Going to Extremes: The influence of unsupervised categories on the mental caricaturization of faces and asymmetries in perceptual discrimination

Andrew T. Hendrickson; Paulo F. Carvalho; Robert L. Goldstone
({athendri, pcarvalh, rgoldsto} @indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
1101 East Tenth Street, Bloomington, IN 47405 USA

Abstract

Recent re-analysis of traditional Categorical Perception (CP) effects show that the advantage for between category judgments may be due to asymmetries of within-category judgments (Hanley & Roberson, 2011). This has led to the hypothesis that labels cause CP effects via these asymmetries due to category label uncertainty near the category boundary. In Experiment 1 we demonstrate that these "within-category" asymmetries exist before category training begins. Category learning does increase the within-category asymmetry on a category relevant dimension but equally on an irrelevant dimension. Experiment 2 replicates the asymmetry found in Experiment 1 without training and shows that it does not increase with additional exposure in the absence of category training. We conclude that the within-category asymmetry may be a result of unsupervised learning of stimulus clusters that emphasize extreme instances and that category training increases this caricaturization of stimulus representations.

Keywords: Categorical Perception, Category Labels, Perceptual Learning, Category Learning, and Language

Introduction

Categorical perception. Our perceptual systems fail overwhelmingly to be precise replicators of reality in the way a camera or a microphone is, because these systems have not evolved to create a veridical representation of reality. Though constrained by overall neural architecture and the inertia of representations in primary sensory areas (Petrov et al., 2005), our perceptual systems consistently learn to create useful, but potentially distorted, representations of reality (Landy & Goldstone, 2005).

Often, this perceptual learning produces experiences that do not reflect the continuous variation of reality. Instead they warp that variability into discrete groupings such that entities that fall within a group are less discriminable than physically equally spaced entities that fall in different groups, a process known as categorical perception (CP; Harnad, 1987).

While some of the focus in CP research has been on assessing if particular categories are innate through cross-cultural studies (Kay & Reiger, 2003; Roberson & Davidoff, 2000; Sauter et al., 2011), early studies of CP focused on phonemes (Liberman et al., 1957) which show systematically different category boundaries based on an individual's native language (Logan et al., 1991).

Learned CP has been shown in the visual modality across a variety of dimensions including hue and saturation (Goldstone, 1994), line drawings (Livingston et al., 1998), and morphs between arbitrarily paired faces (Kikutani et al., 2008; 2010).

Category labels and CP. An alternative framework suggests that the presence of category labels, and not perceptual changes, are responsible for CP effects (Pisoni & Tash, 1974). In this view the category label can be seen as an additional feature: entities in different categories have different labels thus having an additional feature unique for each category. This causes similarity to decrease and discrimination accuracy to rise. Items in the same category have the same label and thus either their similarity increases or remains constant leading to discrimination accuracy that does not increase.

Hanley and Roberson (2011) point out that the accuracy in assigning category labels is not constant across distance to the category boundary. Items farther away from the boundary are more likely to be categorized correctly than items closer to the category boundary. This viewpoint is consistent with many models of category learning that do not incorporate perceptual learning, including decision boundaries (Ashby & Maddox, 1990) and many exemplar-based (Nosofsky, 1986) models of category learning.

Within-category discrimination asymmetries. In perceptual discrimination testing in which a target object (X) must be held in memory and compared to itself and a foil object (A and B, respectively), if A is more likely to be assigned the same category label as X than B, then the probability of selecting A as the answer should increase relative to if A and B are equally likely to be assigned to categories. Therefore, when the target object is farther away from the category boundary than the foil and thus more consistently labeled in the category, accuracy will increase because the target object is more likely to be selected. Similarly, when the foil object is farther away, accuracy will decrease because the foil object will be selected more frequently (compared in both cases to cases in which no labeling asymmetry exists).

