - from separating the colors to generalizing across them - would affect the repulsion bias. Would the bias remain, diminish, or possibly flip resulting in an attraction bias reflecting a merging of the color values?

Participants successfully learned to infer the face pairings as their selection of target face increased over learning rounds ($F(1,25) = 225.3$, $p < 0.00000001$; Figure 2.4B) and by the end of learning they were above chance performance in all conditions (p s < 0.000001). The similarity of the colors did not impact generalization across the object-face pairs as average inference performance across rounds did not vary by similarity condition ($F(2,50) = 0.52$, $p = 0.67$). Despite not being explicitly tested on their ability to discriminate between the competing colors, subjects' color error decreased over learning ($F(1,25) = 101$, $p < 0.0000001$; Figure 2.4C).

After 8 rounds of learning color error varied by similarity condition ($F(2,50) = 6.34$, $p = 0.003$). Follow up t-tests revealed that there were larger color errors for objects in the low similarity condition compared to the both the high similarity ($t(25) = -2.90$, $p = 0.008$) and the moderate similarity conditions ($t(25) = -2.55$, $p = 0.017$). However, color error did not differ between the high and moderate similarity conditions ($t(25) = 1.48$, $p = 0.15$). Turning to our main question of interest we asked if the change in task demand altered the bias in color estimates. As in Experiment 1 color bias varied by similarity condition ($F(2,50) = 5.18$, $p = 0.009$), however, in this experiment we found no color bias in the high similarity condition ($t(25) = -0.48$, $p = 0.63$). Surprisingly, both the moderate similarity and high similarity conditions showed positive bias *towards* the competing color (moderate: $t(25) = -2.09$, $p = 0.047$; low: $t(25) = -5.15$, $p = 0.00003$). Directly comparing the bias measures across Experiment 1 and Experiment 3 revealed a significant difference in color bias for the high similarity ($t(47.42) = 3.52$, $p = 0.001$) and low similarity conditions ($t(46.70) = 3.54$, $p = 0.0009$) conditions but not in the moderate similarity condition ($t(47.31) = 1.61$, $p = 0.11$). Thus, changing the task demand to encourage integration across pairs eliminated the repulsion bias in the high similarity condition and induced an attraction bias in the low similarity condition.

Discussion

Summary

While there have been numerous studies demonstrating that learning reduces interference between similar memories, to date, there have been no studies examining if and how the features of those corresponding memories change over learning. Using a novel behavioral paradigm that assessed feature memory on a continuous

scale, we show that learning induced a repulsion bias in color memory between competing memories such that the colors were remembered as being more different from each other than their true colors. Importantly, this repulsion bias developed over learning and was competition dependent as it was only observed for colors that were highly similar (24 degrees apart). Furthermore, the repulsion of competing colors was adaptive to memory performance as greater repulsion between the colors was associated with reduced interference between them. These findings provide insight into how memory representations get distorted over learning to reduce interference between them.

Our study design in that objects were paired with we were not interested in just whether participants could remember the pairings but how they remembered them.

This research contributes to our knowledge of how competitive retrieval shapes memories. Competition based frameworks of memory retrieval propose that retrieving a target memory employs inhibitory mechanisms that acts to suppress any competing traces that might interfere with successful recollection (Anderson 2003, Levy 2008). However, this inhibition has critical and lasting consequences on the non-target memories it suppresses. Behavioral research has demonstrated that the act of retrieval of a target memory makes it less likely for the competing, non-target memories to be subsequently remembered (Anderson RIF). A recent neuroimaging study confirmed that multiple retrieval attempts of a target memory suppressed the cortical representation of competing memories (Wimber). Intriguingly, in paradigms where participants receive an opportunity to restudy the previous competing items, memory for those items improves relative to items that were studied the same number of times but had not been the competing item during retrieval (Storm 2008, Hulbert 2014). Thus, it has been shown that repeated testing and restudy opportunities of similar events can quantitatively facilitate their future retrieval. The current set of studies adds to this research by demonstrating how competitive learning can qualitatively change those memories. Here, we found that learning exaggerated the features (color) of competing memories apart and furthermore that this exaggeration reduces the interference between them. This suggests that the inhibition processes engaged during competitive retrieval might not fully suppress the entire competing memory trace, but rather selectively inhibit the shared features of the memories (Norman papers) .

