Categorical perception effects reflect differences in typicality on within-category trials

J. RICHARD HANLEY AND DEBI ROBERSON

University of Essex, Colchester, UK.

Keywords: Categorical perception, color, face identity, facial expressions of emotion

Running head: Categorical perception

Address for correspondence:
Prof. D Roberson
Dept. of Psychology
University of Essex
Wivenhoe Park
Colchester
Essex CO4 3SQ
UK
Email: robedd@essex.ac.uk

ABSTRACT

Many studies have shown better discrimination of two stimuli that cross a category boundary than of two stimuli belonging to the same category. This finding, known as categorical perception, is generally assumed to reflect consistently good performance on cross-category relative to within-category trials. However, (Roberson, Damjanovic, & Pilling, 2007) revealed that performance on within-category pairs of morphed facial expressions matched performance on cross-category trials when the target was a good exemplar of its category. Here we investigate the generality of that finding by conducting new analyses of data from a series of studies of categorical perception in facial identity and color domains with speakers of different languages. Consistent with Roberson et al. (2007), the new analyses demonstrate that performance for central targets on within-category trials is as good as performance on cross-category trials. Participants perform badly on within-category items only when the target is closer to the category boundary than the distractor. These results provide no support for the view that categorical perception is associated with increased perceptual sensitivity at category boundaries.

Categorical perception (CP) is manifest behaviorally as a greater ability to distinguish two stimuli that cross a boundary between categories than stimuli that both belong to the same category. CP has been investigated for auditory continua such as musical notes (e.g. (Howard, Rosen, & Broad, 1992; Locke & Kellar, 1973); and phonemes (e.g. (Liberman, Harris, Hoffman, & Griffith, 1957; Repp, 1984), for natural visual continua such as color (e.g. (Gilbert, Regier, Kay, & Ivry, 2006); and for morphed continua between face identities or facial expressions (Angeli, Davidoff, & Valentine, 2007; Gilbert, et al., 2006; Goldstone & Hendrickson, 2009; Huette & McMurray, 2010; Liberman, et al., 1957). It has even been found for morphed continua between human and ape faces (Campbell et al., 1997), and can also be trained for novel category boundaries (Goldstone, 1994, 1998; Özgen & Davies, 2002).

Many studies have used a two-alternative forced choice (2-AFC) procedure to demonstrate CP across different perceptual domains, and it is performance in the 2-AFC task that we investigate in this article. Participants view a target such as a colored patch for a short duration. Shortly afterwards, the original item is displayed next to a distractor and the participant has to indicate which item was presented a few moments earlier. Performance is significantly faster and more accurate when target and distractor are members of different categories than when they both belong to the same category even though the physical separation of the members of each pair is always equated.

One important reason for using this procedure is that it can provide evidence for the existence of categories and for boundaries between those categories. For example, Roberson, Davies and Davidoff (2000) found that speakers of Berinmo showed CP in the 2-AFC task at a boundary between two color categories that does not exist for speakers of English. This finding provides evidence against the idea that color categories are universal and exist irrespective of the color vocabulary that speakers have acquired during language development. Similarly, Kikutani, Roberson and Hanley (2008, 2010) used evidence of categorical perception in the 2-AFC task to discover differences between conditions in which categorical representations for unfamiliar faces do and do not appear to develop.

The focus of this paper, however, is an attempt to understand the reasons why CP is observed in the 2-AFC task. One possibility is that CP has a perceptual basis (Harnad, 1987). Perceptual systems might be fine-tuned by experience to be more sensitive to change at category boundaries (acquired distinctiveness) and/or to be less sensitive to within-category change (Goldstone, Lippa, & Shiffrin, 2001). Roberson, Hanley and Pak (2009) investigated this issue in the color domain with English and Korean speakers who have either one linguistic category boundary in the blue-green region (English) or two (Korean). They found no evidence that participants' absolute discrimination thresholds for just-noticeable differences (JNDs) had become more sensitive at the boundary between these color categories than they were at category centers in either population (see also Pinto Kay and Webster, in press). Nevertheless, a number of studies have observed CP at the boundaries between these same color categories using pairs of well-separated stimuli that could easily be distinguished (Gilbert et al., 2006; Pilling, Wiggett, Özgen, Davies, 2003; Roberson & Davidoff, 2000). CP effects thus appear to occur even when there is no evidence of increased perceptual sensitivity at category boundaries.