Hanley and Roberson (2011; see also Roberson et al., 2007) find this asymmetric within-category advantage for more perceptually extreme targets across a wide array of stimuli for which CP effects have been shown, including color across cultures (Roberson & Davidoff, 2000; Roberson et al., 2000; Roberson et al., 2005), facial emotions (Roberson et al., 2007), morphed celebrity faces and morphed unfamiliar but trained faces (Kikutani et al., 2008; 2010). They failed to find an advantage for more extreme faces among morphed unfamiliar and either untrained (Kikutani et al., 2008) or covertly exposed (Kikutani et al.,

2010) conditions. Recently, Sauter et al. (2011) failed to replicate the within-category asymmetry across cultures for morphed facial emotions despite showing CP effects.

CP within categories. Recent evidence has demonstrated (Gureckis & Goldstone, 2008; Hendrickson et al., 2010) that CP effects emerge not only between categories but also within categories. For example, two objects that belong to the same learned category (receiving the same label) may nonetheless have increased discriminability if they belong to different clusters within the category when compared to the case in which they belong to the same cluster. Within-category CP effects occur when the distribution of members of a category is structured into clusters (sub-groups within each category) rather than distributed uniformly (e.g. Goldstone, 1994) or normally (e.g. Ashby & Maddox, 1990).

These within-category CP effects are consistent with models of categorization in which the discriminability of items is not only affected by their category label but also by the learned clustering of items regardless of their labels (Love et al., 2004; McDonnell & Gureckis, 2011). These learning processes account for both within and between category CP effects through representational change: learning new clusters or prototypes that warp the similarity between entities either within or between categories (Goldstone & Hendrickson, 2009).

Within-cluster discrimination asymmetry. Interestingly, Gureckis and Goldstone (2008; see also Hendrickson et al., 2010) also found that the magnitude of the CP effects on both the category relevant and the category irrelevant dimensions increased as categorization accuracy improved. Importantly, neither of these CP effects were found before training.

Using this kind of stimuli space, the label ambiguity account of CP hypothesizes that the within-category asymmetry should emerge with the CP effect along the category-relevant dimension and is in fact causing the categorical perception effect. This would be for discriminations perpendicular to the category boundary in a two-dimensional space (see Fig. 1).

What remains unclear is if, within each category, a similar asymmetry should emerge parallel to the category boundary. A strict category label account suggests this should not occur because all stimuli would be equidistant to the category boundary and thus categorized equally accurately. This strict viewpoint would need to postulate a second mechanism to account for within-category CP effects.

Conversely, a category label ambiguity theory of CP that allows each cluster within a category to have a unique label would predict that the asymmetry will occur along the category-irrelevant dimension and that the emergence of the asymmetry will cause the within-category CP effect. The main purpose of this work was to investigate the emergence of within-category asymmetries along both the category relevant and irrelevant dimensions. A pre-post design was used to assess the relative timing during training of the emergence of CP effects and within-category asymmetries.

Experiment 1

In the first experiment we tested these predictions by measuring perceptual discrimination accuracy along both relevant and irrelevant dimensions before and after category training. The stimuli and category structures were identical to previous studies (Gureckis & Goldstone, 2008; Hendrickson et al., 2010) that showed CP effects both dimensions. The perceptual discrimination task was a two alternative forced choice (2AFC) XAB task similar to those reported by Hanley and Roberson (2011). The within-cluster asymmetry and the standard CP effect were measured along the category relevant and category irrelevant dimensions before and after category training.

Method

Participants. 80 Indiana University undergraduates participated in this experiment for course credit. 1 participant was excluded from analyses for failing to conclude the experiment within the allotted time (60 min).

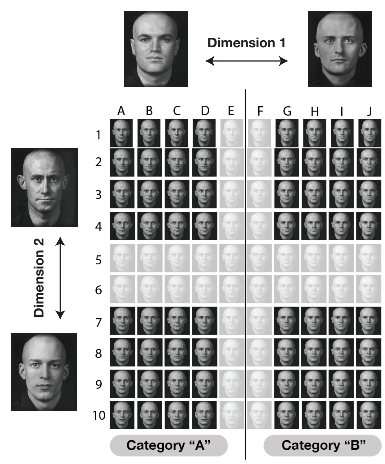

Fig. 1: Stimuli varied along two arbitrary dimensions (1 and 2) forming a 10-by-10 grid of blended faces. The light grey stimuli were not used in discrimination or categorization, introducing a source of within-category structure (two clusters of faces within each category). The vertical line between columns E and F shows an example category boundary used during category learning (the other boundary was a horizontal line between rows 5 and 6).

