Categories affect color perception of only some simultaneously present objects

Marina Dubova (mdubova@iu.edu)
Cognitive Science Program, Indiana University
1101 E. 10th Street, Bloomington, IN 47405 USA

Robert Goldstone (rgoldsto@indiana.edu)
Cognitive Science Program & Department of Brain and Psychological Science, Indiana University
1101 E. 10th Street, Bloomington, IN 47405 USA

Abstract

There is broad empirical evidence suggesting that higher-level cognitive processes, such as language, categorization, and emotion, shape human visual perception. For example, categories that we acquire throughout lifetime have been found to alter our perceptual discriminations and distort perceptual processing. However, many of these studies have been criticized as unable to differentiate between immediate perceptual experience and the arguably concomitant processes, such as memory, judgment, and some kinds of attention. Here, we study categorical effects on perception by adapting the perceptual matching task to minimize the potential non-perceptual influences on the results. We found that the learned category-color associations bias human color matching judgments away from their category ideal on a color continuum. This effect, however, unequally biased two objects (probe and manipulator) that were simultaneously present on the screen, thus demonstrating a more nuanced picture of top-down influences on perception than has been assumed both by the theories of categorical perception and the El Greco methodological fallacy. We suggest that only the concurrent memory for visually present objects is subject to a contrast-from-caricature distortion due to category-association learning.

Keywords: categorical perception; top-down effects; El Greco fallacy; color perception; visual short-term memory

Introduction

A large body of evidence suggests that the categorical structure of our environment affects human auditory and visual perceptual processing. Learning categories, such as a /b/ phoneme in English, a dove, and a high school teacher, biases our perceptual processing as reflected in perceptual similarity judgments, discrimination judgments, same-different judgments, and stimulus matchings (e.g. de Leeuw, Andrews, Livingston, & Chin, 2016; Dubova & Goldstone, in press). In particular, category acquisition makes us more sensitive to the sensory stimulations that differentiate one category from another, and biases our perceptual matchings toward or away from the feature distributions of the item's category. For example, it has been suggested that knowing that bananas are typically yellow might make a new banana appear a bit more yellow than it actually is (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006). Categorical perception reflects human adaptations across a wide range of domains, and it has been suggested as partially underlying our perceptual expertise (Goldstone & Hendrickson, 2010).

Methodologies that have been commonly used to study higher-level (e.g. categorical) effects on perception have been recently criticized. Firestone and Scholl (2016) suggested that standard XAB discrimination, same-different judgment, similarity rating, and matching paradigms are unable to differentiate between memory, attention, and judgment-based adaptations from the adaptations that directly affect contents of immediate, conscious perceptual experience. This critique, however, is based on the assumption that the "pure" immediate perceptual experiences exist in isolation from the memory, judgment, and attention processes, and that the "pure" perceptual experiences could in principle be measured without being "contaminated" by these other processes. Even though we are skeptical that the "pure" perceptual experiences exist and can be studied, we eliminated the potential experimental artifacts in our study. Thus, we investigate top-down categorical effects on perception by minimizing potential influences from long-term memory, judgment, and some kinds of attention in a perceptual matching task. We found a systematic categorical effect even under these conditions, but this effect biased human active perception in a more nuanced way than it has been previously assumed.

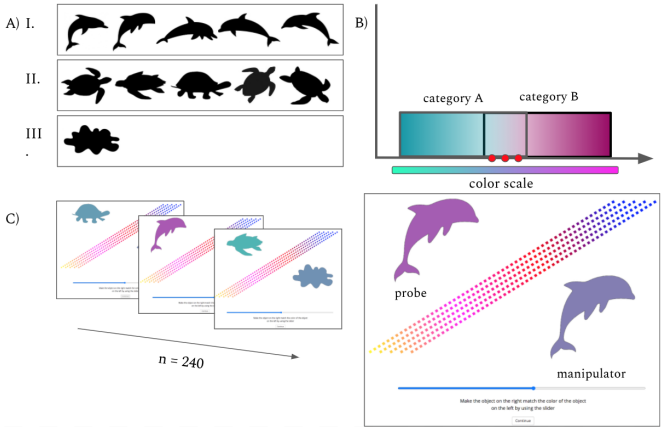

Figure 1: A) Stimuli contours used in the experiment: I. Category 1 ("dolphins"), II. Category 2 ("turtles"), III. "Neutral" category not associated with any color distribution. B) Color distributions of the probe items' hues for two categories. Red dots in the middle of the color scale represent the hues that were used for both categories. C) Experimental procedure.