Roberson and Hanley (2009, 2010) instead proposed that CP for color in the 2-AFC task occurs because of the effectiveness of categorical codes in cross-category conditions. They argued that, in principle, participants could perform the 2-AFC task using either categorical or perceptual information about the target item. Their explanation therefore bears some similarity to the dual code model originally proposed by Pisoni and colleagues to explain CP in speech perception (Pisoni & Lazurus, 1974; Pisoni & Tash, 1974), that was subsequently applied to color by (Bornstein & Korda, 1984). The categorical code contains less information but may be easier to retain than the perceptual code unless verbal interference takes place during the retention interval. Targets and distractors can be easily distinguished when they cross a category boundary because they differ at both the categorical and the perceptual level. For within-category pairs, however, the categorical code cannot distinguish target and distractor. They can be distinguished only if the participant retains an accurate representation of the target's perceptual features. This account can also readily explain Roberson and Davidoff's (2000) finding that when verbal interference was interposed between presentation of the target and the test pair, CP was abolished both for colors and for facial expressions. Presumably verbal interference disrupts participants' ability to retain the category label, and so performance for both within and cross-category targets relies on the retention of a visual representation.

There is, however, a finding reported by Roberson, Damjanovic and Pilling (2007) in a study of CP for facial expressions that is less easy to explain in terms of either a perceptual account or the categorical account suggested by Roberson & Hanley (2010). Most published studies that have used the 2-AFC task have reported accuracy or response time for the cross-category trials compared to aggregated within-category trials. However, Roberson et al. (2007) also analyzed performance on different types of within-category pairs. On half of the trials, the target (T) was a good exemplar of its category (further from the boundary than the distractor), while the distractor (D) was a less good exemplar (e.g. T = 90% Fearful; D = 70% Fearful). On the remainder, the poor example (closer to the boundary than the distractor) was the target (e.g. T = 70% Fearful; D = 90% Fearful). Roberson et al.'s analysis revealed an asymmetric pattern *within the same pairs of stimuli* whereby performance was significantly better when the target was the good exemplar than when it was the poor exemplar. Moreover, accuracy for good targets on within-category trials was not statistically different from accuracy on cross-category trials. Only on within-category trials in which the target was a poor exemplar of the category was identification significantly worse than for cross-category pairs (see figure 1).

(figure 1 about here)

Clearly, the CP effect in Roberson et al.'s (2007) study did not result from greater perceptual sensitivity to change around the category boundary because the same within-category pairs were sometimes discriminated just as accurately as cross-category pairs and sometimes significantly worse, depending only on which member of the pair was the target and which the distractor. Since the effect occurred within the same pairs of stimuli, it could not have been caused by unequal separations of within- and cross-category pairs. Neither can the account put forward by Roberson and Hanley (2009, 2010) easily accommodate such good performance on half the within-category trials. On that view, the availability of categorical coding should produce consistently better performance on cross-category pairs even when the within-category target is closer to the category center than the distractor.

It is therefore important to investigate whether Roberson et al.'s (2007) findings apply in domains other than facial expression judgments for both theoretical and methodological reasons. If so, there would be important implications for most published theoretical accounts of CP. It would also follow that future studies of CP should investigate whether performance is consistently poor on all within-category comparisons, or only on those where the target is a more peripheral category member than the distractor.

To address this issue, we conducted new analyses of performance on within-category pairs in a series of published experiments by Roberson and her colleagues that used the 2-AFC task to investigate CP. The new analyses compare accuracy on within-category trials where performance with good exemplars (GE) and poor exemplars (PE) had previously been aggregated. They reveal a strikingly consistent pattern of performance in a range of different populations across the domains of color, facial expression and face identity. As in Roberson et al. (2007), the GE and PE pairs always contained *the same* stimuli but differed in terms of which member of the pair was the target (T) and which the distractor (D). This applies to all the experiments described below. Exact details of the stimuli used in each experiment can be found in the original published reports.

Study 1: Color discrimination in English speakers (Roberson and Davidoff, 2000). Roberson and Davidoff (2000) used the 2-AFC task to demonstrate CP for color categories in 72 speakers of English at three retention intervals by showing significantly better performance on 32 cross-category than on 32 within-category items. A new 3 x 3 ANOVA compared performance on the 32 cross-category items (CR) with the 16 within-category central targets (GE) and the 16 within-category boundary targets (PE) following retention intervals of 10 seconds, 5 seconds, or 0 seconds (see figure 2).