Materials. A 2-dimensional 10 by 10 matrix of bald male faces was created using a standard morphing technique (Steyvers, 1999). Each dimension was created by morphing between two faces selected from Kayser (1997). The two selected faces were roughly equally spaced in the multi-dimensional space based on a pilot similarity judgment task. The 100 stimuli that constitute the full matrix were created by equally morphing between all unique pairs of 10 faces in each of the two dimensions of faces (see Fig. 1).

Procedure. The task consisted of a block of 192 discrimination trials (pre-categorization phase), followed by 8 blocks of 16 categorization trials (categorization phase), and a second block of 192 discrimination trials (post-categorization phase).

Each discrimination trial followed the XAB pattern: a target stimulus (X) was presented for 500 ms in the center of the screen followed by a response screen containing a target and a foil (A and B) stimulus presented horizontally until a response was made. A 500 ms blank screen was presented between the two screens and between trials there was a pause of 1000 ms. Participants were instructed to determine whether A or B was identical to X. The "target" is the option identical to X, and the "foil" is the other choice.

Target and foil face stimuli were selected such that they were identical along one dimension and were separated by 2 face stimuli in the 10 X 10 stimulus space along the other dimension. This spacing was determined by pilot studies to avoid ceiling or floor performance. The two central rows and columns were not used as either targets or foils.

Participants completed 384 discrimination trials broken up into the two blocks). Each block of 192 trials consisted of 12 unique trials in each row (and each column): the first and fourth stimuli in the row were compared four times, the fourth and seventh were compared four times, and the seventh and the tenth stimuli were compared four times (see Fig.1). Within each pair each stimulus was the target twice and with the target occurring equally often on the left and right position. These comparisons were made for 8 rows (excluding the middle two) and 8 columns (or rows), both parallel and perpendicular to the category boundary.

Each categorization trial consisted of a face stimulus appearing in the center of the screen. The two category labels appeared below the stimulus indicating which key ("q" or "p") should be pressed to indicate that category label. The assignment of labels to keys was randomized on each trial. After participants respond, feedback indicating the correct category label was presented for 1000 ms followed by a pause of 1000 ms between trials. Each non-grey stimulus from Fig. 1 was presented twice in random order during category training.

Results

Categorization Performance. A repeated measures ANOVA with block as a factor revealed a significant effect on categorization accuracy $F(7,546) = 21.75$, $p < .0001$, categorization accuracy improved throughout training.

A linear regression between distance to the center of the category space and categorization accuracy was performed separately for each dimension (category relevant and irrelevant). There was a significant improvement in accuracy for stimuli more distant on the category relevant dimension, $F(1,236) = 73.7$, $p < .0001$ but no significant change in categorization accuracy as a function of distance along the irrelevant dimension $F(1,236) = 1.47$, $p = .23$.

Discrimination Performance. All discrimination trials were coded in three ways. Half the trials varied along the category relevant dimension (perpendicular to the category boundary) and half along the irrelevant dimension. Discrimination trials were also coded on the relative extremeness of the target and foil objects: an equal number of trials were coded as "foil more extreme", "target more extreme" and "equal." Finally, for traditional CP analyses, the "foil more extreme" and "target more extreme" trials were grouped as Within trials, "equally" extreme trials were coded as Between trials. Between and Within trials could be relative to the category relevant or irrelevant dimension.