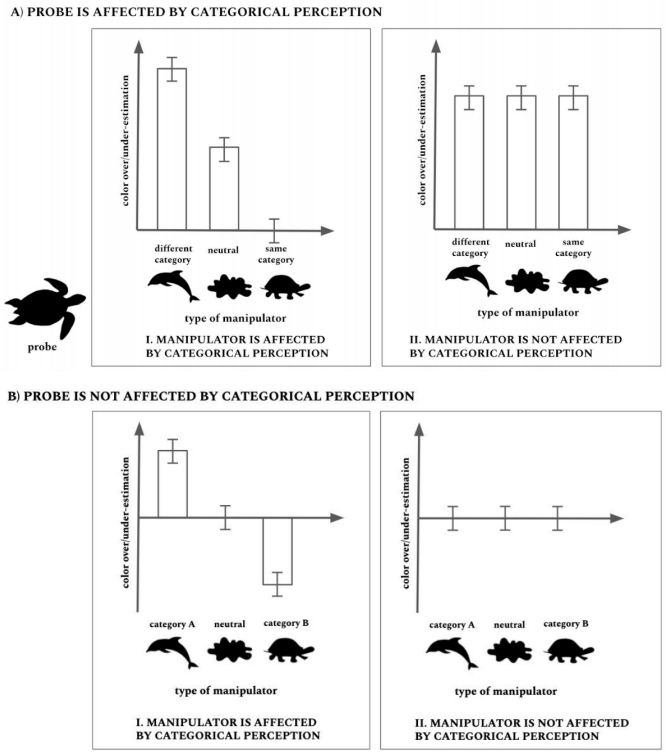

Figure 2: Qualitative predictions of the category-based matching bias patterns for each of the conceptual alternatives: A) I. Probe and manipulator perception are both subject to category-based biases; II. Probe perception is affected by a category-based bias, but the manipulator perception is undistorted; B) I. Manipulator perception acquires a category-based bias, but probe perception does not; II. Neither probe nor manipulator is affected by category learning. Note that the prediction patterns are identical for categorical effects of the contrastive and assimilative nature, as long as the effect direction stays the same for stimuli appearing as a manipulator and as a probe.

Minimizing the influences of non-perceptual processes in a perceptual matching experiment

Here, we use a computer-based perceptual matching task to study top-down categorical influences on visual perception. In this experimental paradigm, participants need to actively adjust a *manipulator* item to make it identical to a *probe* item on a given perceptual dimension (e.g. Goldstone, 1995). This experimental setting facilitates collecting rich data on the direction and magnitude of perceptual biases and also minimizes some potential non-perceptual influences on the results. In our experiment, participants actively adjusted a slider on the screen to change the color of the manipulator stimulus to match the color of the probe stimulus. On each trial, both the manipulator and the probe items were constantly present on the screen until participants made their final matching response (Fig.1C).

There are several reasons why this perceptual matching task would be expected to minimize potential influences from putatively non-perceptual processes on the results. First, both the probe and manipulator items for each trial simultaneously appeared on the screen while participants were making their adjustments. Thus, the only memory-related process that can affect the matching results is the kind of memory that operates at the scale of moving the eyes between two objects that are in front of the participants (we will further call this kind of memory *concurrent*). This kind of memory is rapid, low-level, and is present in all human perceptual interactions with the world. Second, the scale of the slider that participants use to make adjustments corresponds to the continuous physical scale of the matched dimension (color). This eliminates the potential judgment biases that result from translating the physical scale of the stimuli to the scale of responses. Moreover, we randomly varied the range of the judgment slider and its direction (normal or reverse) between trials, so that the participants could not associate the items' categories with particular positions (or ranges of positions) on the response slider. Third, the matched color dimension is suffused throughout the stimulus, so that covert spatial attention is not needed for completing the task. Thus, we suggest that the categorical effects that could be found using this experimental paradigm are unlikely to be attributable to spatial attention, judgment scaling, or short-term memory processes outside of the scope of human visual perception.

"El Greco" criticism of top-down effects on perceptual matching: invalidating or invalid?

