Title
Representational Shifts Towards the Prototype in Memory for Hue

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Kelly, Laura
Heit, Evan

Publication Date
2013

Peer reviewed

Representational Shifts Towards the Prototype in Memory for Hue

Laura Jane Kelly (lkelly@ucmerced.edu)
Evan Heit (eheit@ucmerced.edu)
Cognitive and Information Sciences, University of California, Merced
5200 N. Lake Road, Merced, CA 95343 USA

Abstract

Representational shifts in memory have been a recent topic of interest and debate (Blanco & Gureckis 2012; Lupyan, 2008; Richler, Gauthier & Palmeri, 2011; Richler, Palmeri & Gauthier, 2012). Whether there are true systematic biases in memory due to a stimulus being labeled has been proposed and contested. The fundamental proposal that representations shift toward the prototype has not previously been demonstrated. In the present experiment, participants judged colored silhouettes by color category or by preference, then were asked to remember the hue of the original silhouette among five narrowly distinct options. By using the single dimension of hue, we are able to show prototypical representational shifts in memory for colored silhouettes after a few minutes. We did not observe a difference between color labeled and preference judged silhouettes, refuting the claim that labeling is the source of prototypical representational shifts.

Keywords: Concepts and Categories; Representation; Perception; Memory; Labels; Color

Introduction

A representational shift is a spatial metaphor for a systematic difference between the representation as measured and the original stimulus that inspired the representation. The representation is said to have shifted from the original sensory input somewhere along its cognitive path before being measured. The representational shift hypothesis (Lupyan, 2008) suggests that when explicit labels are used while perceiving an object, the encoded representation shifts from what it would be without explicit labeling to a more prototypical representation.

The original experiments (Lupyan, 2008), described in more detail below, did not directly address the main predictions of the representational shift hypothesis: the existence of a systematic shift or the direction of that shift. The analysis relied on inferring a representational shift from a pattern of non-directional forgetting. The data only directly indicated worse memory for labeled objects. The existence of a systematic shift and the direction of the potential shift is simply not shown by the data collected.

In the experiment presented here, we use color to test memory for recently presented silhouettes of animals and objects. In a paradigm similar to Lupyan (2008), the objects are labeled or not labeled followed by a surprise memory task. Instead of a yes-no recognition task, we present an array of 5 hue variations for the participants to choose among. By switching from multidimensional objects to unidimensional hue, and testing fine variations of hue memory rather than course-grained recognition memory, we can see whether there are systematic shifts of hue memory and their direction within the dimension of hue. The pattern of false alarms is used to look for representational shifts.

Why do these representational shifts matter? We already know there are top-down influences on memory (e.g., Heit, 1997). Two particularly dramatic examples of the imperfection of memory are eyewitness testimony (Wells & Olson, 2003) and flashbulb memories (Schmolck, Buffalo, & Squire, 2000). Small differences between the experience of an object or color and its representation in memory may seem minor in comparison to changing the race of a shooter based on stereotypes or radically rewriting how you heard about a defining national moment over the course of a few years. However, taking into account how pervasive these small differences would be, representational shifts could have extensive effects on how meaning is build and supported. Prototypes can be conceived of as resulting from the build up of exemplars over time (Nosofsky, 1986; Palmeri & Nosofsky, 2001). As a ramification of the representational shift hypothesis, exemplars in memory would not accurately reflect experience alone but would also have a bias related to when the example of the category was experienced. The stronger the category, the stronger the pull of the prototype, and as a result, the more new exemplars are biased towards categories as they already stand.

The Representational Shift Hypothesis

Lupyan (2008) proposed that there would be a difference in encoding and retrieving memories of objects if the categorization was explicit rather than only implicit at the time of encoding. Previous work had shown better category learning with labels than without labels (Lupyan, Rakison & McClelland, 2007), suggesting that there are differences in how categories with labels are activated. The difference in activation of the category could influence the encoding of exemplars of the category. Specifically, the representational shift hypothesis asserts that concurrent labels activate a more prototypical representation of the category than would the perceived exemplars activate alone. The stored encoding is hypothesized to be a mixture of the exemplar activation and the label activation; the final representation has shifted toward the prototype as a result of the interaction of perception and semantic memory.

