A Bayesian Approach to Understanding the Perceptual Magnet Effect in Color Perception

Rachel Chen (qc898@nyu.edu)
Department of Psychology, New York University

Helen Hu (fh986@nyu.edu)
Department of Psychology, New York University

Avery Liang (ll5234@nyu.edu)
Department of Psychology, New York University

Abstract

The perceptual magnet effect describes the phenomenon where the perceptual space is shrunken near an existing category, such that stimuli close to a prototype are perceived as more similar to the prototype than they actually are. This study extends the perceptual magnet effect from the well-studied domain of speech perception to color perception. We propose that, as in speech perception, the perceptual magnet effect can be explained by a Bayesian model which assumes that individuals use their knowledge of color category to optimally infer the specific shade of a color while compensating for uncertainty in the sensory signal. In two experiments, we investigated the influence of the green component (G-value) in the RGB color space on color categorization and discrimination. Our results demonstrate that the perceptual magnet effect is strongest near the centers of color categories and weakest at the category boundaries. The Bayesian model accurately predicted participants' ratings for their perceived difference between colors, highlighting the role of category knowledge in perceptual inference. These findings suggest that the perceptual magnet effect is not limited to speech perception but can also be applied to color perception, emphasizing the generalisability of Bayesian approaches to understanding human cognition. Our study provides novel insights into the interaction between color categories and the influence of category structure on color perception.

Keywords: perceptual magnet effect; color perception; color categorization

Introduction

Our perceptual experience is not a mere reflection of the physical world, but is profoundly shaped by the categorical knowledge we acquire throughout our lives (Feldman, Griffiths, & Morgan, 2009). The process of organizing stimuli into distinct categories fundamentally alters the way we perceive these stimuli, a phenomenon known as categorical perception. This effect is widespread in cognition and perception and is characterized by both an enhanced discrimination across category boundaries and a reduced discrimination between within-category stimuli (Feldman et al., 2009). Categorical perception has been widely observed across various domains, including the perception of speech sounds (Liberman, Harris, Hoffman, & Griffith, 1957), colors (Bornstein & Korda, 1984) , facial expressions (Etcoff & Magee, 1992), and even artificially created categories (Goldstone, 1994).

A prominent example of categorical perception in speech is the perceptual magnet effect, first reported by Kuhl et al. (1991) in vowel perception. In this study, participants were asked to provide ratings of category goodness for over 100 synthesised /i/ sounds. As presented in Figure 1, two vowels were identified as a result of these ratings: one that listeners consider as the best instance for the /i/(referred to as the prototype, P), and one that was consistently judged as a relatively poor exemplar of an /i/ vowel (referred to as the nonprototype, NP). Kuhl and colleagues then constructed a set of variants surrounding P and NP across the vowel space, and asked listeners to 1) rate the category goodness of the variants on the scale from 1 to 7, and 2) discriminate between pairs of vowel sounds that varied in their proximity to prototypical vowel categories. They found that the perceived category goodness of /i/ vowels declined systematically as stimuli were further away removed from P. More importantly, discrimination was found to be poorer for sounds near P compared to sounds further away from P, and that this generalization was greater around P than NP. Taken together, these results supported the hypothesis that the distance between P and surrounding members is effectively decreased. Viewing phonetic discrimination in spatial terms, the perceptual space appears to be "warped", effectively pulling category members towards the prototype. The perceptual magnet effect has been replicated and extended in multiple studies (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Iverson & Kuhl, 1995), establishing it as a robust phenomenon in speech perception.

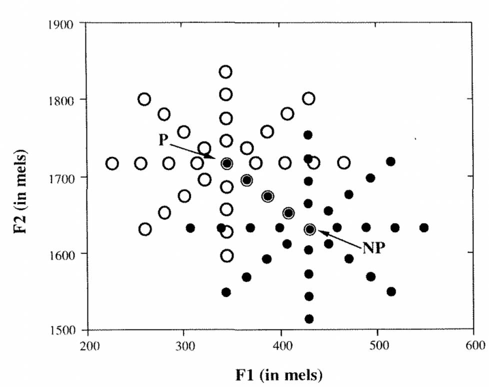

Figure 1: The prototype /I/ (P) and its 32 variants (open circles) and the nonprototype (NP) and its 32 variants (closed circles).

