## Title
Using a Task-Filled Delay During Discrimination Trials to Examine Different Components of Learned Visual Categorical Perception



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

**Authors**
de Leeuw, Joshua R
Andrews, Jan

**Publication Date**
2015

Peer reviewed

# Using a Task-Filled Delay During Discrimination Trials to Examine Different Components of Learned Visual Categorical Perception

**Joshua R. de Leeuw (jodeleeu@indiana.edu)**
Department of Psychological and Brain Sciences, Program in Cognitive Science
Indiana University, Bloomington, IN

**Jan Andrews (andrewsj@vassar.edu)**
Department of Cognitive Science
Vassar College, Poughkeepsie, NY

**Abstract**

The evidence concerning the level at which learned CP effects occur is complex. The goal of this study was to use a different approach to this question by manipulating the abstractness of the information available for distinguishing pairs of items in an XAB task, and the presence or absence of a short task-filled delay between X and AB. Participants engaged in XAB trials containing a mixture of trials with and without the delay task before and after standard training to classify visual texture stimuli into two categories. Training improved discrimination of pairs differing on the category-relevant dimension whether within- or between-category, but not on pairs differing only on non-category relevant low level features. In addition, only successful learners in the post-training trials avoided decreased discrimination accuracy due to the delay task, suggesting that they formed more stable representations. However, this effect was not limited to pairs varying in category-relevant ways.

**Keywords:** Categorization; categorical perception; compression; expansion; learning; discrimination; bottom-up; top-down; dimensions; interference.

## Introduction

Learning to place objects into novel categories affects the way those objects are judged and how well they can be discriminated from one another. These so-called learned categorical perception (CP) effects are well established (e.g., Goldstone, 1994; Livingston, Andrews, & Harnad, 1998), but there are several major unresolved issues concerning their fundamental nature. For example, learned CP effects always involve better sensitivity to variations in items that are from different categories than variations in items that are from the same category. However, there are several distinct patterns of learning that result in the relative advantage for between-category pairs. The changes in sensitivity could be localized to the boundary between categories, resulting in heightened sensitivity to variations that cross a category boundary or decreased sensitivity to variations that do not cross a boundary. Or the changes could be at the level of perceptual dimensions, with increased sensitivity to *any* changes along category-relevant dimensions and/or decreased sensitivity to changes along category-irrelevant dimensions, regardless of whether or not the stimuli are in different categories. Some researchers apply the term "learned CP" only to the boundary specific effects known as compression (reduced sensitivity to differences among within-category stimuli) and expansion (enhanced sensitivity to differences among between-category stimuli) (e.g., Folstein, Palmeri, & Gauthier, 2014). Dimensional sensitivity changes that are not restricted to the category boundary are often referred to as acquired equivalence (reduced sensitivity to a category-irrelevant dimension) and acquired distinctiveness (enhanced sensitivity to a category-relevant dimension). In what follows, we will take the term "learned CP" to include all of these.

An important issue in learned CP research is the locus of the learning effect: is it a perceptual effect – do people actually see stimuli in different ways after learning to categorize? – or is the effect post-perceptual? There is mixed evidence as to whether CP is perceptual. We believe the multifaceted nature of learned CP is complicating efforts to understand the mechanism(s) underlying the phenomenon. Part of the apparent controversy may simply be that different studies of CP are isolating different kinds of learning, such as boundary effects vs. dimension-wide effects, which may differ in the level at which they occur.

For example, Goldstone and Hendrickson (2009) argue that effects occur at multiple levels based on evidence from studies using vastly different methodologies. For example, studies of speakers of languages with distinct color terms such as "blue" and "green" show faster discrimination of boundary-straddling stimuli than speakers of languages that lack distinct terms, but the fact that these effects appear to be strongest when stimuli are presented to the right visual field/left hemisphere and are disrupted by verbal interference suggests that the effects are not deeply perceptual. On the other hand, evidence that objects in the same category are judged to be more similar not only to each other but also to a neutral, non-categorized object suggests a representational change that is not label-based (Goldstone et al., 2001). In addition, Notman, Sowden, and Özgen (2005) showed that a strong expansion effect observed for participants who learned to categorize oriented line gratings did not transfer to stimuli of varying orientations, suggesting that the learned CP effect occurred