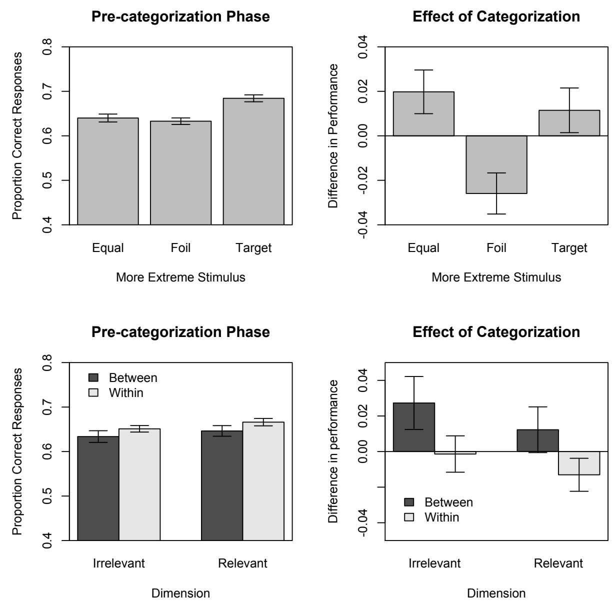

Fig. 2: Experiment 1 Results. Top-left: pre-categorization target-foil extremeness. Top-right: change in target-foil extremeness after categorization. Bottom-left: pre-categorization CP effects, split by dimension. Bottom-right: change in CP effects after categorization, also split. Error bars represent standard errors.

Pre-categorization phase. The graph in the top-left panel of Fig. 2 depicts the pre-categorization results, divided by extremeness condition. A 3 x 2 repeated measures ANOVA with relative extremeness (Equal vs. Foil vs. Target) and dimension (Relevant vs. Irrelevant) revealed a main effect of stimulus extremeness, $F(2,156) = 16.15$, $p < .0001$, but no main effect of dimension, $F(1,78) = 2.06$, $p = .16$, or interaction, $F\ (2,156) < 1$. Pairwise comparisons revealed that discrimination accuracy is higher when the target is the more extreme stimulus when compared to when the foil is more extreme, $p = .0001$, and when they are equally extreme, $p = .001$. The last two types of discrimination trials did not differ, $p = 1$. All p values were adjusted for multiple comparisons using a Bonferroni correction.

The results from the pre-categorization task considering the traditional CP analyses are depicted in the bottom-left panel of Fig. 2. A 2 x 2 repeated measures ANOVA with CP type (Within vs. Between) and dimension (Relevant vs. Irrelevant) as factors revealed a main effect of CP type, $F(1,78) = 4.34$, $p = .04$, with Within more accurate than Between, but no effect of dimension, $F(1,78) = 1.68$, $p = .20$, or interaction between the two variables, $F(1,78) < 1$.

Change in discrimination performance after learning. The pre-categorization analyses of extremeness and CP type were performed on the change in discrimination performance due to categorization. The change was computed by subtracting the pre-categorization discrimination accuracy from the post-categorization. The top-right panel in Fig. 2 depicts the results for the extremeness effects and the bottom right depicts the same results in terms of CP type.

A 3 x 2 repeated-measures ANOVA with relative extremeness and dimension as factors revealed a main effect of extremeness, $F(2,156) = 6.89$, $p = .001$ but no main effect of dimension, $F(1,78) = 1.33$, $p = .25$ or interaction, $F(2,156) < 1$. Pairwise comparisons revealed that performance changed equally for target more extreme and equally extreme, $p = 1$. The change in accuracy for the foil more extreme condition was significantly less than the other types: target more extreme ($p = .03$) and equal ($p = .005$).

To further investigate if accuracy performance improved with categorization, we performed a series of one-sample t-tests for each one of the extremeness conditions. The change in accuracy did not significantly differ from 0 for the target more extreme condition, $t(78) = 1.11$, $p = .27$ but was significantly lower for foil more extreme, $t(78) = -2.75$, $p = .007$. The change in accuracy of the equal condition was marginally greater than 0, $t(78) = 1.87$, $p = .06$.