The repulsion bias observed for the features of memory, strongly parallels recent fMRI findings that initial overlap in hippocampal patterns triggers of repulsion of their patterns with learning. Prior studies have found that the hippocampus can come to exaggerate the differences between similar memories by coding them more distinctly from each other than to unrelated memories. A key component in all these studies is that this effect emerged after extensive learning and experience with the stimuli. Indeed, studies that directly measured the change in hippocampal patterns with learning found that similar memories initially had overlapping activity patterns within the hippocampus but after learning those patterns were repelled away from each other (Chanales). This accrual of repelled representations over learning parallels the emergence of repelled color memories in this study. Whereas initial color estimates tended to be biased towards the competitor, after learning color estimates were biased away from the competing color. Furthermore, fMRI studies have demonstrated that repulsion occurs when competition between memories is maximal. Similarly, in this study, the repulsion bias in color memory occurred in the high similarity condition in which the competing colors were very similar. when the initial colors were very similar. and the fMRI Future work will need to test if the two results are related.

An intriguing result from these studies is that the repulsion bias depends on the task demand to separate. When the task in between study rounds was changed to encourage integration across the face-object pairs, the repulsion bias in the high similarity condition disappeared. This suggests that top-down signals interact with the mnemonic processes engaged during learning to shape the organization of information in memory storage. This corroborates research suggesting that memory representations flexibly store the features most

relevant to current goals (Mack, Love). Thus distancing the features of memory should only be adaptive in as much as discrimination between the memories is beneficial for future behavior. Although we demonstrated that the demand to separate is critical to mechanistically repel the features of memories, it is unclear at what stage of processing this demand contributed to the repulsion. It could be the goal to separate altered the attentional weighting afforded to the diagnostic feature relevant for separation (i.e. the color of the objects) which in turn enhanced its encoding (Aly, Mack). Or possibly the task demand influenced downstream processes known to influence the degree of separation such as the strength of reactivation of the competing memory (Norman, Kim) or relatedly, the level of inhibition employed during retrieval (Anderson). Future work will need to further investigate how the top down goals dictated by the task demands contribute to the repulsion effect.

An important question this research raises

Lastly, this research contributes to a growing body of research demonstrating that repulsion of similar information is a common coding scheme employed by the mind and brain. While this is the first evidence of a repulsion bias in the featural content of long term memory, similar biases have been found at many levels of processing from perceptual representations in visual working memory (Lae, Rademaker) to estimates of temporal duration (Ezzyat, Huesser), to judgements of social categories (Krueger, Forster, Wilder). This confluence of findings across a wide range of fields strongly suggests that distortions exaggerating the differences between similar representations is not an error arising from computational or resource limitations but rather is a feature of the system that improves discrimination performance between otherwise confusable information. Future work will need to explore whether repulsion biases across these different domains arise from a common mechanism or whether each domain developed a distinct mechanism to arrive at a shared repelled representational structure.

Methods

Experiment 1

Exclusions

Since the task was long and monotonous many participants we created objective measures of poor performance that were unrelated to our measure of interest. The first was

Subjects

Materials

Stimuli consisted of 18 object images and 36 face images. The object images were selected from a set of images designed to be color-rotated (Brady et al. 2013). Objects were chosen that had no strong association with a particular color or set of colors. To alter the color of each object, the hue of an image was rotated through a circular colorspace ranging from 0-360 degrees. Colors were altered by independently rotating every pixel through an equiluminant circle in $L^*a^*b^*$ space. Face images were pictures of white males. For each subject, object images were randomly assigned to each of the three color difference conditions (24 degrees, 48 degree, 72 degree). Each object was then assigned two colors separated by the hue angle

degree difference of their respective condition. To do so, the colorspace was sub-selected to include 45 colors each separated by 8 degrees. Colors were then assigned to objects with the following constraints 1) within a condition, there was an even distribution of colors across thirds (120 degrees) of the colorspace, and 2) each color was only assigned once across all objects. Each colored object was then assigned a unique face associate.