Using a Bayesian framework, Feldman et al. (2009) has provided a computational account of the perceptual magnet effect. In this model, the listener's goal is to infer the intended target production of a speaker based on the noisy speech signal they receive. The model assumes that listeners have prior knowledge of the distribution of speech sounds within each phonetic category, which is modeled as a Gaussian distribution. The listener then uses this prior knowledge to optimally infer the intended target production. The key insight of the model is that the optimal inference is biased towards the center of the inferred category (i.e., the prototype) due to the uncertainty in the speech signal. As a result, perceptual magnet effect arises because this bias is stronger for sounds near the category center and weaker for sounds near the category boundary, leading to a warping of perceptual space that mirrors the perceptual magnet effect (see Figure 2).

Building on this work, the current project aims to extend the Bayesian framework to explain categorical effects in color perception. Like speech sounds, colors are organized into distinct categories (e.g., red, blue, green) with a graded structure, and cross-linguistic studies have provided compelling evidence that color perception is influenced by linguistic categories. For example, Roberson et al. (2000) found that speakers of Berinmo, a language that makes a categorical distinction between nol (yellowish-green) and wor (greenish-blue), showed better discrimination of colors across this boundary compared to English speakers. Similarly, in the classic experiment by Winawer et al. (2007), Russian speakers were found to be faster at discriminating different shades of blue when they fall into different linguistic categories in Russian (volubly and siniy) than when they were from the same category, whereas English speakers showed no such category advantage. Despite the clear category effect in color perception, the existence and nature of a perceptual magnet effect in color perception remains unclear. Like speech perception, color perception is inherently ambiguous because we need to infer the true reflective properties of surfaces in different environment and lighting conditions. Therefore, we propose to study perceptual magnet effect in color perception by extending Winawer et al.'s (2007) color discrimination paradigm to the blue-turquoise distinction. If a Bayesian model, similar in structure to the model of Feldman et al. (2009), can account for the influence of color categories on perception, we predict that English speakers will show poorer discrimination of colours near prototypical blue and prototypical turquoise compared to colors at the blue-turquoise boundary.

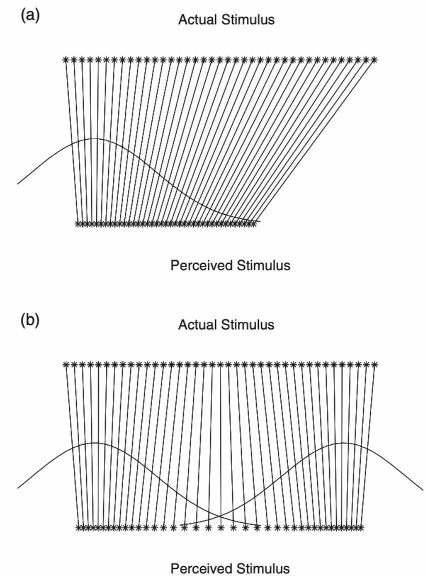

Figure 2: Predicted relationship between acoustic and perceptual space in the case of (a) one category and (b) two categories.

Experiment 1

Introduction

This experiment examines how the green component (G-value) in the RGB color space affects the perception and categorization of blue and turquoise colors. By systematically varying the G-value of a blue color, we aim to investigate the perceptual boundaries between blue and turquoise and the category goodness of each shade. Our findings will identify the prototypes for blue and turquoise, establish continuous and graded category memberships for different shades in the blue-turquoise spectrum, and generate essential parameters for Bayesian modeling in Experiment 2.

Methods

Participants A sample of 11 adult participants (7 females; age: mean = 26, SD = 4.8) was recruited to take part in the present study. All participants were fluent in English, but only 2 were native English speakers. All participants self-reported to have normal color vision.