The original work (Lupyan, 2008) used an experimental paradigm consisting of a presentation of chairs and lamps followed by a surprise recognition test including both the old objects and matched new object lures that were very similar to the old objects. There were two conditions, category judgments and preference judgments. During the initial presentation of the objects, each object was either judged to be a chair or lamp, or was judged to be liked or disliked. The results were then analyzed in terms of both hits and false alarms. Lower hit rates for the categorically judged objects than for the preference judged objects were taken to indicate a distorted or shifted memory. High false alarm rates, on the other hand, would have been taken to indicate overall poor memory. The results showed the predicted lower hit rate without a higher false alarm rate in the category judgment condition but not in the preference judgment condition. This pattern of forgetting was taken to be evidence for the representational shift account.

Challenges to a Representational Shift Account

Not all researchers accept the representational shift hypothesis. An alternative hypothesis to explain the low hit rates (Lupyan, 2008) is depth of processing. A depth of processing account predicts both high false alarm rates and low hit rates for the category judgment condition because only minimal exemplar specific information would be encoded for categorical judgments but more detailed information about each exemplar would be processed and encoded for preference judgments. Follow-up experiments explored this idea using additional judgment conditions of location (Richler, Gauthier & Palmeri, 2011) and of orientation (Blanco & Gureckis, 2012). They tested whether a preference judgment simply forces more fine-grained processing of an object than a category judgment does, leading to more detailed encodings. The location and orientation judgments were not expected to require as much processing of the actual item as preference judgments. According to the representational shift hypothesis, if labeling forces a more typical encoding than non-labeling, then a category judgment should result in the lower hit rates without a change in false alarms while preference, location and orientation judgments should all have higher false alarm rates. The labeling conditions did not uniformly create lower hit rates than non-labeling conditions. The hit rates were only lower in comparison to the preference condition supporting depth of processing.

Richler, Palmeri, and Gauthier (2012) tested the representational shift hypothesis, or labeling effect, against a paradoxical production effect (MacLeod, Gopie, Hourihan, Neary & Ozubko, 2010) which is characterized by more distinct and accurate memory for vocally named words and objects. Explicit category responses by button push, by silent labeling, and by verbal labeling were contrasted with preference judgments using the same surprise recognition memory test paradigm as Lupyan (2008). Richler et al. found varying levels of memory strength. Verbal labeling was remembered most accurately, then preference, followed by silent naming, with button press categorization being the least strong. The pattern of forgetting along with the pattern of accurate memory was used to support the distinctiveness of processing account which emphasizes the uniqueness of features over depth of processing

In each of these follow-up experiments, the pattern of forgetting was used to support a depth of processing account.

But What About the Shifts?

There is now one set of experiments in favor of the representational shift hypothesis and three sets opposed. However, none of these experiments truly get at the main prediction of the representational shift hypothesis: a stimulus processed with overt categorization will undergo a shift where the encoded representation in memory will be more prototypical than the original stimulus. The literature accepts an effect that is only implied, arguing about its cognitive mechanism rather than its existence. Rather than looking at the absence of recognition and trying to infer what processes could result in poor recognition memory, we can instead look at patterns of recognition to see if the encoding has shifted and by how much which should lead more straightforwardly to possible underlying processes. One way to go about looking at the shifts is to move the categorization and subsequent shifts onto a single dimension, in this case the hue dimension, rather than trying to infer shifts in chair-versus-lamp space, with multiple unknown dimensions.

Hue Perception and Memory

In psychological color research, it has been found that there is a near universal progression of basic color names (e.g., Berlin & Kay, 1969) and optimized focal or prototypical shades of color within a named category (e.g., Regier & Kay, 2009). This has led to further research looking at categorical perception where colors are labeled differently based on their relation to the focal colors and the color boundaries of their language's color categories (e.g., Kay & Kempton, 1984). An implication of this literature is that colors are not always experienced the same even if they are objectively the same wavelength of light. A speaker of Russian will see two shades as more different from each other when they cross the boundary between the Russian *light blue* and *dark blue* basic categories than an English speaker who would categorize them both as *blue* (Winawer et al., 2007). Color space has a distorted topography of similarity and difference based on the categories applied to it.