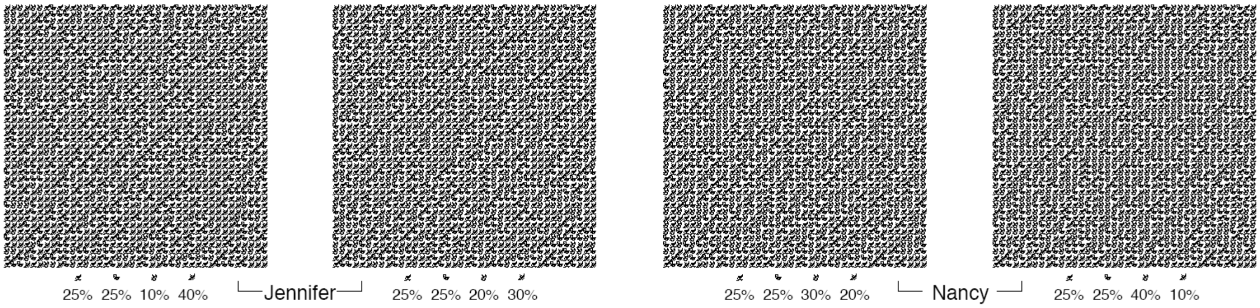


Figure 1: Example stimuli. The proportion of each microfeature is shown below the stimulus. The relative proportion of microfeature 3 to microfeature 4 (the category relevant dimension) increases from left to right. The two leftmost stimuli were categorized as being created by Jennifer, while the rightmost stimuli were created by Nancy.

relatively early in visual processing. Yet Roberson, Hanley, and Pak (2008) found no evidence of greater sensitivity in terms of absolute discrimination thresholds for JNDs at color category boundaries for groups demonstrating category boundary CP effects.

This issue has been framed in a variety of ways: auditory vs. phonetic (category-based) processing in speech CP (e.g., Gerrits & Schouten, 2004), visual vs. verbal/categorical codes in color CP (e.g., Hu, Hanley, Zhang, Liu, & Roberson, 2014), strategic judgment bias vs. altered object description (Goldstone, Lippa, & Shiffrin, 2001), bottom-up vs. top-down effects of category knowledge, and so on. As suggested above, a great deal of empirical evidence can be found to support both sides of this opposition, but it is rarely definitive for a number of different reasons. For example, if top-down effects are extremely rapid, they may be very difficult to distinguish behaviorally at least from bottom-up effects, and if both types of process are occurring simultaneously, they may be difficult to disentangle. Effects found with known categories may not occur in the process of learning new ones.

It seems likely that one variable of importance in producing learned CP at different levels is the type of stimulus variation that occurs. This study represents a preliminary attempt to clarify the role of different types of information in learned CP by distinguishing among different levels of variation within the stimuli themselves and measuring how sensitivity to these kinds of variations changes as a result of category learning. At least three levels can be identified: low-level (essentially random) details completely unrelated to category membership; abstracted perceptual dimensions that may or may not be related to category membership; and category membership itself. These form a hierarchy of abstractness, with each level incorporating the one(s) below it; for example, items differing in category membership will also differ both dimensionally and in their randomly varying low-level details. The low-level details depend most heavily on bottom-up processing and category labels are most clearly a top-down information source, while abstracted perceptual dimensions fall somewhere in between. Changes at the levels of perceptual dimensions and category membership could produce learned CP, but the underlying mechanism of learning might look different in each case. Learning based solely on category membership might produce category-boundary-specific effects, while changes in the salience of a perceptual dimension would result in an acquired distinctiveness effect. Our stimulus set allows both kinds of learning to take place. Whether both kinds of learning actually happen is an empirical question.

To further bolster the ability of the experiment to distinguish between different components of learned CP, we sometimes use a task-filled delay during the discrimination task used to test for effects of category learning. The rationale for this is that the delay manipulation should interfere most with remembering bottom-up information and least with remembering category-level variation. If learned CP is the result of acquiring strictly category-level information, then the delay should not affect trials where category-level information varies, but should affect other trials. If learned CP involves changes in perceptual representations of the stimuli, performance after training on trials with category-relevant feature variation should be less sensitive to the delay. An early demonstration of this sort was provided by Pisoni (1973), who showed that within-category vowel discrimination accuracy was inversely related to the length of a delay of up to 2 seconds, while between-category accuracy was unaffected, because the former relied on auditory short-term memory while the latter relied on phonetic categories. More recently, Pilling, Wiggett, Özgen, and Davies (2003) tested color CP effects using a 5 second delay during discrimination trials that contained no task, a visual interference task, or a verbal interference task and found that only verbal interference ever disrupted CP for between-category comparisons, and then only if different delay conditions were blocked. Pilling et al. took their results to support the view that linguistic/categorical codes caused the observed CP effects.