Remote and In-person Testing The experiment was conducted either remotely (n = 1) or in-person (n = 10). In the in-person setting, participants viewed and responded to stimuli on the same computer that was used to present the experiment stimuli. In the remote experimental condition, the researchers conducted the experiment on their own computer while sharing their screen with the participant via Zoom, a remote conferencing platform. In both experimental settings, experimenters manually entered and recorded the participants' responses to the computer used to generate the experimental stimuli.

Apparatus All participants (including those tested remotely and those tested in-person) completed the experiment using a MacBook with Liquid Retina XDR Display, with the True Tone feature turned off and screen brightness set to maximum. All participant responses were recorded and stored using the same computer system that

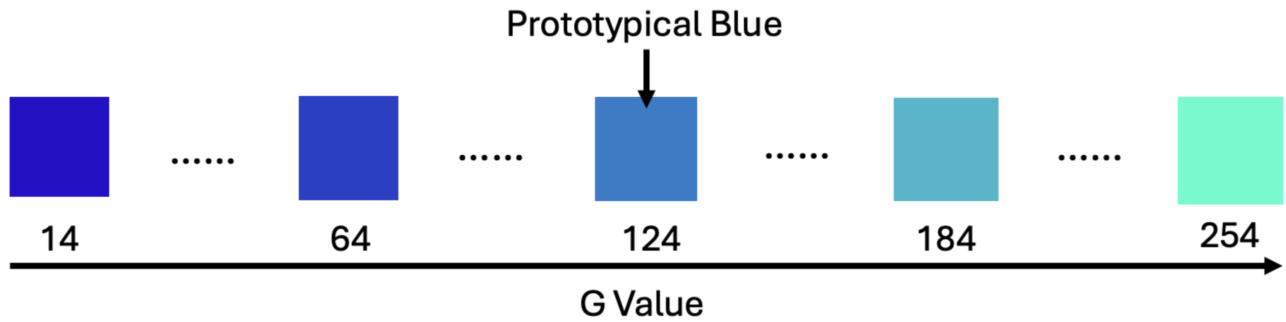

Figure 3: Examples of the experimental stimuli. The prototypical blue is defined in Regier et al., 2005. Different variations of blue are created by setting the R value as 38, B value as 203, and varying the G-value in the RGB triplets for the colors.

was used to present the experimental stimuli.

Stimuli Twenty-five color blocks were generated for the present experiment. These include one prototypical blue color and 24 variations of it. The prototype is identified by the World Color Project (Regier, Kay, & Cook, 2005), which described the prototypical blue with the following parameters in the Munsell Color Palette: Hue (or color) = 5PB, Value (or lightness/darkness) = 5, Chroma (weak/strong) = 12. This color can also be described by the RGB triplet [38, 124, 203] (*Munsell color palette*, n.d.). Twenty-four variations of blue were created by varying the "G," or the green component, in the RGB triplet for the blue prototype. The 25 G values ranged from 14 to 254 with increments of 10 (for examples of colors used, see Figure 3).

Procedure During the experiment, the participants were shown the 25 color stimuli, one at a time, in random order. For each color, they answered two questions: 1) "Is this color blue or turquoise?", and 2) "How good is this color an example of the category you chose?" For the first question, they were asked to respond by indicating "Blue" or "Turquoise". For the second question, they were asked to provide a score from 1 (worst representation of the color) to 7 (best representation of the color). Participants were given an unlimited viewing time for each color stimulus, allowing them to observe the color for as long as necessary before providing their response. Each participant viewed and responded to each color once.

Data Analyses For each color stimulus, the number of participants who responded Blue to the first question were recorded and then divided by the total number of participants, yielding the possibility of identifying each color as Blue. Then, for all trials where participants responded Blue, Goodness ratings were averaged across participants to produce a mean Goodness score for each color stimulus. Similarly, for all trials where participants responded Turquoise, Goodness ratings were averaged across participants to produce a mean Goodness score for each color.

Results

Identification of the Color Blue We observed large variations in participants' perception of blue as a function of the G-value in the RGB color space. As depicted in Figure 2A, the identification of blue remained relatively high and stable for G-values ranging from 14 to 150. Specifically, the probability of identifying a color as blue was consistently above 90% for G-values up to 150 but sharply declined to below 20% as G-values approached 200. This observation suggests a perceptual boundary or threshold around a G-value of 130, beyond which the color is less frequently recognized as blue, aligning closely with our defined prototype.