(figure 2 about here)

There was a significant effect of target type [$F_{(2,138)} = 21.66$, MSe = .06, $p<.01$] but no effect of interval [$F_{(2,69)}<1$] and no significant interaction [$F_{(2,138)}<1$]. Newman-Keuls pair-wise comparisons showed no significant difference between performance on CR and GE targets, but significantly worse performance on PE targets than on CR and GE targets (both $p<.01$, see figure 2).

So, as in Roberson et al. (2007), poor performance on within-category pairs when the target item was closer to the boundary than the distractor accounted for all of the difference between cross- and within-category recognition rates (the hallmark of CP). The effect that Roberson et al. reported with facial expressions clearly occurs also with color recognition.

The only other CP study to have examined performance on within-category items also investigated color discrimination. It was carried out by Pilling, Wiggett, Özgen, & Davies (2003) who failed to find any significant differences in performance on within-category pairs. However, the findings from the two studies are not directly comparable. In Roberson and Davidoff's (2000) experiments, stimuli varied only in hue so their relative position between the category boundary and the category center fell along a linear continuum. Pilling et al.'s stimuli varied in both hue and lightness. Pilling et al.'s participants likely integrated the two varying dimensions, reducing the overall difference in their perceived distance from the category center for pairs lighter than best examples. So, the physical differences between the targets that were closer to and further from the boundary on within-category trials were non-linear. As a consequence, Pilling et al.'s design may not have been quite sensitive enough to detect differences between performance on within-category targets.

Study 2: Color discrimination in Berinmo speakers (Roberson, Davies, & Davidoff, 2000).

Roberson et al. (2000) investigated CP for color in speakers of Berinmo, a traditional hunter-gatherer culture in the upper Sepik region of Papua New Guinea. The Berinmo have a different set of basic color terms from speakers of English. 8 Berinmo speakers showed significantly better performance on 32 cross-category items than on 32 within-category items at the boundary between two Berinmo color categories (*nol* and *wor*) that does not exist for English speakers.

(figure 3 about here)

A new one-way ANOVA compared performance by Berinmo participants on the 16 within-category trials where the target was a better example of the category than the distractor (GE) with the 16 trials where distractor was a better example of the category than the target (PE). There was a significant effect of target type [$F_{(2,22)} = 7.97$, MSe = .032, $p<.01$]. A Newman-Keuls pair-wise comparison showed that the Berinmo recognized both CR and GE targets significantly better than PE targets (both $p<.05$). They showed no significant difference between performance on CR and GE targets (see figure 3).

Study 3: Color discrimination in Himba speakers (Roberson, Davidoff, Davies, & Shapiro, 2005).

Roberson et al. (2005) investigated CP for color in Himba participants who are semi-nomadic cattle-herders in northern Namibia in South-West Africa. The Himba also have a different color vocabulary from English speakers, and once again CP was demonstrated at a color boundary that does not exist in English (Roberson et al., 2005). A new ANOVA that compared performance by the 12 Himba participants on 32 CR pairs, 16 GE pairs and 16 PE pairs showed a significant effect [$F(2,22) = 9.343$, MSe = .026, $p<.01$]. Newman-Keuls comparisons showed that only on PE targets (where the target was closer to the category boundary and the distractor closer to the category centre) was performance significantly worse than in the cross-category condition ($p<.01$). These results are illustrated in figure 4.

It is clear from these new analyses of the CP data from Himba and Berinmo participants that the within-category effect reported by Roberson et al. (2007) generalizes to members of different cultures and to category boundaries that do not exist for speakers of English.

(figure 4 about here)

Study 4: Facial identity (Kikutani, Roberson, & Hanley, 2008).

Kikutani et al. (2008) demonstrated categorical perception of pre-experimentally familiar faces using morphs of 2 well-known celebrities. They reported that performance was better in the 2-AFC task when the target and distractor crossed the category boundary between the 2 face identities than when both morphs were from within the same face category. Similar results were obtained with pre-experimentally unfamiliar faces so long as the faces and their names were learnt before the 2-AFC task began. All of Kikutani's experiments employed 16 cross-category trials, 12 central within-category trials (GE) and 12 boundary within-category trials (WR) for both famous and unfamiliar face continua.

In a new analysis of the CP data from Kikutani et al.'s (2008) first experiment that found CP only for familiar faces, the 20 participants showed a significant effect of target type [$F_{(2,38)}$ = 6.93, MSe = .027, p<.01]. A Newman-Keuls pair wise comparison showed no significant difference between CR and GE targets, but both CR (p<.01) and GE targets (p<.01) were recognized significantly better than PE targets.