The application of the perceptual matching paradigm in studies of top-down influences on perception has recently been criticized as suffering from "El Greco" methodological fallacy. Some studies have found categorical perception effects in matching experiments where the manipulator and probe belong to the same category (e.g. the color of one letter is adjusted to match the color of a letter with the same shape in Goldstone, 1995), which made Firestone and Scholl (2014) question the result: why have the "perceptual" effects been found at all in this condition? To illustrate the methodological fallacy, Firestone and Scholl alluded to the "El Greco" phenomenon in art history. Imagine that El Greco is drawing from life, but his astigmatism makes all figures look more stretched than they actually are. Will El Greco's painting depict the distorted or non-distorted figures? The authors suggest that the visual distortions that affect the immediate perceptual experiences should result in non-distorted figures on the painting: El Greco would perceive the figures on the canvas as more stretched too, thus the biases affecting the perception of the real figures and the biases affecting their depictions on the canvas should cancel each other out. Similarly, it has been suggested that categorical influences that affect immediate perceptual experience should equally bias both the probe and the manipulator items when they are drawn from the same category. Thus, categorical perception biases (for example, a "perceptual magnet effect" which distorts the exemplars of a

category towards their respective category prototypes) should lead participants to produce unbiased matchings, which contradicts the systematic categorical effects that have been reported in previous studies.

The "El Greco" criticism makes a strong case against the results that top-down (e.g. categorical) effects bias human immediate visual perception, as some studies report significant effects even when none should be expected. This criticism, however, is based on the assumption that the biases in immediate perception produce equal distortions for the scene and the canvas for El Greco, or the probe and the manipulator in the matching task. We suggest that this assumption will often not be psychologically plausible. Firestone (2013) notes this possibility, observing that even the originator of the astigmatic El Greco hypothesis argued that the visual distortion must vary with viewing distance, and therefore unequally influence the real life figures and their depictions on canvas.

Here, we suggest that there are many reasons why the top-down perceptual biases might unequally affect human perception of physically separated items. For example, we typically focus on a particular stimulus in the scene, while encoding the other items in concurrent memory between eye-movements. By "concurrent memory" we mean the memory for an object that is currently present. Most of the time, we do not even consider concurrent memory to involve memory at all. We simply think of it as our perceptual encoding of persistently available information. In the discussed examples, one of the compared items is usually present in concurrent memory while the other one is immediately perceived: for example, at a given moment El Greco is directly looking at the canvas while remembering the human figure that he just focused on and which is still perceptually available for him to study whenever he chooses. We suggest that the biases affecting human active scene perception might follow a nuanced picture. For example, some biases may selectively distort only the items that are currently being focused on, while other biases may only affect the concurrent memory encoding of them. Thus, it is possible for El Greco to draw distorted figures or for participants to produce distorted matches for the same-category items due to perceptual biases – if there is at least some difference between the effects biasing the concurrent encoding and the items that are immediately perceived.

Procedure

Here, we test whether categorical effects on perception uniformly affect the visible items. We independently varied categories of both manipulator and the probe items between trials. This design allows us to delineate between the acquired biases that affect only the probe, only the manipulator, both of them, and none of them (Fig.2).

The entire experiment consisted of 240 color matching trials (Fig.1C), where the probe items were drawn from the color distribution assigned to a given category. Thus, learning of the color distributions and bias assessment were happening synchronously at each trial.

Predictions

In the matching task used in this study, both the probe and the manipulator items are simultaneously present at two different quadrants of the screen (Fig.1C). Given the physical distance of the two items and the presence of the colorful dividing pattern, it is difficult to solve the matching task without repeatedly moving the eyes from manipulator to probe and back to manipulator. Moreover, the relation between the manipulator's hue and its specific location along the slider varies across the trials, so the participants have to actively look at the manipulator to reconstruct a desired hue. Thus, we hypothesize that the participants are most likely to directly look at the manipulator item when they are changing its color, and that they store the probe in concurrent memory.

The potential categorical perception effect is supposed to bias the middle-colored items from two categories in the opposite directions, given that the color distributions of two categories are located to the opposite sides from these middle test hues (Fig.1B). This allows us to differentiate between the categorical effects that bias the probe (concurrent memory), manipulator (immediate perception), both probe and manipulator (concurrent memory and immediate perception), and neither of them (Fig.2). If both probe and manipulator are affected by category-based color associations, adjusting a manipulator of a different, opposing category to match a color of the probe would produce the highest color matching distortion, as the participants would need to match the colors that are distorted in the opposite directions. By the same logic, the category-based effect should be the lowest (or non-different from zero) when the manipulator is from the same category, and the "neutral" modulators' effect will lay somewhere in between. These distortion patterns would indicate that categorization affects both concurrent memory encoding and the item that is currently focused on, as suggested by the "El Greco" fallacy (Fig.2A.I). If categorization only biases the concurrent memory encodings, we expect to see significant category-based matching distortion, but no effect of the manipulator type (Fig.2A.II). If categorization distorts the item that is directly perceived but does not affect its concurrent memory encoding, we expect to see different directions of perceptual distortion for different manipulator categories (Fig.2B.I). Finally, if categorization does not affect any of these processes, we expect to see no category-based distortion in perceptual matchings (Fig.2B.II).