Goodness Ratings The assessment of the Goodness ratings for both blue and green hues, as a function of G values, is illustrated in Figure 2B. For blue, the goodness ratings begin near a maximum (approximately 6 on a scale from 1 to 7) for G-values less than 50, followed by a gradual decline in perceived quality as G-values increase. This downward trend continues until a G-value of 150, where the ratings sharply drop, reaching their lowest at approximately G-value 190.
Conversely, the goodness ratings for green show an inverse relationship. They start at their lowest when the G-values are below 100, gradually increasing as the G-values rise, with a significant increase noted past a G-value of 150. This rise continues, reaching a peak for the highest G-values tested, demonstrating a strong correlation between higher G-values and the perceived quality of green.

Discussion

The behavioral data indicates that not all shades in a category are equally good exemplars of the category. The results highlight a perceptual gradient in the recognition and qualitative assessment of colors, influenced markedly by the modulation of the green component in the RGB spectrum. As the G-value increases, there is a clear perceptual shift from blue to green, which is not only recognized by participants but also reflected in their quality assessments of the colors. This shift emphasizes the importance of the green component

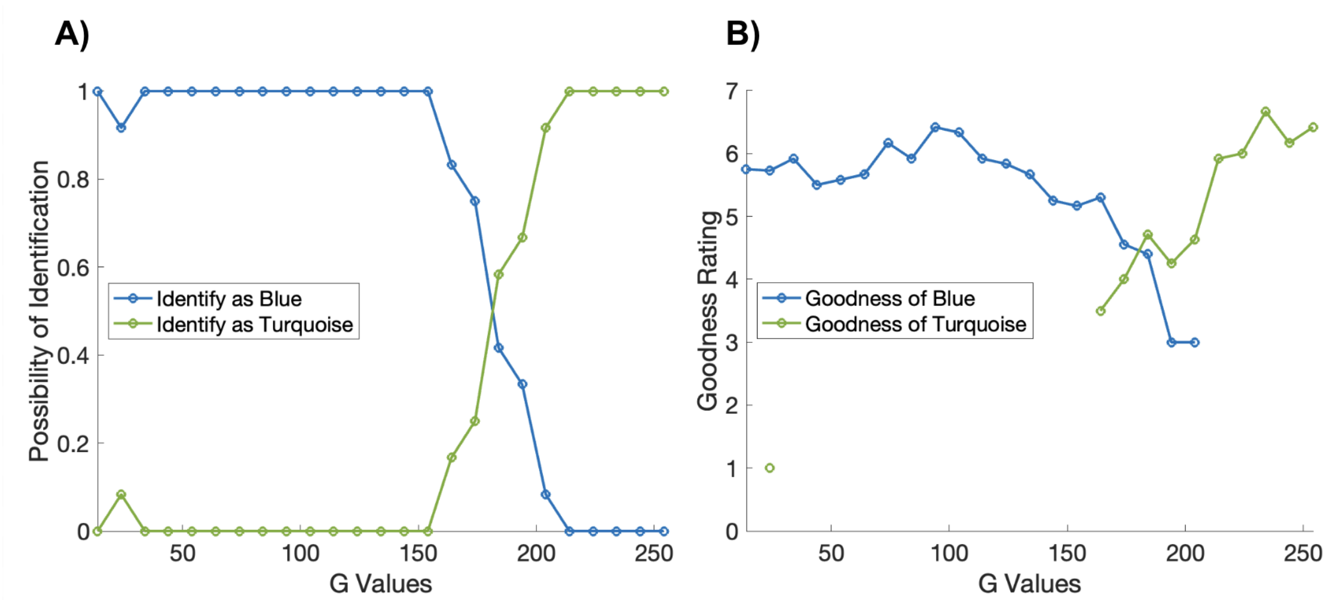

Figure 4: A. Effect of G values on Identifying the stimuli as blue. X axis shows G values used to create the stimuli, ranging from 14 to 254. Y axis shows the percentage of participants who identified the stimuli as blue. B. Effect of G values on average goodness ratings. The rating for how good is a stimulus a representative of blue, ranging from 1 to 7, is averaged across participants to produce the average goodness of rating. X axis shows the G values and y axis shows the average goodness ratings.

in color categorization and suggests a competitive interaction between the two color categories in terms of perceptual quality.