As implied by the term categorical perception, it is tempting to conclude that there is always an issue of categorization at play with color perception. We routinely perceive colors to be different than they actually are as in these color perception tasks. However, in the psychophysics literature, hue is highly memorable. Hue is one of three dimensions along with lightness and saturation that are used to describe color. Under some circumstances, participants can quite accurately reproduce the hue that they have just seen (Pérez-Carpinell, Baldoví, de Fez, and Castro, 1998). These results contrast with the psychological literature

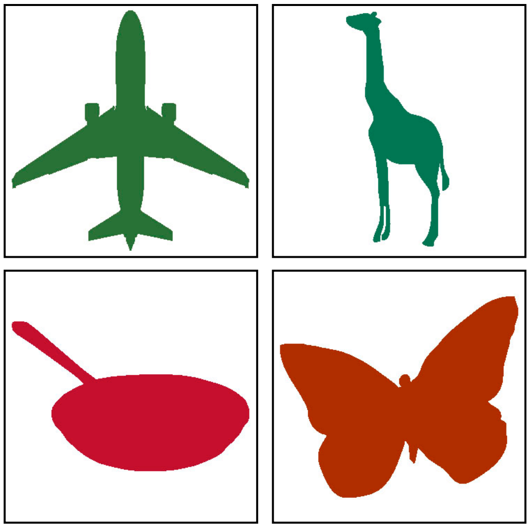

Figure 1: Example colored silhouettes. The silhouettes are representative of the two color categories: red and green, as well as the two animacy categories: living and non-living.

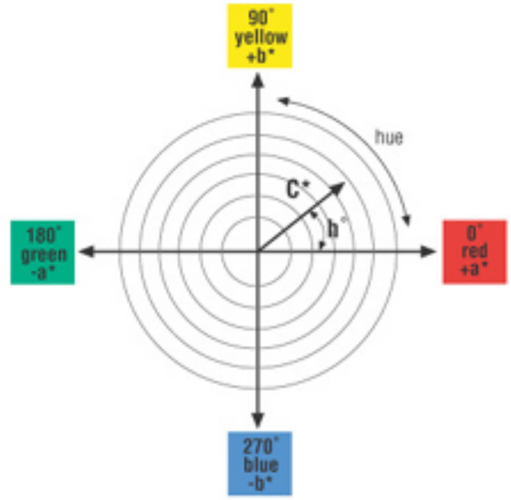

Figure 2: A representation of CIE L*C*h color space. The variations in hue are calculated in degrees on the plane of lightness and the radius of saturation. (http://nyman.netsolution.ch/IT8FujiProvia.htm)

Focused on distortion, instead suggesting that we do not always influence the colors we see with top-down knowledge but truly record hues as they objectively are in the world.

Color is a domain where memory has proven accurate to experience and is a domain where categories affect perception. The representational shift hypothesis can be tested in a domain where memory has shown to be accurate to experience without labels. By adding labels alongside hue perception we can see if representations of the colors do indeed shift.

Experiment

The present experiment was designed to test the predictions of the representational shift hypothesis: There is a systematic prototypical shift of memory for overtly labeled objects, and preference-judged objects are not subject to the same shift. Preference-judged objects could lack a representational shift or at the least demonstrate a less strongly prototypical shift.

Method

Participants 39 participants were recruited through UC Merced's participant pool. All participants were monolingual or early (by age 10) bilingual English speakers. Participants had normal or corrected to normal vision. They also had normal color vision, tested using the CITY color vision test (City University, 2002) at the conclusion of the experimental session.

Stimuli 40 colored silhouettes were created in Adobe Photoshop for the study phase of the experiment. All colors were calculated in device independent CIE L*a*b* color space and converted to RGB device dependent color space via unique monitor profiles created by a X-rite i1 Display Pro color calibrator to ensure color constancy across testing stations. 20 silhouettes were living things such as a giraffe and a butterfly, and 20 silhouettes were non-living objects such as a pan and an airplane (see Figure 1). 8 main hues, 4 reds and 4 greens were selected. The colors had the lightness and saturation values of their category's focal color (Sturges & Whitfield, 1995). Each object was randomly assigned a hue of red and a hue of green. The silhouettes were randomly assigned to 4 groups of 10 objects, 5 living and 5 non-living. The groups were then assigned a color category between subjects (i.e. Participant 1 saw groups 1 & 2 in their red hues and groups 3& 4 in their green hues while Participant 2 saw groups 1 & 3 in red and groups 2 & 4 in green, etc.) The semi-random creation of colored silhouettes preserves color and animacy balances while counterbalancing the color/shape pairings across participants.