A 2 x 2 repeated-measures ANOVA revealed a main effect of CP type, $F(1,78) = 5.75$, $p = .02$ but no main effect of dimension, $F(1,78) = 1.38$, $p = .24$, or interaction between the two variables, $F(1,78) < 0$ (Fig. 2, bottom-right).

Discussion

The results of Experiment 1 are not consistent with the hypothesis that category label ambiguity causes CP patterns. The pre-categorization phase in Experiment 1 indicates that the asymmetries seen in 2AFC tasks do not depend on the category or verbal codes assigned. More specifically, the results show that discrimination accuracy is higher when the target is more extreme than the foil alternative in the absence of any previous categorization learning. CP patterns were not observed before categorization despite the presence of the within-category asymmetry; in fact the reverse of the CP effect was marginally significant before categorization.

That the asymmetry exists before categorization suggests that it is a result of unsupervised learning processes rather than explicit category labels (Gureckis & Goldstone, 2008; Love et al., 2004). It remains unclear if the unsupervised mechanism is cluster labeling or perceptual change. We revisit this point in the general discussion.

Extremeness along the category relevant dimension predicted categorization accuracy but extremeness along the irrelevant dimension did not. This suggests that the asymmetry along the irrelevant dimension, both before and after categorization training, was not produced by differences in category labeling accuracy.

Categorization training did produce the expected CP effect: Between improved more than Within. This effect was modulated by an asymmetry among the Within trials, the foil more extreme trials showed decreased performance and the target more extreme showed significantly higher change. This asymmetry is consistent with the category label ambiguity hypothesis and occurred after category training.

The changes in discrimination performance differ between the relevant and irrelevant dimensions. This may have been due to the extensive opportunity for unsupervised learning of cluster structure during pre-categorization discrimination.

Experiment 2

One hypothesis that must be tested is if the asymmetric change in discrimination performance found after categorization training in Experiment 1 can be accounted for by the increased exposure to the stimuli instead of learning categories. This hypothesis is tested in Experiment 2, which is similar to Experiment 1 in that it consists of two critical blocks of discrimination judgments. However, another block of discrimination trials was substituted for the categorization task. Thus, by comparing performance in the first and last blocks of discrimination trials, which had roughly the same number of exposures to the stimulus as in Experiment 1, we can test the effect of experience with the stimulus space in the absence of categorization experience.

Method

Participants. 76 Indiana University undergraduate students participated in this experiment for course credit. Two participants were excluded from analyses because they did not conclude the experiment in the allotted time (60 min).

Procedure. This experiment followed the same general procedure of Experiment 1 except for the exclusion of the categorization phase. Participants completed 3 blocks of discrimination trials. Each block was identical to those in Experiment 1.

Results

All discrimination trials were coded similar to Experiment 1 but collapsed across dimension because no category boundary was learned. Discrimination trials were coded on the relative extremeness of the target and foil stimuli: an equal number of trials were coded as “foil more extreme”, “target more extreme” and “equal.” Finally, to compare to traditional CP analyses, the “foil more extreme” and “target more extreme” trials were grouped as Within trials, “equally” extreme trials were coded as Between trials.

1st discrimination block. The accuracy results for the first block of discrimination by extremeness condition are shown in the upper left corner of Fig. 3.

A repeated measures ANOVA revealed a main effect of stimulus extremeness, $F(2, 146) = 9.03$, $p < .0001$. Pairwise comparisons further revealed that trials in which the target was more extreme resulted in better discrimination than trials in which the foil was more extreme, $p = .006$, and also trials in which the two stimuli were equally extreme, $p = .01$. Finally, there is no difference in accuracy between trials in which the foil was more extreme and trials in which both stimuli were equally extreme, $p = 1$ (Fig. 3, top-left panel).

In the traditional CP analysis (despite no category training occuring), discrimination accuracy is higher for Within than Between discrimination trials, $t(73) = -2.01$, $p = .05$ (see Fig. 3, bottom-left panel).