Method

Participants

We collected data from 80 undergraduate students of Indiana University who received course credit for their participation. The sample size was decided a priori, based on the other studies with most similar design (e.g. Goldstone, 1995). The participants and the research assistant collecting the data were not aware of the experimental hypothesis. Three participants did not finish the experiment, and their data were not included in further analyses.

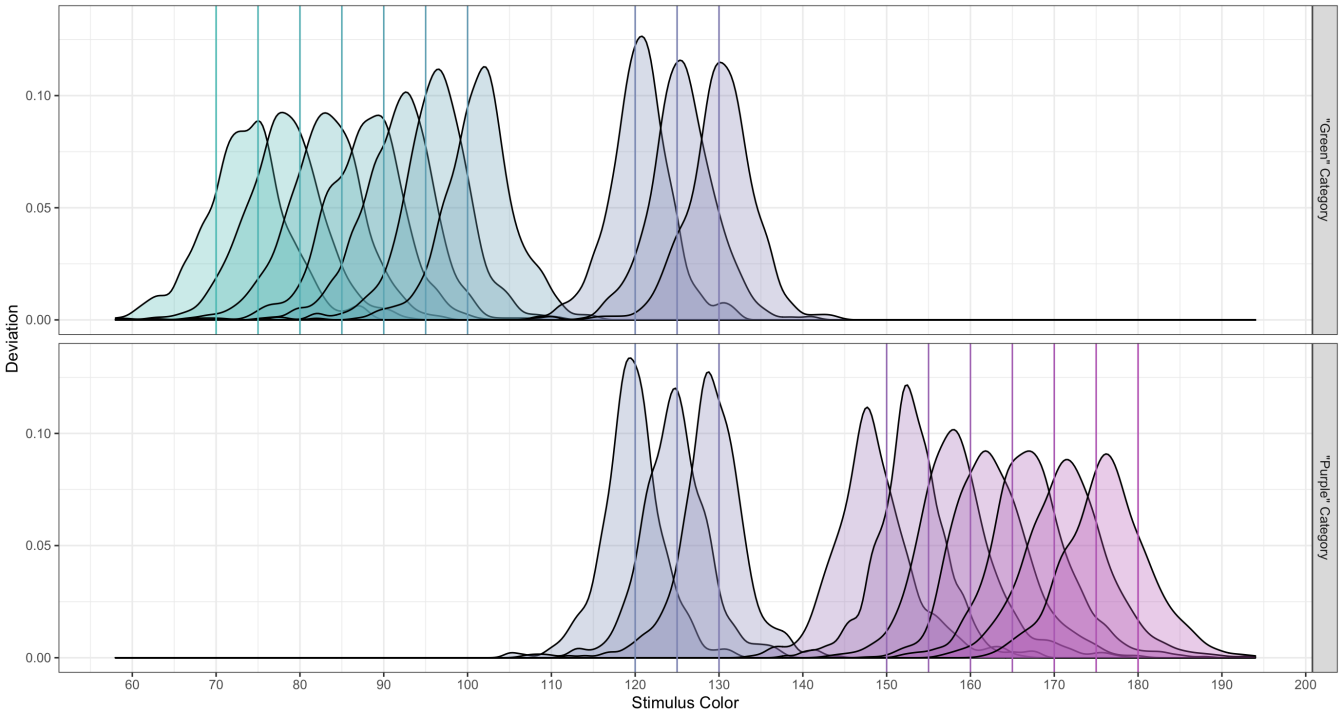

Figure 3: Matching distributions for "Green" (top) and "Purple" (bottom) category. Vertical lines represent the actual hues of the probe on the color dimension, and the densities reflect the matching distributions for these hues. The physical colors of the densities and vertical lines reflect the actual colors that were presented to participants for each of these hue conditions. Notice that the matching distributions are skewed, with the skewness increasing for more peripheral items on the color dimension. Also, notice that all matching distributions are biased away from the caricatures/extremes of the item's category color.