Experiment 2

Introduction

Building on the findings from Experiment 1, which established the boundaries and prototypes for blue and turquoise, Experiment 2 aims to replicate and model the perceptual magnet effect in color perception. By presenting participants with pairs of colors that differ in G-values and asking them to rate the perceived difference between the colors, we seek to demonstrate that discrimination is enhanced near category boundaries and reduced near category centers. A comparison of our behavioral data with predictions from a Bayesian model will provide a computational explanation for the observed effects.

Methods

The participants, experiment settings (in-person vs. remote), apparatus, and stimuli for Experiment 2 were the same with those for Experiment 1.

Similarity Discrimination Task In the course of the experiment, participants were presented with pairs of color stimuli. The G values of each color pair differed by 10, resulting in a total of 24 stimulus pairs. The G values of these pairs ranged from 14 and 24, to 24 and 34, and so on, up to 244 and 254. The stimuli pairs were presented in random order and participants were asked to rate, with a scale from 1 to 7, on how different the two colors were, with 1 being the most similar and 7 being the most different. Participants were given an unlimited viewing time for each color stimulus, allowing them to observe the color for as long as necessary before providing their response. Each participant viewed and responded to each color once.

Data Analyses For each color pair, participants' answers were averaged to produce a mean difference score for that pair of color stimuli. To explain the behavioral data, we used the Bayesian model developed by Feldman and Griffiths (2007) to simulate the similarity ratings. The expected value of the perceived color given the G-values tested, aggregating over all categories, could be modeled as Equation 1 below:

$$E[T|S] = \frac{\sigma_c^2}{\sigma_c^2+\sigma_s^2}S + \frac{\sigma_s^2}{\sigma_c^2+\sigma_s^2}\sum_c p(c|S)\mu_c \tag{1}$$

where T represents the perceived G-value; S represents the G-value of the stimuli; c represents a given category (in our case, blue or turquoise); μ_c and σ_c represent the mean and standard deviation of the category, respectively; and σ_S represents the noise of the stimuli. The term describing the posterior probability in the above equation was estimated from the following logistic equation that describes the identification of a given category:

$$p(c|S) = \frac{1}{1+e^{gS+b}} \tag{2}$$

where $g = \frac{\mu_1 - \mu_2}{\sigma_c^2 + \sigma_s^2}$ and $b = \frac{\mu_1^2 - \mu_2^2}{2(\sigma_c^2 + \sigma_s^2)}$. To minimize the number of free parameters, the model's parameters were derived, to the greatest extent possible, from empirical measurements. Among these parameters, μ_1 and μ_2 were the G-value for the prototypical blue and the G-value that yielded the largest Goodness rating for turquoise, respectively. The sum of the other two parameters, σ_c and σ_S, were chosen to maximize the fit of the model.

To make a direct comparison between the Bayesian model and the behavioral data, we first used Equation 1 to estimate the expected G-values for all G-values tested, and then found the estimated perceptual distance, i.e., the difference of expected values between neighboring G-values tested. Assuming linear relationship between perceptual distance and participant's similarity ratings, we derived the model estimates from the following equation:

$$E_{rating} = m * D + n \tag{3}$$

where D is the estimated perceptual distance and m and n are fitted constants.

Results

The results from Experiment 2 demonstrated a nuanced perception of color differences as G-values incrementally changed within the stimulus pairs. Figure 3 shows perceptual difference rating from participants as well as predictions from the Bayesian model.

Behavioral Data Participants' ratings indicated a distinct pattern of perceived differences between adjacent color pairs. Initially, for G-values from 14 to about 100, the perceived differences were moderate, averaging between ratings of 3 to 4. This suggests that participants could discern subtle variations in hue at lower G-values. As G-values increased beyond 100, perceived differences peaked, particularly between G-values of 150 to 200, where ratings approached an average of 6, indicating a heightened sensitivity to changes in this range.