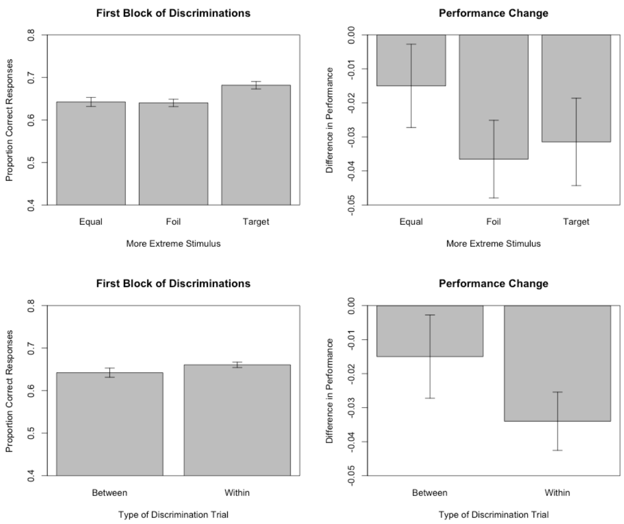

Fig. 3: Experiment 2 results. Top-left: Block 1 target-foil extremeness. Top-right: change in target-foil extremeness after prolonged exposure. Bottom-left: Block 1 CP effects. Bottom-right: change in CP effects after exposure. Error bars represent standard errors.

Change in discrimination performance. We computed the difference in accuracy between the last and first blocks to analyze the possible learning effect through successive exposure to discrimination trials. The top-right panel in Fig. 3 depicts the results considering the extremeness analysis while the bottom-right panel shows the results organized in terms of CP analyses.

A within-subjects ANOVA performed on these data revealed no main effect of stimulus extremeness, $F(2,146) = 1.22$, $p = .30$. Similarly, when analyzing the change in performance between the last and first blocks of discrimination trials for Within and Between discriminations (see bottom right panel of Fig. 3) there are no significant differences in performance between the two types of discrimination trials, $t(73) = -1.65$, $p = .1$.

Categorization (Exp. 1) vs. Exposure (Exp. 2) Categorization training (M = 0.011) led to significantly higher change in discrimination performance relative to exposure (M = -0.031) for target more extreme, $t(151) = 2.62$, $p = .01$, as well as for equal trials, (cat. M = 0.020, exp. M = -0.015, $t(151) = 2.16$, $p = .03$). On the contrary, there was not a significant difference between the change in discrimination accuracy for categorization (M = -0.026) and exposure (M = -0.036) for foil more extreme trials, $t(151) = 0.72$, $p = .47$).

Discussion

The results from the first block of discrimination trials replicate those found in the pre-categorization phase of Experiment 1. The asymmetry between the target and foil more extreme trials existed without category training and when CP patterns were not found.

Without category training, however, comparing the first and last blocks of discrimination in Experiment 2 did not show a change in performance consistent with the CP effect or a change in the difference between equal, target or foil more extreme trials. Performance for all trial types decreased in a consistent way across all trial types. This is likely due to fatigue considering the great number of trials participants complete without any feedback.

Finally, the categorization in Exp. 1 resulted in significantly different performance change for equal and target more extreme trials than what was seen with exposure alone (Exp. 2). However, this was not the case for foil more extreme trials. This suggests that category training improves discrimination for between-category judgments as well as for within-category judgments in which the target is more extreme than the foil.

General Discussion

The presence of the within-category asymmetry before categorization and for each of the clusters refutes the hypothesis that the asymmetry alone can account for CP patterns or that the asymmetry is a direct result of explicit category labels. Instead these results are consistent with an unsupervised learning mechanism that is sensitive to the distribution of items within categories (Love et al., 2004; McDonnell & Gureckis, 2011) and a decision process for discriminations that distorts extreme exemplars to produce category caricatures in the distribution of items (Goldstone, 1996; Goldstone et al., 2003; Roberson et al., 2007).