Stimuli

We used 5 dolphin-shaped contours and 5 turtle-shaped contours to represent examples of two categories. We used one amoeba-shaped contour for the "neutral" category trials (see Fig.1A). There were two color distributions for the two target animal categories (dolphins and turtles), defined by their position on the rgb color spectrum: rgb(x, (255-x), 180), where x = [70, 75, 80, 85, 90, 95, 100, 120, 125, 130] for one category and x = [120, 125, 130, 150, 155, 160, 165, 170, 175, 180] for the other category. Thus, across the experiment, one shape-based category was generally more green, while the other shape-based category was generally more purple (Fig.1B). The two categories' color distributions overlapped at three colors (x=[120, 125, 130]), which were used for the main hypothesis testing. Any systematic difference in the matching of these colors based on animal category can not be attributable to a general distortion in the perception of the hues.

The experiment consisted of 240 color matching trials and can be found here. We created the experiment using the jsPsych library (de Leeuw, 2015) with a custom written plugin to implement the interactive color matching paradigm. The experiment was executed in the Chrome browser in a laboratory context where each of 12 Apple iMac 21.5 inch computers was enclosed in a sound-dampening cubicle with a constant 5-watt, overhead illumination.

On each trial, participants adjusted the hue of the manipulator stimulus on the bottom-right corner to make it identical to the hue of the probe stimulus on the top-left corner of the screen. A colorful diagonal filler appeared between the two stimuli to prevent participants from using strategies that do not involve moving the eyes from one item to another. Participants were not limited in the time that they could use to match the hues. Participants pressed the "Continue" button on the bottom of the screen to finalize their matching response and progress to the next trial (Fig.1C).

The probe stimuli had two possible shape categories: dolphins and turtles. The manipulator items belonged to four shape-based categories depending on the shape of the probe on a given trial: identical, same category, opposite category, and neutral (the "amoeba" shape that never appeared as a probe). The categories of the probe and the manipulator were randomly chosen on each trial. The probe colors were randomly drawn from category-based color distribution. For each participant, one category (e.g. dolphin) generally laid on the "green" side of the color spectrum, whereas the other category (e.g. turtle) belonged to the "purple" side of the color spectrum (Fig.1B). The category-to-color assignment was randomly chosen for each participant when they started the experiment, so that the prior color associations that partic-

ipants have for each of two categories (dolphins and turtles) do not confound the results.

Participants used a slider on the bottom of the screen to make adjustments to the manipulator. The slider contained 190 hue increments, and its beginning color varied from 0 to 60. The starting position of the slider on each trial was exactly in the middle. The slider direction was randomly chosen between normal (from green to purple) and reverse (from purple to green).

Analysis

Data Exclusion Criteria

Data were excluded according to the following pre-registered exclusion criteria:

1. All of the data from a participant were excluded if the average absolute deviation of their reconstructions from the actual colors is larger than 13 (1 subject).

2. We also excluded data from individual trials in which the absolute deviation of the actual hue from the matched hue was larger than 18 (128 trials).

3. For pre-registered hypothesis testing, we only used the effects on the perception of the controlled hues that were present in both categories (e.g. the same 3 color values that appeared in both categories).

The analysis was replicated with a data-driven subject and trial exclusion criteria. In particular, we removed participants with an average absolute matching error more than 3 standard deviations higher than the average per-subject matching error, as well as trials with an absolute matching error higher than 3 standard deviations from the average matching error.

Hypothesis testing

We applied a pre-registered mixed linear model and an alternative, Bayesian mixed linear model to test the main effects of category shape and type of manipulator on color estimations. We used the following formula:

Signed deviation from the actual stimulus' color $\sim$ probe category + manipulator type + probe category * manipulator type + trial number + (1|subject)

We expected a significant effect of probe and manipulator type on the reconstruction deviations: in particular, we anticipated that reconstructions would be biased towards the mass of their respective category color distribution (as reported in Goldstone, 1995; Cibelli, Xu, Austerweil, Griffiths, & Regier, 2016; Bae, Olkkonen, Allred, & Flombaum, 2015).