Interestingly, the perceived differences declined once again for G values exceeding 200, stabilizing at a lower rating of around 2 to 3. Worth-noticing, the peak in G-values between 160-200 is the critical value between the blue category and green category. This reduction in perceived difference ratings at higher G-values suggests a stage in sensitivity to changes, potentially indicating a saturation point in hue differentiation or a perceptual grouping of higher G-values under a single color categorization.

Bayesian Modeling Parameters in the Bayesian model employed in Experiment 2 were derived from empirical values. The model's parameters were chosen to maximize the fit between the simulated predictions and the behavioral data collected from participants. Below we outline the parameters used in the simulation and their implications:

- μ_1 = 124: represents the prototypical G-value for the color blue.
- μ_2 = 255: represents the G-value at which the highest goodness rating for turquoise was observed.
- σ_c = 50: reflects the variability within the color categories.
- σ_s = 55: measures the noise level associated with the stimulus.
- M= 60.95 and N = 2.60: constants fitted to linearly transform the model's output to match the scale used in participant ratings.

The effectiveness of these parameters was validated through their ability to closely simulate the actual difference ratings given by participants, as illustrated in Figure 3. This alignment suggests that our model accurately reflects the cognitive processes underlying color discrimination tasks.

Discussion

The results of this experiment provide compelling evidence for the perceptual magnet effect. The observed pattern of perceived difference ratings, which peak near the middle G values and decrease towards the extremes of the blue and turquoise ranges, suggests that the perceptual space is distorted, with colors near the category center being perceived as more similar to each other. This compression of perceptual space is most pronounced in regions of unambiguous color categorization, typically near the prototypical colors or the centers of color categories. In these areas, the perceptual magnet effect is strongest, leading to a reduction in the perceived difference between colors that are close to the category prototype. Conversely, the effect is weakest at the category boundaries, where colors are more ambiguous and less clearly associated with a specific category. The diminished perceptual magnet effect at category borders allows for better discrimination between colors that fall on either side of the boundary, as these colors are perceived as more distinct from each other.

The simulation results align remarkably well with the observed behavioral data. However, it is important to note that the model parameters used in this simulation should be considered as initial estimates rather than definitive values for color perception. Given the variability in individuals' perceived color categories (Regier et al., 2005), it is probable that the actual parameter values may differ from those used in the simulation. Furthermore, it is conceivable that these parameters exhibit inter-individual differences.

General Discussion

The two experiments reported in this study provides evidence for the perceptual magnet effect in color perception, where uncertainty in color stimuli leads participants to infer a color that is closer to the mean of a color category than the specific shade they actually viewed. Experiment 1 shows that not all colors are perceived as equally good exemplars of their color categories. Experiment 2 demonstrates that people are better at discriminating colors that are far from the