## Method

### Participants

We recruited 59 undergraduate students to participate in the experiment in exchange for course credit.

### Materials

Stimulus materials were modeled on those used by Pevtzow and Harnad (1997). Each stimulus consisted of 1600 microfeatures arranged in a 40 x 40 grid. The total size is 320 x 320 pixels. Each microfeature is an 8 x 8 image, with 30 pixels (out of 64) colored black. Each colored pixel is adjacent to at least one other colored pixel. A total of four different microfeatures were used (see Figure 1). For each stimulus, half of the microfeatures consisted of equal numbers of microfeatures 1 and 2 (so 400 of each, or 25% and 25% of the total). Two categories of stimuli were created by varying the proportions of microfeatures 3 and 4 in each stimulus as follows: for category 1 (art by Jennifer), 10%/40% or 20%/30%; for category 2 (art by Nancy), 30%/20% or 40%/10%. Thus category 1 stimuli contained fewer instances of microfeature 3 and more of microfeature 4 than category 2 stimuli. See Figure 1 for sample items from each category.

The location of the microfeatures varied from presentation to presentation and was random except for the top 3 and bottom 3 rows, which were kept identical in all stimuli to discourage fixation strategies. Each feature was equally represented in these 6 rows and the particular pattern of the six rows was randomly generated for each individual subject but consistent within a particular subject.

Due to the nature of the stimulus features in relation to the categories, items can differ from one another in up to three different ways: (1) low level only ("L"), i.e., specific location of all microfeatures, which is not relevant to the category distinction in any way; (2) value on the category-relevant dimension ("L+D"), which is the relative proportions of microfeatures 3 and 4, and (3) category membership ("L+D+M"). Note that all items differ on the irrelevant variation (hence the "L" in all the pair type names) and only between-category pairs differ on all three types of variation.

### Procedure

The experiment was created using the jsPsych platform (de Leeuw, 2015). Subjects completed the experiment using the Chrome web browser in a laboratory setting. The browser was displayed in full-screen mode so that only experiment-relevant material was on the display.

The task used to assess learned CP effects consisted of displaying one stimulus (X) for 1500 ms followed by a pair of stimuli (AB) side by side, one of which was identical to X. The position of the stimulus that was identical to X varied randomly. AB was shown until the participant responded.

Participants were told that they would be viewing and judging pieces of digital art. A pre-training, training, post-training design was used, with pre- and post-training consisting of XAB trials, half of which were standard (i.e., exactly as described above) and half of which contained a 6 s delay between X and AB. During the delay the participant was asked to track the location of a dot in a 5 x 5 grid and click on the last square in which it appeared.

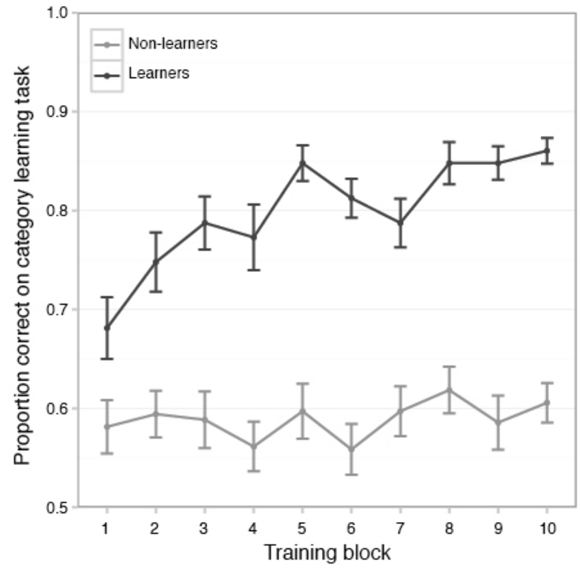


Figure 2: Classification accuracy of learners and non-learners over the course of training.

Pre-training consisted of 3 blocks of 24 XAB trials, where each block contained equal numbers of three pair types defined according to the types of variation described above: pairs differing only on irrelevant variation (L), pairs differing also on the category relevant dimension (L+D), and between-category pairs (L+D+M). For each of those pair types, there were equal numbers of trials with and without a task-filled delay, presented intermixed in a random order.