Procedure

The experiment consisted of an associative learning task. Participants learned the object-face pairs through 8 rounds of iterative study and test. During a study block, the face and object appeared on screen side by side and subjects had 2.5 seconds to study the pairing. All 36 pairs were presented in a round. Next participants were tested on their color memory for each object. On each trial a studied face was presented along with its paired object presented in greyscale. Color memory was assessed using the method of adjustment. A trial started with a cursor in the center of the object image. Once participants moved the mouse the object would appear in color. The hue of the object was determined by the angle between the mouse and the center of the object image. A line marker moved around a ring surrounding the object image to indicate the current hue angle. Once participants rotated to the desired color they clicked the mouse to finalize their choice. The color wheel was randomly rotated across trials so there was no correspondence between spatial position and color. Following the color memory test, participants were tested on their associative memory for all object-face pairs. On each trial, a colored object was presented at the top of the screen. Participants were instructed to select the face paired with the image from a choice of four faces presented below. The target face was always presented along with the competitor face paired with the object's competitor (i.e the same object in a different color). Thus participants had to discriminate between the objects' colors in order to select the correct face. The other two non-competitor foils were two randomly selected faces. Each colored object served as a non-competitor foil on exactly 2 trials. Participants made a selection using a computer mouse. Participants had unlimited time to answer. They then indicated confidence in their response by clicking either a 'sure' or 'unsure' button using the mouse. Feedback was given indicating whether their choice was correct along with the correct colored object-face pairing displayed on screen for 1.25 seconds.

Following the 8 rounds of learning participants completed a post test which consisted of an extended color memory test. Trials in this color memory test were identical to those in the learning phase, however in this test, each object was tested 5 times. The order of trials was randomized across the entire test with the constraint that an object and its competitor were not tested on successive trials. Memory estimates were averaged across the 5 trials to reduce the noise associated with individual trials. Participants returned 24 hours later for a second post test. This test was identical to the previous day's test except the order of trials was re-randomized.

Analysis

The error in color memory was computed by taking the absolute hue angle difference between a participant's color estimate and the true color. Bias in color memory was then assessed by measuring the error in color estimates in relation to the object's competing pairmate color. Specifically, each color estimate was coded as towards or away from the competitor color and then we computed a bias measure by taking the percentage of estimates across all trials that fell away from the competing color.

Experiment 3 (Integration)

Subjects

Materials

The stimuli used were identical to those used during Experiment 1.

Procedure

The procedure was the same as Experiment 1 except for a few changes. The critical change from the previous experiments was the associative memory test following each round. On each trial in this test a learned face appeared at the top of the screen. Participants were instructed to select the face paired with the same object using a computer mouse. The target face was always presented along with 3 other learned faces. All the faces were tested once per round and each face served as a non-target lure on exactly 3 trials. Feedback was given indicating whether their choice was correct along with the correct face-face pairing displayed on screen for 1 second. In addition to this change, a few other minor improvements were made to the previous study. First, to address the problem of participants not finishing the experiment within the allotted time, a trial time limit of 10s was imposed for the associative and color memory tests. Second, since we did not find any differences in color memory after a 24 h delay, we excluded the delayed color memory test from this experiment.

Experiment w(Ultrasim)

Subjects

Materials

The stimuli used were identical to those used during Experiment 1 and 3.

Procedure

The procedure was exactly the same as Experiment 1. The only differences was that a trial time limit of 10s was imposed for the associative and color memory tests and there was no delayed color memory test.

Experiment 2

In Experiment 2 we tested if practice of reporting the colors during learning was necessary for the emergence of the repulsion bias. Since we only found a repulsion bias in the high similarity condition, here we dropped the moderate and low similarity conditions to increase power in our condition of interest. We also increased the number of study rounds to 14.

replicate effect. eliminated other condition to give more power. one possibility is that the test of color memory.

color retrieval itself is not necessary

As in Experiment 1, participants successfully learned to discriminate between the similar colors as evidenced by their performance on the associative memory test. At the end of 14 rounds participants successfully chose the target face over the competitor face.

Turning to the post-test we found a repulsion bias in both the signed error and percentage away measure. Thus, the repulsion bias emerged despite not having an explicit color memory test during learning.

Additionally, there was also a correlation between participants performance on the associative memory test and the magnitude of the repulsion effect.