Results

As expected, we found a significant effect of the probe category on the matching of the three overlapping colors across the two categories ("purple" category: est. = -0.68, 95% CI [-1.16,-0.18], t = -2.66, p = 0.008). Contrary to our expectations, the test hue matchings were biased away and not towards their respective category color distributions (Fig.3, Fig.4). The effect of the manipulator category (had 3 layers: "neutral", "green" category, "purple" category) was inconclusive: the effect for the "purple" category was significant (est. = 0.52, 95% CI [0.03,1.01], t = 2.09, p = 0.037), but the effect for the "green" category was not (est. = 0.3, 95% CI [-0.12,0.73], t = 1.41, p = 0.158). The effects of the interaction of the manipulator and probe category were not significant ("purple" category probe and "green" category manipulator: est. = -0.5, 95% CI [-1.15,0.15], t = -1.5, p = 0.134; "purple" category probe and "purple" category manipulator: est. = -0.51, 95% CI [-1.16,0.14], t = -1.53, p = 0.125). The effect of trial number was not significant (est. = -0.0001, 95% CI [-0.002,0.002], t = -0.14, p = 0.893). Adding a random slope of the probe category (probe category—subject) to the model did not lead to qualitative changes in the results.

We added the predictors of the response slider direction (normal or reverse) and slider starting position on each trial to the model. The significance of the reported effects and

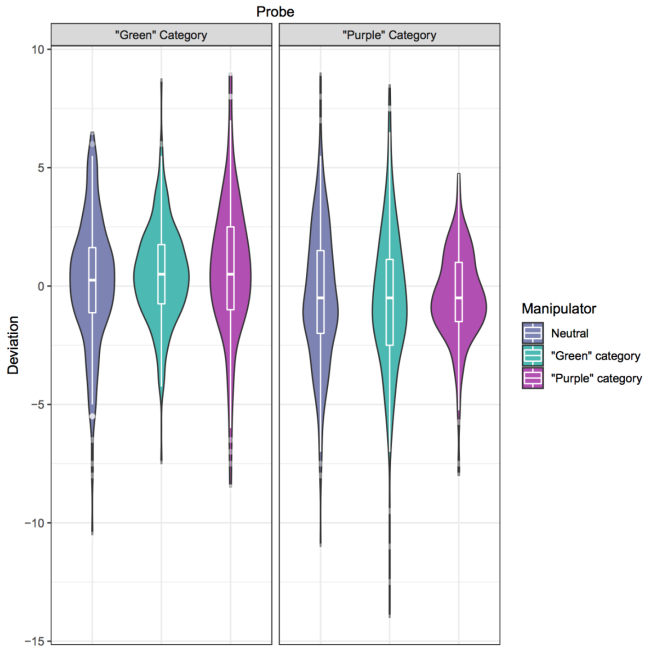

Figure 4: Effects of the shape-based category and manipulator type on matching errors for the three testing hues that were present in both categories. Negative matching error represents distortion to the left (green) side of the color dimension, positive matching error reflects distortion to the right (purple) side of the color dimension. Left: matching error distributions for the shape-based category (either dolphins or turtles) associated with the "green" side of the color continuum. Notice that the items of this category were more likely to be reconstructed as more purple than they were. Right: matching error distributions for the category that was associated with the "purple" side of the color continuum. Notice that the items of this category were more likely to be reconstructed as more green than they were.

their signs did not change, but the effect of slider direction was significant ("reverse": est. = 0.27, 95% CI [0.02,0.51], t = 2.15, p = 0.032).

The same analysis with the data-driven exclusion criteria led to the same results, except that the effect of the "purple" manipulator category was not significant (est. = 0.43, 95% CI [-0.07,0.93], t = 1.67, p = 0.094). Results of the Bayesian mixed linear regression analysis in Stan (Carpenter et al., 2017) were fully consistent with the results of frequentist mixed linear regression analysis. All the posterior distribution samples for the effect of "purple" probe category were negative (mean = -0.88, 95% CI [-1.28,-0.49]), and 97% of the posterior distribution samples for the effect of "purple" manipulator category were positive (mean = 0.38, 95% CI [-0.02,0.78]).

The observed categorical bias pattern (Fig.4) is consistent with one of our predictions (Fig.2A.II: probe is distorted by categorical bias, but the manipulator is not).