To facilitate learning the categories, participants were initially shown 5 different pairs of stimuli described as artwork by two different artists, Nancy and Jennifer, to familiarize them with the category-relevant and irrelevant variation among the stimuli. This was followed by 200 training trials, divided into 10 blocks of 20 trials each, with each of the 4 values of the category-relevant dimension (10%/40%, 20%/30%, 30%/20%, 40%/10%) represented 5 times and the 20 items presented in a random order. On each trial, a stimulus was presented until the participant responded by pressing one of two keys on a computer keyboard to indicate if the artist was Nancy or Jennifer. Feedback was provided on each trial and at the end of each block.

At the conclusion of training, the post-test was conducted with a procedure identical to that of the pre-test and using the exact same item pairs randomly reshuffled.

## Results

### Category Learning

There was substantial variability in performance on the category-learning task. Some subjects learned the correct categorization almost immediately, presumably from the examples shown immediately before the first block of training. Some subjects never learned to categorize the items correctly, and remained at chance levels of performance throughout the entire training session. Some subjects showed low initial performance, but improved throughout the training and were competent categorizers by the end of training.

Because the presence of CP effects depends on successful category learning, we divided the subjects into a learning group and a non-learning group. The criterion for being in the learning group was an overall accuracy of 80% or above on the last two blocks of the categorization training. There were 24 subjects in the learning group (mean accuracy = 85.4%, *SD* = 4.5%), and 33 subjects in the non-learning group (mean accuracy = 60.2%, *SD* = 11.4%); see Figure 2.

### Delay Task

Accuracy on the intermediate delay task was quite high overall (*M* = 92.7%, *SD* = 11.7%). Two subjects had an accuracy level below 80% (65.3% and 18.1%). These subjects were removed from the subsequent analysis.

### CP Effects and Effects of Delay

Using the dependent variable of proportion correct on XAB trials, a 2 (learner/non-learner) X 3 (pair type: L/L+D/L+D+M) X 2 (pre/post) X 2 (delay task/no delay task) ANOVA was conducted:, with repeated measures on all but the first factor. This yielded main effects of learning, $F(1, 55) = 8.966$, $p = .004$, $\eta_p^2 = .140$, with learners being more accurate (*M* = .671, *SD* = .098) than non-learners (*M* = .593, *SD* = .098); pair type, $F(2, 110) = 3.855$, $p = .024$, $\eta_p^2 = .065$, with L+D+M pairs more accurate (*M* = .651, *SD* = .121) than L+D pairs which were more accurate (*M* = .633, *SD* = .113) than L pairs (*M* = .612, *SD* = .113); and delay, $F(1, 55) = 17.107$, $p < .001$, $\eta_p^2 = .237$), with delay causing lower accuracy (*M* = .608, *SD* = .098) than no delay (*M* = .656, *SD* = .121).

Of greater interest were two interactions that were obtained: Pair type interacted with pre/post, $F(2, 110) = 3.746$, $p = .027$, $\eta_p^2 = .064$). As shown in Figure 3, discrimination performance after category training was higher on L+D+M and L+D pairs but not L pairs, a pattern consistent with expansion and/or acquired distinctiveness learned CP effects.

In addition, there was a significant three-way interaction between learning, pre/post, and delay, $F(1, 55) = 4.283$, $p = .043$, $\eta_p^2 = .072$). As shown in Figure 4, relative to no delay, the delay lowered discrimination accuracy for non-learners both before and after category training, but only lowered discrimination accuracy for learners before category training. However, this was true regardless of pair type as the four-way interaction was not significant.

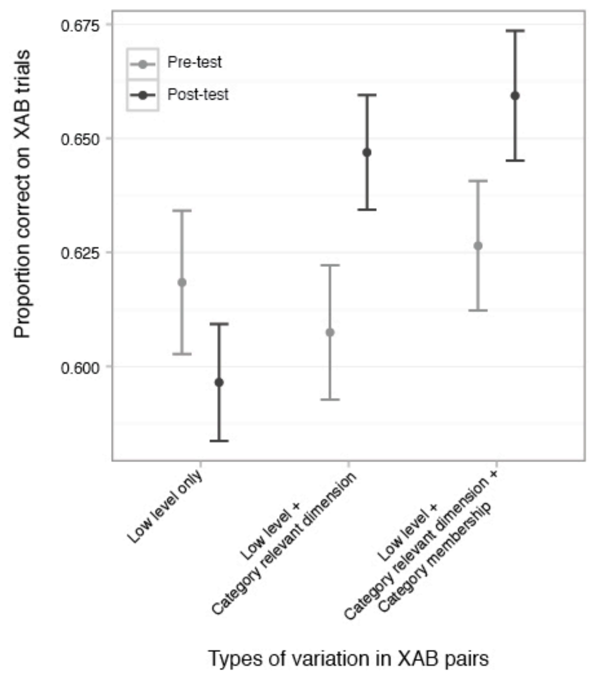


Figure 3: Discrimination accuracy as a function of pair type both pre and post category training.