For the testing phase of the experiment, four variations of each of the 8 main hues were calculated in CIE L*C*h color space (Figure 2) with a distance of 4° along the hue dimension from its adjacent hue. A test scale for each silhouette of two steps more typical, one step more typical, the original, one step less typical, and two steps less typical was created.

Procedure

The experiment consisted of two main parts: a judgment phase and a memory test. Each participant encountered 80 judgment trials followed by 40 memory trials. For the judgment phase, the participants had been instructed to

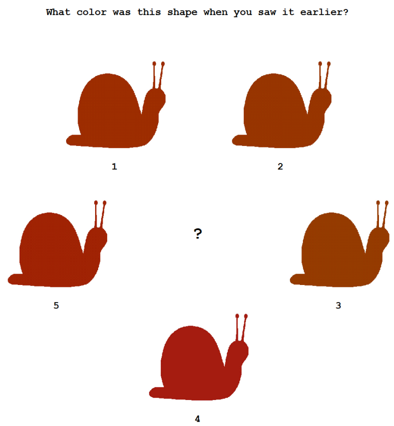

Figure 3: The memory test display of 5 hue variants.

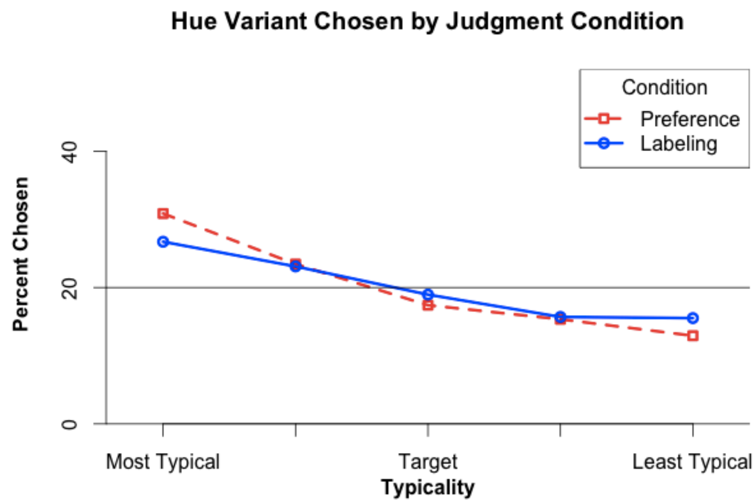

Figure 4: Average percent of each hue typicality chosen by participants. 20% chosen would be expected if participants were choosing at random.

remember the silhouettes as they would show up more than once but were not explicitly told of a memory test.

Initially, participants were instructed in the two types of judgments: color ("1" for red or "2" for green) and preference ("3" for like or "4" for dislike). Tags were placed above the keys on the keyboard to remind participants of the mappings mid-task. The trials were presented in alternating judgment blocks with 10 trials per block. Each silhouette was judged twice by each participant, both within the same judgment type (e.g., if the giraffe was judged for color the first time it appeared, the giraffe would be again judged for color the second time it appeared). The silhouettes were judged in each judgment condition between participants.

Each trial consisted of a fixation cross (1500ms), the silhouette to be judged (300ms), a question mark eliciting the judgment for that block (700ms), followed by a blank screen (1000ms).

After the judgment phase, participants were then tested for hue memory. The memory trials consisted of a circular array of the 5 hue variants of a particular silhouette (Figure 3). The array had the hues in graded clockwise order with the most typical hue rotated to a random position by trial resulting in a consistent appearance of selecting from a gradient of hues but avoiding position effects that would be present in a line. Each of the 5 positions had a location label 1 through 5 that participants entered on the keyboard to make their selection. There was no time limit imposed on the memory test responses with an intertrial interval of 1500ms.

Results

Not all participants proved equally skilled or motivated to complete this task. The following criteria needed to be met in order to include a participant's data in the analysis: at least 80% accuracy for color judgments (4 participants did not meet this criteria), and responded using at least 3 of the 5 memory test positions (6 participants did not meet this criteria). As a result, 29 participants were included in the analyses.