In Kikutani et al.'s (2008) second experiment, pre-training with names successfully induced CP for novel as well as for familiar faces. A new 2 (Face type: familiar vs. novel) x 3 (Target type: CR vs. GE vs. PE) repeated-measures ANOVA with 22 participants revealed a significant effect of Face type [F(1,21) = 14.74, MSe = .02, p<.01], as well as a significant effect of Target [F(2,42) = 19.26, MSe = .04, p<.01], but no significant interaction. A Newman-Keuls pair wise comparison showed no significant difference between CR and GE targets for either familiar or novel faces. Both CR targets and GE targets were recognized significantly better than PE targets for familiar (p<.01), and for novel faces (p<.05) (see figure 5).

(figure 5 about here)

The results of these new analyses of Kikutani et al.'s data display the same pattern of performance for both familiar and newly learned face identities as was seen in the color studies. Because it applies to faces, the effect is clearly not restricted to categories that are established early in life.

Study 5: Facial identity (Kikutani, Roberson, & Hanley, 2010).

Kikutani et al. (2010) investigated CP for previously unfamiliar faces following either an overt or a covert learning procedure. Of particular interest is an analysis of data from a 2-AFC task that followed covert training for two previously unfamiliar faces (Experiment 1). 20 participants completed a pre-experimental classification task in which they judged 300 faces to be either famous or unfamiliar. Two unfamiliar faces were repeated 20 times at random intervals. Participants showed evidence of learning these faces, becoming significantly faster at correctly rejecting them across the course of the session, but this was not enough to establish CP in a subsequent 2-AFC test. A new 2 (Face type: familiar vs. novel) x 3 (Target type: CR vs. GE vs. PE) repeated-measures ANOVA revealed no significant difference between familiar and unfamiliar faces [F<1], but a significant effect of Target [F(2,38) = 6.37, MSe = .02, p<.01], and a significant interaction [F(2,38) = 7.22, MSe

= .04, p<.01]. A Newman-Keuls pair wise comparison of the interaction showed no significant difference between CR and GE targets for either familiar or novel faces. For familiar faces, both CR and GE targets were recognized significantly better than PE targets (p<.01). For unfamiliar faces, the difference between CR and PE targets failed to reach significance (figure 6).

In contrast, in an overt learning condition (Experiment 2), training did successfully induce CP for novel faces. Not only did the ANOVA reveal no significant effect of Face type [F(1,27) = 1.18, MSe = .03, p>.1], a significant effect of Target [F(2,54) = 12.87, MSe = .04, p<.01], and no significant interaction [F<1], but now there was no significant difference between CR and GE targets for either familiar or novel faces. Both targets were recognized significantly better than PE targets (p<.01). These results are illustrated in figure 6.

(figure 6 about here)

Study 6: The effects of verbal interference (Roberson and Davidoff, 2000; Roberson et al., 2007)

Given that it abolishes the advantage for cross-category pairs over within-category pairs, it follows that verbal interference should remove differences in performance between PE items and GE items. We therefore conducted a new analysis of the effect of verbal interference on accuracy for colors in Roberson and Davidoff (2000), and on accuracy for facial expressions in Roberson et al. (2007).

A new 2 (Interference: none vs. verbal) x 3 (Pair type: cross-category vs. within-central target vs. within-boundary target) ANOVA on Roberson and Davidoff's (2000) color data revealed a significant interaction [F(2,46) = 4.08, MSe = .04, p<.05]. The accuracy advantage for CR and GE items relative to PE items in standard conditions disappeared under verbal interference. In fact, tests of simple main effects showed that whereas cross-category pairs accuracy decreased significantly under verbal interference [F(1,23) =8.43, MSe =.03, p<.05], accuracy for boundary targets showed a trend towards *improvement* [F(1,23) =3.53, MSe =.03, .05 < p <.08].

(figure 7 about here)

A similar analysis of Roberson et al.'s (2007) data that investigated the effect of verbal interference on facial expressions also revealed a significant interaction between target type and interference [F(2,34) = 5.11, MSe = .01, p<.05]. Once again, the accuracy advantage for CR and GE items relative to PE items in standard conditions disappeared under verbal interference. Under verbal interference, accuracy for cross-category pairs again decreased significantly [F(2,34) = 3.41, MSe = .02, p<.05], while accuracy for boundary targets again showed a trend towards improvement [F(2,34) = 3.09, MSe = .02, .05 < p <.07] (see figures 7 and 8).