Discussion

We found that participants' perceptual matchings of colors were biased by the category of the probe, but not the manipulator, stimulus. Both items were simultaneously present on the screen, and the possibility for slider position-category associations was eliminated by assigning a random starting hue for each trial's slider. There are several differences between manipulator and probe items that could cause their differential susceptibility to a top-down categorical bias. In particular, the probe color remains constant while the manipulator color is actively changed by the participants. The probe item is likely stored in a concurrent memory while the manipulator is focused as its color is adjusted. We suggest that the category-based bias most likely affects the concurrent memory encoding that operates at the timescale of moving the eyes from one part of the screen to another. Even though we did not directly measure eye movements, we suspect that the concurrent memory that affects the probe could be similar to transsaccadic memory which has been proposed to store the contents of objects on the scene during active perception (note that there are many similarities between transsaccadic and visual short-term memory: Hollingworth, Richard, & Luck, 2008). Again, we note that the kind of memory that participants employ to solve our matching task likely participates in most realistic perceptual activities, and that there is still extensive (and probably unresolvable) debate on whether such visual short-term memory constitutes our conscious visual experiences (for a discussion, see O'Regan & Noë, 2001).

Our main result shows that learned categorizations affect human perception in a more nuanced way than has previously been recognized. In particular, not all of the items in a scene are equally affected by learned categories. We suggest that further studies clarify the properties that make some items in a scene more susceptible to categorical biases than others.

Our work urges caution for using the "El Greco" criticism when interpreting top-down effects on perception. Our results cast doubt on this criticism's main premise that the probe and manipulator (or the canvas and the scene, for El Greco) are equally affected by top-down effects on perception. Valenti and Firestone (2019) recently reported similar results: they found inconsistent categorical influences on color matchings for different combinations of the object and its background categories. The authors interpreted this result as indicating the *non-perceptual* nature of the effect. Our interpretation differs from Valenti and Firestone's conclusion: we suggest that it might not be possible to differentiate between "perceptual" and "non-perceptual" effects beyond what can be done by minimizing clearly non-perceptual artifacts in a given task. The "El Greco logic", however, can be used for testing whether a given bias equally affects different types of objects in the task, as we did in this study.

We found that participants' concurrent memory color encodings were contrasted away from the probe category's idealization on the color scale, instead of being assimilated to the category-based color distribution (Fig.3 & 4). Imagine that dolphins are generally greener than turtles. An idealized, or caricatured, color representation of dolphins would be shifted away from the prototypical color of the dolphins in the direction opposite of the contrasting turtle category's purple hue. Accordingly, the caricature of dolphins would be greener than the prototypical dolphin, perhaps as extreme as color=60 (Fig.3), whereas the turtle's caricature would be around the purple extreme (e.g. at color=190). Even though we expected that participants' matching judgments would be distorted by an assimilative category bias, the contrast-from-caricature effect that we found is consistent with previous results on interrelated category representations. In particular, when simultaneously learning categories that are frequently alternated, people often form caricature-based category representations that emphasize the diagnostic criteria that distinguish one category from the others (Goldstone, 1996; Palmeri & Nosofsky, 2001; Goldstone, Steyvers, & Rogosky, 2003; Ameel & Storms, 2006; Davis & Love, 2010; Levering & Kurtz, 2006). Our results indicate that the stimuli get encoded relative to such idealized representations.

The contrast-from-caricature effect on perceptual concurrent memory encodings is inconsistent with many rational models of perceptual memory. In particular, a rational perceptual encoding is commonly formalized as assimilated towards the relevant priors (e.g. the respective category color distribution). Our data suggest that the concurrent memory encodings of the testing stimuli are contrasted away, instead of being biased towards, the perceptual prior distribution. Two recent studies used a similar immediate color matching paradigm to study the effects of language labels on color reconstruction (Cibelli et al., 2016; Bae et al., 2015). These studies showed a significant effect of prototypes associated with English color labels on perceptual matching, and this evidence has been construed as supporting the rational models of adjusting hue perception towards the label-associated category prototype (see also Huttenlocher, Hedges, & Dun-

can, 1991; Feldman, Griffiths, & Morgan, 2009). However, close examination of the results of both studies indicates a bias away from caricatures for one out of two tested colors in Cibelli et al. (2016) and the biases away from the prototype for several color categories in Bae et al. (2015). Taken with our current evidence, these studies suggest that contrast from category caricatures is a commonly observed, if neglected, phenomenon, and further studies should clarify the role of category distributions in biasing concurrent visual memory encodings in different directions.

Open practices statement

Pre-registration can be accessed at https://osf.io/maf93. De-identified data and scripts for reproducing analyses and figures are posted at https://osf.io/uprch/.