As can be seen in Figure 4, there was a systematic shift of responses in the hue memory test. As described above, the position of the hues was randomized, disambiguating position responses from typicality. The judged stimuli selected were both clockwise and counterclockwise to the focal color in color space disambiguated hue space direction and the direction of typicality. Only when taking into account the relative direction in hue space of the focal color does the systematic bias emerge from the data.

There are three main questions to be addressed in the data analysis to test the representational shift hypothesis: Is there a shift of memory for color judged silhouettes? Is there a shift for preference judged silhouettes? Is there a difference between the two shifts should they both be observed?

To test whether there is a shift for items in the color judgment condition, a one sample t-test with a comparison value of 3 was performed on the average typicality value of the hue chosen at memory test. Responses lower than 3 are more typical of the color category, a response of 3 is true to the original color viewed, and response values higher than 3 are atypical of the color category. The mean response typicality for color judged silhouettes is 2.70. The mean is significantly lower than 3, indicating that the shift is in the prototypical direction, $t(28)=-3.93$, $p<.001$. In the color judgment condition, there is an observed representational shift is the prototypical direction as predicted by the representational shift hypothesis.

To test whether there is a representational shift for items in the preference condition, another one sample t-test with a comparison value of 3 was preformed. Again the mean response typicality is less than 3 at 2.56 indicating a prototypical representational shift, $t(28)=-5.82$, $p<.001$. The representational shift hypothesis does not explicitly reject such a shift, but it does predict that the shift due to labeling should be stronger.

Finally, the mean is significantly greater for the color condition (2.70) than the preference condition (2.59), paired $t(28)=2.05$, $p<.05$. The direction of the difference is actually the opposite of the prediction made by the representational shift hypothesis. There is more of a prototypical shift for preference-judged stimuli than color judged stimuli. Even if the two conditions being statistically different were a matter of chance, it is quite unlikely that they are truly different in the opposite direction.

Discussion

As designed, this paradigm allows us to directly view representational shifts within hue color space. Participants were exposed to colored silhouettes then tested on 5 variations of the original silhouette. Participants on average chose hues that were slightly more prototypical of the basic color category than the hue that they had originally seen. This is the prototypical representational shift predicted by the representational shift hypothesis. However, the type of judgment made on the silhouette did not make a difference in whether a prototypical representational shift occurred. In contrast to the predictions of the representational shift account, participants chose a more prototypical hue in even larger proportions when they had made preference judgments about a silhouette rather than labeling the color category.

Depth of Processing

In previous experiments there have been measurable differences between conditions (Blanco & Gureckis, 2012; Lupyan, 2008; Richler et al. 2011, 2012) that were interpreted as superior memory for preference-judged items. The restriction to a single dimension for variation limited the ways in which preference-judged items could be uniquely encoded into memory to benefit recognition. For the dimension tested, the depth of processing predictions could have been turned around with color being more deeply processed with category judgments than with preference judgments. While making preference judgments, participants were not limited to opining on the color. It is plausible that they paid more attention to whether they liked giraffes and pans than the particular hue. Thus, a depth of processing account of the present results appears most plausible.

Implications of Representational Shifts

Earlier we discussed a potential implication of representational shifts: a pervasive influence of past experience and existing categories on new representations. These representations then become part of the categories that proceed to influence representational shifts in future experiences. Rather than categories being an accumulation of raw experience, these distortions in how new exemplars are encoded support existing category structures and discourage new categorization schemes from developing in well-categorized domains. As categorization schemes mature, the representational shifts would cascade, potentially reaching the extreme of the characteristically discrete looking categorical perception effect (Harnad, 1987; *but see* Huette & McMurray, 2010). These implications hold for the representational shift effect regardless of the mechanism behind the shifts.

Online Influence of Categories

Categories, regardless of overt labeling, affect memory. In the context of this experiment, hue memory was not as accurate as it has been reported in other research (e.g., Pérez-Carpinell et al., 1998). Given that categorization of objects is fairly automatic (Grill-Spector & Kanwisher, 2005), having category judgments in the course of the experiment could have created a 'category-relevant' context where regardless of explicit categorization responses, categories were activated for all stimuli. The implication is that categories alter memory formation online rather than being a permanent perceptual bias consistent with recent research (e.g. Landau, Dessalegn, & Goldberg, 2010). A non-category relevant context with added overt labels is a possible scenario where labeling may have an effect on representational shifts.