(figure 8 about here)

Discussion

The new analyses reported here reveal that cross-category targets were not consistently better identified than within-category targets. Instead, there was a tendency to misidentify boundary targets in within-category pairs when distractors were closer to the category center than targets. As noted above, this effect could not have resulted from differences in the physical separation of pairs of stimuli, because it was found within the same pair of stimuli, depending on which member of the pair was the target and which the distractor. The asymmetries of performance for within-category pairs were observed in several different perceptual domains and for members of a diverse range of cultures. When verbal interference was introduced, the effect disappeared, suggesting that poor performance on boundary targets in within-category pairs cannot be a perceptual phenomenon, but results specifically from the encoding of category labels.

Nevertheless, as noted earlier, these results are problematic for the suggestion (Roberson and Hanley, 2009, 2010) that CP occurs because a categorical code distinguishes cross-category targets from distractors, but cannot distinguish within-category targets from distractors. In that case, availability of categorical coding should produce consistently better performance on cross-category pairs than within-category pairs even when the target is closer to the category center than the distractor. Below we put forward a modified account of the way in which participants might use categorical codes during the 2-AFC task that can explain performance on within-category trials. This explanation can be seen as building on Pisoni and Tash's (1974) model of categorical perception of speech by incorporating the idea that a categorical code will prove most effective when the target is a good exemplar of the category.

Categorical accounts of CP in the 2-AFC task assume that at test participants do not always have available an accurate representation of the visual/auditory features of the target. In the absence of perceptual information, they must rely instead on a categorical code to select the target item. The new explanation derives from the fact that central targets are categorized and named faster, more accurately and more consistently than less central targets (Agrillo & Roberson, 2009; Brown & Lenneberg, 1954; Huette & McMurray, 2010; Rosch Heider, 1975). If a target is named correctly at presentation and at test on a within-category trial, but the distractor is labeled incorrectly, then the target can be selected reliably on the basis of the categorical code alone. If a target is named inconsistently at encoding and test, then the categorical code will be an unreliable guide to which is the target item. Since peripheral targets in within-category pairs are less likely to be categorized consistently at presentation and test, the categorical code is less likely to be reliable than it is for central targets. Hence performance will be poorer for the peripheral than the central targets.

Furthermore, this asymmetry is likely to be exacerbated in the 2-AFC task by a context effect. In the case of color categorization, it is known that an ambiguous color on the boundary between pink and yellow is likely to be named “pink” when it appears in the context of a flamingo and “yellow” when it appears in the context of a school bus (Kubat, Mirman, & Roy, 2009; Mitterer, Horschig, Müsseler, & Majid, 2009). Crucially, when a boundary color appears against a background that is a good example of that color category, it is more likely to be given a different name from the background than when it appears alone (Hansen, Walter, & Gegenfurtner, 2007). It follows that when a good target appears next to a boundary distractor in the 2-AFC test, it is more likely that the distractor will be given a different category label from the target. This bias will improve performance when the target is a good exemplar as the distractor can be rejected because it elicits the wrong category label. However, this bias will reduce performance when the target is a boundary item. For instance a greenish-blue boundary target may be categorized as ‘blue’ when it appears on its own. However the presence of a distractor that is a good example of blue at test may mean that the boundary target elicits a different label (green) in the AFC test. The outcome will be that the category label given to the target at encoding is now elicited only by the distractor, which will be selected in preference to the target as a consequence.

It also follows from this account that performance on boundary targets might actually improve under verbal interference (see figures 7 and 8). This is because the proposed bias against selecting the more peripheral member of a pair at test will only apply when a participant has access to a categorical code. In this regard, it is also interesting to note that in other studies where verbal interference has abolished CP for color, an unexplained improvement in within-category performance has occurred (Gilbert, et al., 2006; Winawer, et al., 2007).

Bearing in mind the results of the analyses reported above, it is important that other researchers also investigate whether better discrimination of cross category pairs comes about because performance was particularly poor when the target was a more peripheral category exemplar than the distractor. For example, Gureckis and Goldstone (2008) reported a study that investigated CP for previously unfamiliar faces using the 2-AFC task. Participants were better able to distinguish faces from different perceptual categories after than before training. Participants were also worse at distinguishing faces from the same category after than before training. These results were observed regardless of whether the categorical distinction was associated with a categorical label during training. Gureckis and Goldstone demonstrated that their findings could be simulated by connectionist models which assumed that a layer in the network had been reorganized by the demands of categorization. However, a quite different type of explanation would be required for these data if it turns out that they came about because discrimination was poorer following categorical training only when the target was a more peripheral category member than the distractor.