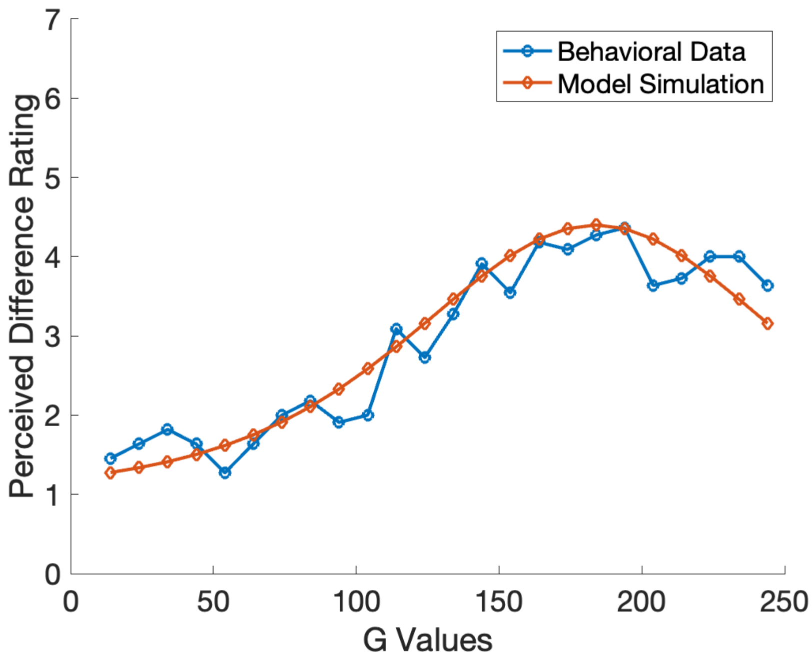

Figure 5: Perceived difference of colors with adjacent G-values. X values for the data points represent G-values used to create the color with the lower G value in the stimulus pair shown to the participant. The other color in the stimulus pair always had a G-value that was larger by 10. The blue line shows perceived difference ratings given by the participants, while the red line shows ratings predicted by the Bayesian model.

centers of perceptual categories, directly demonstrating the expanded perceptual space at category borders. We have also detailed a Bayesian model of color perception, which fits the empirical data closely and provides an explanation for better discrimination colors with subtle changes in green component at perceptual boundaries.

The interest in the influence of categorization on color perception is not new. In their study, Winawer et al. (2007) found that Russian speakers exhibited faster discrimination between colors that belonged to different linguistic categories in Russian, compared to colors within the same category. In contrast, English speakers did not show any category advantage when tested on the same stimuli. Here, we propose that these results could be explained by the perceptual magnet effect. The current study reveals that individuals are more adept at perceiving differences between colors that lie near the boundaries of two categories, as opposed to colors situated near the center of a single category. This observation can be attributed to the expansion of perceptual space at category borders, allowing for enhanced discrimination between colors on either side of the boundary, as they are perceived as more distinct from one another. Conversely, the compression of perceptual space at category centers makes it more challenging to discern differences between colors within this range. In another study, Huttenlocher et al. (2000) reported that the presence of category structure in visual stimuli allowed participants to compensate for memory trace uncertainty. We propose that color perception confronts a similar computational challenge. Similar to inferring a visual stimulus value while correcting for memory uncertainty, participants in our study must infer the specific shade of a color while compensating for uncertainty in the RGB values presented.

The results from our experiments demonstrate that the perceptual magnet effect, typically discussed in the context of speech perception, can also be applied to visual color perception. This effect can be understood as the consequence of optimally solving the statistical problem of color perception using knowledge about the structure of color categories. By employing Bayesian modeling, we accurately predicted the participants' ratings of color similarity and difference, highlighting the influence of prior knowledge and category means on perceptual inference. However, while the current parameter estimates serve as a solid foundation for understanding color perception, further research is necessary to refine these values and account for potential individual differences. Future studies could focus on collecting more extensive behavioral data across participants from different language backgrounds, to better characterize the variability in goodness ratings and to derive more precise parameter estimates. Additionally, exploring the impact of different parameter settings on the model's predictions could provide valuable insights into the robustness and generalizability of the perceptual magnet effect across various color categories.

In summary, this study underscores the significant overlap and interaction between color categories, revealing how subtle changes in a single color component can shift perceptual judgments largely. Future research could extend this model to other aspects of visual perception or apply similar frameworks to different sensory modalities, further exploring the universal applicability of Bayesian approaches to understanding human cognition.

References

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.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227–240.

Feldman, N. H., & Griffiths, T. L. (2007). A rational account of the perceptual magnet effect. In *Proceedings of the annual meeting of the cognitive science society.*

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. doi: https://doi.org/10.1037/a0017196

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

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220–241. doi: 10.1037//0096-3445.129.2.220

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, *97*, 553–562.

Kuhl, P. K. (1991). Human adults and human infants show a ”perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception Psychophysics*, *50*, 93–107.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

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*, *54*, 358–368.

Munsell color palette. (n.d.). Retrieved 2023-05-07, from `https://pteromys.melonisland.net/munsell/`

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, *102*(23), 8386–8391. doi: 10.1073/pnas.0503281102

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*, 369–398.

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*, *104*(19), 7780–7785. doi: 10.1073/pnas.0701644104