Time Course

Our previous research (Kelly & Heit, 2012) has shown that atypical representational shifts—shifts away from the prototype—are found with immediate (half second and five second delay) recognition tests. At some point between five seconds and a few minutes representations go from being distorted atypically to being distorted typically. Representations are not veridical to the stimulus initially or subsequently.

It's possible that these shifts are symptomatic of competing needs of working memory and long term memory. Working memory could privilege differentiating information in case details are important in the moment. Detailed information that will go unused would deplete resources unnecessarily when being encoded into long term memory, making relying on the category general information to supply a complete representation upon recall more advantageous. Variables that change depth of processing could be indicative of how likely specific differentiating information is to be needed in the future. These explanations for the mobility of representational shifts are purely speculative and need to be researched further.

Push or Pull?

Lupyan (2008) hypothesized the mechanism of blended representations between exemplar and category prototype as the cause of prototypical representational shifts. Another proposed mechanism to account for this bias is boundary truncation (Huttenlocher, Hedges, Lourenco, Crawford, &

Corrigan, 2007). Rather than having a pull toward the prototype via prototype activation, there is a push toward the prototype by disregarding extreme information. The shift into the category would be consistent with both theories and needs to be explored further.

Conclusion

Representational shifts exist in memory for hue. There are prototypical shifts in memory for colored silhouettes encountered minutes before test. These shifts can be seen and measured on a unidimensional testing ground. Conceptual space has long been known to be contorted in color (Kay & Kempton, 1984) and beyond (Goldstone, 1992). Our results support the idea that the creation and maintenance of these contortions could be due to representational movement.

References

Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Blanco, N. and Gureckis, T.M. (2012). Does category labeling lead to forgetting? *Cognitive Processing.*

City University (2002). A web-based colour vision test. http://www.city.ac.uk/avrc/colourtest.html.

Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, *49(1),* 585-612.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 151-160.

Harnad, S. R. (Ed.). (1990). *Categorical perception: The groundwork of cognition.* Cambridge, UK: Cambridge University Press.

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts, and Categories*, 7-41. London: Psychology Press.

Huette, S. & McMurray, B. (2010). Continuous dynamics of color categorization. *Psychonomic Bulletin & Review, 17*(3), 348-354.

Huttenlocher, J., Hedges, L. V., Lourenco, S. F., Crawford, L. E., & Corrigan, B. (2007). Estimating stimuli from contrasting categories: Truncation due to boundaries. *Journal of Experimental Psychology: General, 136,* 502-519.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis?. *American Anthropologist*, *86(1),* 65-79.

Kelly, L.J. & Heit, E. (2012, November). *Representational Shifts in Memory for Hue*. Poster presented at the 53rd Annual Meeting of the Psychonomic Society, Minneapolis, Minnesota.

Landau, B., Dessalegn, B., & Goldberg, A.M. (2010). Language and space: Momentary interactions. In P. Chilton and V. Evans (Eds.), *Language, cognition and space: The state of the art and new directions. Advances in Cognitive Linguistics Series* (Ed, V. Evans, B. Bergen, J. Zinken). London: Equinox Publishing.

Lupyan, G. (2008). From chair to "chair:" A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General, 137(2)*, 348-369.

Lupyan, G., Rakison, D.H., & McClelland, J.L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science, 18(12)*, 1077-1083.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115(1),* 39.

Palmeri, T. J., & Nosofsky, R. M., (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology, 54,* 197-235.

Pérez-Carpinell, J., Baldoví, R., de Fez, M. D. and Castro, J. (1998). Color memory matching: Time effect and other factors. *Color Research and Application, 23,* 234–247.

Regier, T. & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Science, 13,* 439-446.

Richler, J.J., Gauthier, I., & Palmeri, T.J. (2011). Automaticity of basic-level categorization accounts for labeling effects in visual recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1579-1587.

Schmolck, H., Buffalo, E. A., & Squire, L. R. (2000). Memory distortions develop over time: Recollections of the OJ Simpson trial verdict after 15 and 32 months. *Psychological Science*, *11(1),* 39-45.

Sturges, J. & Whitfield, T.W.A. (1995). Locating basic colours in the munsell space. *Color Research and Application, 20*, 364-376.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*(1), 277-295.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy Sciences 104*, 7780 –7785.