## Discussion

This study was designed to provide evidence on two questions: First, how does category training change people's sensitivity to different kinds of features, i.e., features at different levels of abstraction and category-relatedness? And second, are people better at remembering different kinds of features after category training (suggesting that they formed more stable strategies for encoding these aspects of the stimuli)? The three kinds of features examined here were hierarchically nested and consisted of low-level details (L), the category-relevant dimension (L+D), and category membership (L+D+M).

On the first question, category training led to improved discrimination performance on L+D and L+D+M pairs but not L pairs, suggesting greater reliance on higher level than purely bottom up information in the stimuli. This isn't surprising since the low level information wasn't relevant to the categories, but does show that experience with that low level information, which is relevant to the XAB task, does not lead to improvement based on its use alone. In addition, this effect was obtained irrespective of category learning success. Also, the fact that both L+D and L+D+M pairs showed improvement to about the same extent may mean that sensitization to the category-relevant dimension, rather than the use of category-level information per se such as the label, was driving the improved performance because the additional feature of category membership did not improve

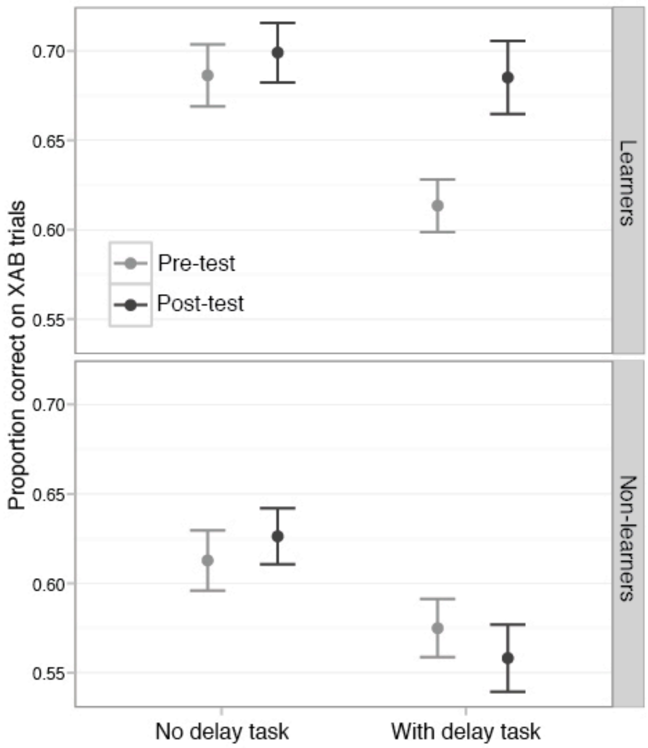


Figure 4: Discrimination accuracy as a function of presence or absence of a delay task for learners and non-learners both pre and post category training.

learning more than just the variation along the category relevant dimension.

On the second question, the data suggested that learners formed more stable representations than non-learners, resulting in no effect of the delay task on XAB performance in the post-category training test for learners. Yet this effect did not differ according to pair type, occurring equally for L, L+D, and L+D+M pairs. We originally expected that the delay would interfere most directly with the discrimination of L pairs given the previous findings that a delay interfered most with lower-level information (Pisoni, 1973). However, it's possible that successful category learning also enhanced sensitivity to the low-level irrelevant variations, perhaps because they overlap with the category-relevant dimension, in such a way as to allow for a more effective means of remembering the stimuli during the XAB task. One hypothesis along these lines is that learners may have learned to selectively attend to the microfeatures that define the category relevant dimension, making it easier to encode the low-level variation of the stimulus by reducing the number of microfeatures that were encoded. This would suggest that learners did not simply encode the abstract category-relevant dimension, but rather encoded perceptual features of the stimulus that were directly relevant to the category-relevant dimension. This hypothesis could be tested by creating additional pairs for the XAB task that varied only in the location of the microfeatures that are not part of the category-relevant dimension. This new condition would allow us to distinguish between the learning of different low-level features depending on whether or not those features are related to the category-relevant variation.