In conclusion, our new analyses of performance on the 2-AFC task consistently indicate that good performance is not confined to items that cross category boundaries. Discrimination of within-category pairs is sometimes as good as that for cross-category pairs. The results can be explained by our new account of performance in the 2-AFC test which claims that more peripheral category members are less likely to be coded consistently at presentation and test. If this account is correct, however, CP appears to have come about because of task-specific categorization strategies that participants adopted to retain information about a stimulus they had just been shown. These CP effects were not the result of any increased perceptual discrimination at category boundaries. This outcome reinforces Roberson et al.'s (2009) conclusion that there is no evidence of increased perceptual sensitivity at locations in the color spectrum where CP has previously been observed.

Researchers investigating perception of speech stimuli have also expressed worries about the generality of CP effects. One reason is that relatively minor modifications to the experimental task (such as changes in the auditory quality of stimuli) can make CP effects for phonemes suddenly disappear (Gerrits & Schouten, 2004; Schouten, Gerrits, & van Hessen, 2003). When task variables such as these are controlled, there is limited evidence of discontinuous processing of speech stimuli around category boundaries. Experimental demonstrations of CP can provide important information about the presence of perceptual categories in different cultures (e.g. Roberson et al., 2000; Winawer et al., 2007) and the circumstances in which they are acquired (Kikutani et al., 2010: Roberson et al., 2004). However, the present study raises similar concerns as to whether investigations of CP are likely to yield deeper insights into processes that are used in normal circumstances to perceive color or faces.

Acknowledgements:

We would like to thank Bill McMurray and Cathleen Moore for their helpful comments on an earlier version of this article.

References:

Agrillo, C., & Roberson, D. (2009). Colour language and colour cognition: Brown and Lenneberg revisited. . *Visual Cognition, 17*, 412-430.

Angeli, A., Davidoff, J., & Valentine, T. (2007). Face familiarity, distinctiveness and categorical perception. *The Quarterly Journal of Experimental Psychology, 61*, 690-707.

Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research, 46*, 207-222.

Brown, R., & Lenneberg, E. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology, 49*, 454-462.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*(3), 363-376.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the USA, 103*, 489-494.

Goldstone, R. L. (1994). Influences of Categorization on Perceptual Discrimination. *Journal of Experimental Psychology-General, 123*, 178-200.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

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

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

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4).

Harnad, S. (Ed.). (1987). *Psychophysical and cognitive aspects of categorical perception: A critical overview.* Cambridge: Cambridge University Press.

Howard, D., Rosen, S., & Broad, V. (1992). Major minor triad identification and discrimination by musically trained and untrained listeners . *Music Perception, 10*(2), 205-220.

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

Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in the name? Categorical Perception of unfamiliar faces can occur through labelling. *Psychonomic Bulletin & Review,, 15*, 787-794.

Kikutani, M., Roberson, D., & Hanley, J. R. (2010). Categorical perception for unfamiliar faces: effect of covert and overt face learning. . *Psychological Science* (21), 865-871.

Kubat, R., Mirman, D., & Roy, D. (2009). *Semantic context effects on color categorization.* Paper presented at the 31st Annual Cognitive Science Society Meeting, Austin, TX.

Liberman, A. N., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Locke, S., & Kellar, L. (1973). Categorical perception in a nonlinguistic mode. *Cortex, 9*(4), 355-369.

Mitterer, H., Horschig, J. M., Müsseler, J., & Majid, A. (2009). The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1557-1562.

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General, 131*, 477-493.

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

Pinto, L., Kay, P., & Webster, M. A. (in press). Perceptual grouping and categorical color perception. *Journal of Vision*.

Pisoni, D. B., & Lazurus, J. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55*, 328-333.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*, 285-290.

Repp, B. H. (1984). The role of release bursts in the perception of S -Stop clusters I. *Journal of the Acoustical Society of America, 75*(4), 1219-1230.

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition, 35*, 1814-1829.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*, 977-986.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. (2005). Colour categories in Himba: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50*, 378-411.

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, 129*, 369-398.

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

Rosch Heider, E. (1975). Cognitive Reference Points. *Cognitive Psychology, 7*, 532-547.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*, 70-80.

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 of Sciences of the United States of America, 104*(19), 7780-7785.