The change in discrimination performance after categorization shows an increase in the asymmetry in Experiment 1 but not in Experiment 2. The fact that the asymmetry increases on the category relevant dimension as well as the irrelevant dimension is a challenge to the category label ambiguity hypothesis (Hanley & Roberson, 2011). To account for this behavior, the labeling hypothesis must be expanded to allow individual clusters within categories, learned via unsupervised mechanisms, to be assigned unique labels as in SUSTAIN (Love et al., 2004) or other semi-supervised learning models (McDonnell & Gureckis, 2011).

However, the fact that the effect of extremeness in a 2AFC task is observed before any category learning has taken place points to a biasing effect of extremeness within a stimulus set rather than category learning *per se*. Consequently, the relative change in performance seen after category learning might result from category learning processes that produce warped caricatures by shifting perceptual representations toward the stimulus extremes (Goldstone, 1996; Goldstone et al., 2003). This account is consistent with the relative improvement in between-category (and cluster) judgments as well as target more extreme judgments after categorization relative to exposure alone for both the relevant and irrelevant dimensions.

We believe the strongest message from this work is the critical importance of measuring the change in perceptual discrimination performance to understand the learning mechanisms that underlie CP. Going forward, we plan to expand this analysis to look at changes in within-category asymmetries under conditions of verbal interference that may impair label usage (Hendrickson et al., 2010; Roberson et al., 2007) and formalize the unsupervised learning predictions in an extension of the SUSTAIN computational modeling framework (Love et al. 2004; Gureckis & Goldstone, 2008).

Acknowledgments

This research was supported in part by National Science Foundation REESE grant 0910218 and Department of Education IES grant R305A1100060. PFC was supported by Graduate Fellowship SFRH/BD/68554/2010 from the Portuguese Foundation for Science and Technology (FCT). All data was collected by Jeremy Falkmann.

References

Ashby FG, Maddox WT. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*.

Goldstone, R. (1994). Influence of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*.

Goldstone, R. L. (1996). Isolated and Interrelated Concepts. *Memory & Cognition*.

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical Perception. *Interdisciplinary Reviews: Cognitive Science*.

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*.

Gureckis, T. M., & Goldstone, R. L. (2008). The effect of internal structure of categories on perception., *Proceedings of the 30th Annual Conference of the Cognitive Science Society*.

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

Harnad, S. (Ed.). (1987). *Categorical perception: The groundwork of cognition.* New York: Cambridge University Press.

Hendrickson, A. T., Kachergis, G., Gureckis, T. M., & Goldstone, R. L. (2010). The effect of verbal interference and the internal structure of categories on perceptual discrimination., *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in the name? Categorical perception of unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review*.

Kikutani, M., Roberson, D., & Hanley, J. R. (2010). Categorical perception for unfamiliar faces: Effect of covert and overt face learning. *Psychological Science*.

Kay, P. and Regier, T. (2003) Resolving the question of color naming universals. *Proceedings of the National Academy of Science*.

Kayser, A. (1997). *Heads.* New York: Abbeville Press.

Landy, D., & Goldstone, R. L. (2005). How we learn about things we don't already understand. *Journal of Experimental and Theoretical Artificial Intelligence*.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*.

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

Logan, J., Lively, S., & Pisoni, D. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*.

McDonnell, J. and Gureckis, T.M. (2011). Adaptive Clustering Models of Categorization. in *Computational Models of Categorization*. edited by Pothos and Willis, Cambridge University Press, Oxford, UK.

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

Petrov, A., Dosher, B., & Lu, Z. L. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological Review*.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*.

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition*.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. (2005). Colour categories in Himba: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*.

Roberson, D., Davies, I. R. L., & Davidoff, J. (2000). Colour categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*.

Sauter, D. A., LeGuen, O., & Haun, D. B. M. (2011). Categorical Perception of Emotional Facial Expressions Does Not Require Lexical Categories. *Emotion*.

Steyvers, M. (1999). Morphing techniques for generating and manipulating face images. *Behavior Research Methods, Instruments, & Computers*.

Webster, M. A., & Kay, P. (2011) Color categories and color appearance. *Cognition*.