The main puzzle posed by the results is that category training seems to result in an overall boost in the discriminability of pairs that vary along the category-relevant dimension but not pairs that vary in only category-irrelevant ways, yet learners show an overall increase across all pair types in discriminability with the delay task. One explanation is that the distinction between learners and non-learners is inexact; any arbitrary cutoff between learners and non-learners will be noisy. It's likely that there are a non-trivial number of people who learned some aspect of the category structure in the non-learning group. This, coupled with the fact that the effect sizes for changes related to specific kinds of variation were relatively small, could partially explain why there was no interaction between successful learning and improvement on each of the types of variation.

While we found evidence that successful category learning altered the way that individual stimuli are remembered, the mechanism for this change remains unclear in light of the non-interaction between the delay task and the type of variation of the XAB pairs (L vs. L+D vs. L+D+M). In addition to the idea of introducing another kind of pair type variation, two other aspects of our methodology could be easily modified to further probe this pattern of results: the task during the delay and the sequence of XAB trials.

The task used in this study during the delay was a visual task; the role of labeling in category-trained performance on the XAB task could be addressed by using a verbal task during the delay. Pilling et al. (2003) showed that verbal but not visual interference during discrimination trials removed the between-category advantage. However, they also showed that this was only true if the different types of interference were blocked; when they were intermixed, the learned CP effect remained. This suggests that people are strategic in their use of category information. When they have expectations that category labels will be challenging to remember, they opt for using the non-verbal aspects of the stimulus to do the discrimination task. As Pilling et al. note, however, non-verbal does not necessarily mean low level, a point that receives support from studies showing that categories can be learned, and CP effects produced, in the absence of verbal labels (e.g., Andrews, Livingston, Sturm, Bliss, & Hawthorne, 2011; Wolff & Holmes, 2012.)

We can manipulate potential strategic uses in our study by altering the sequence of XAB trials. In our design, trials with the visual task were intermixed with trials having no delay, and this may have influenced the strategies participants used. If trials were blocked, participants might develop strategies specifically suited to immediate versus delayed discrimination. If participants know that they will have to remember the stimulus during a delay task, they may choose to focus more on easily compressible aspects of the stimulus, such as higher-level category features. Conversely, if they expect to be able to make their judgment immediately, they may focus on more concrete perceptual details (especially since category-level information is more inferential and easier to get wrong).

## Acknowledgments

This material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1342962. We thank Tom Murphy and Nick Boyd for assistance with data collection.

## References

Andrews, J., Livingston, K., Sturm, J., Bliss, D., & Hawthorne, D. (2011). Category learning research in the interactive online environment *Second Life*. In T. Pinelli (Ed.), *Selected Papers and Presentations Presented at MODSIM World 2010 Conference & Expo*. NASA.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1-12. doi:10.3758/s13428-014-0458-y

de Leeuw, Andrews, & Livingston. (2014). Learned visual categorical perception effects depend on method of assessment and stimulus discriminability. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 391-396). Austin, TX: Cognitive Science Society.

Folstein, J., Palmeri, T., & Gauthier, I. (2014). Perceptual advantage for category-relevant perceptual dimensions: The case of shape and motion. *Frontiers in Psychology, 5*, 1394.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–76.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200.

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical perception. *Interdisciplinary Reviews: Cognitive Science, 1*, 69-78.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43.

Hanley, J.R. & Roberson, D. (2011) Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin & Review*, *18*(2), 355-363.

Holmes, K., & Wolff, P. (2012). Does categorical perception in the left hemisphere depend on language? *Journal of Experimental Psychology: General*, *141*, 439-443.

Hu, Z., Hanley, J.R., Zhang, R., Liu, Q, & Roberson, D. (2014). A conflict-based model of color categorical perception: Evidence from a priming study. *Psychonomic Bulletin & Review*, *21*, 1214-1223.

Livingston, K. R., & Andrews, J. K. (2005). Evidence for an age-independent process in category learning. *Developmental Science*, *8*(4), 319–25.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732–753.

Notman, L., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: a psychophysical approach. *Cognition*, *95*(2), B1–14.

Pevtzow, R., & Harnad, S. (1997). Warping similarity space in category learning by human subjects: The role of task difficulty. In M. Ramscar, U. Hahn, E. Cambouropolos, & H. Pain (Eds.), *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization* (pp. 189–195). Department of Artificial Intelligence, Edinburgh University.

Pilling, M., Wiggett, A., Özgen, E., & Davies, I. (2003). Is color "categorical perception" really categorical? *Memory & Cognition, 31*(4), 538-551.

Pisoni, D. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*(2), 253-260.

Roberson, D., Hanley, J.R., & Pak, H.S. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition, 112,* 482-487.