Acknowledgement

Authors thank Reina Munoz for the help with data collection; Arseny Moskvichev, Gunar Epping, Jerome Busemeyer, and Indiana University Statistical Consulting Center for their suggestions on modeling and data analysis; Chaz Firestone, Eleanor Schille-Hudson, and 4 anonymous reviewers for their helpful suggestions on the earlier version of this manuscript; all the attendees of the IU PCL and HSE Cognitive Research lab meetings, especially Madeleine Ransom, Peter Todd, Vladislav Khvostov, and Igor Utochkin for discussions that were helpful for interpreting the results.

References

Ameel, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language*, *55*(3), 402–421.

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: a probabilistic programming language. *Grantee Submission*, *76*(1), 1–32.

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The sapir-whorf hypothesis and probabilistic inference: Evidence from the domain of color. *PloS one*, *11*(7), e0158725.

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science*, *21*(2), 234–242.

de Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*(1), 1–12.

de Leeuw, J. R., Andrews, J. K., Livingston, K. R., & Chin, B. M. (2016). The effects of categorization on perceptual judgment are robust across different assessment tasks. *Collabra*, *2*(1).

Dubova, M., & Goldstone, R. L. (in press). The influences of category learning on perceptual reconstructions. *Cognitive Science*.

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.

Firestone, C. (2013). On the origin and status of the "el greco fallacy". *Perception*, *42*(6), 672–674.

Firestone, C., & Scholl, B. J. (2014). "top-down" effects where none should be found: The el greco fallacy in perception research. *Psychological science*, *25*(1), 38–46.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and brain sciences*, *39*.

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, *6*(5), 298–304.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & cognition*, *24*(5), 608–628.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78.

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, *31*(2), 169–180.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature neuroscience*, *9*(11), 1367–1368.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*(1), 163.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological review*, *98*(3), 352.

Levering, K., & Kurtz, K. J. (2006). The influence of learning to distinguish categories on graded structure. In *Proceedings of the 28th annual conference of the cognitive science society* (pp. 26–29).

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, *24*(5), 939.

Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 197–235.

Valenti, J., & Firestone, C. (2019). Finding the "odd one out": Memory color effects and the logic of appearance. *Cognition*, *191*, 103934.

Characterizing Variability in Shared Meaning through Millions of Sketches

Molly Lewis
mollyllewis@gmail.com
Dept. of Psychology
Carnegie Mellon University

Anjali Balamurugan
abalamur@andrew.cmu.edu
Dept. of Stats. & Data Science
Carnegie Mellon University

Bin Zheng
binzheng@andrew.cmu.edu
Dept. of Stats. & Data Science
Carnegie Mellon University

Gary Lupyan
lupyan@wisc.edu
Dept. of Psychology
University of Wisconsin-Madison

Abstract

The study of mental representations of concepts has historically focused on the representations of the "average" person. Here, we shift away from this aggregate view and examine the principles of variability across people in conceptual representations. Using a database of millions of sketches by people worldwide, we ask what predicts whether people converge or diverge in their representations of a specific concept, and which kinds of concepts tend to be more or less variable. We find that larger and more dense populations tend to have less variable representations, and concepts high in valence and arousal tend to be less variable across people. Further, two countries tend to have people with more similar conceptual representations when they are linguistically, geographically, and culturally similar. Our work provides the first characterization of the principles of variability in shared meaning across a large, diverse sample of participants.

Keywords: concepts and categories; drawings; cultural variability; large scale data

Introduction

Understanding how the human mind represents concepts is a fundamental question for psychologists, philosophers, and linguists (Margolis & Laurence, 1999). Researchers have developed a wide range of theories characterizing how people represent relatively simple concepts that are shared across people, like "chair" and "tree." Such theories make predictions about how an "average" person should perform in behavioral tasks, such as rating how typical members of the category are. Here we shift the focus from the representation of the "average" person, to *differences* in representations between people. This allows us to ask (1) what predicts whether people converge or diverge in their representations of a specific concept, and (2) which concepts are more vs. less variable. To answer these questions, we use a novel method: analyses of millions of sketches drawn by participants worldwide.

The question of variability in shared meaning across people has received remarkably little attention in part because it is difficult to study. One reason for this is that people tend to engage in "good enough" processing (Ferreira & Patson, 2007) when faced with behavioral tasks, thus often failing to reveal to outside observers underlying differences in their conceptual representations. Second, the psychological paradigms that are used to study individuals' concept representations, like typicality (e.g., Rosch, 1975) and word association tasks (e.g., De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019), are relatively course-grained. And, third, the differences in conceptual representations between people are likely small, requiring a large dataset to observe variability.

One way to get a window into people's shared meaning is through drawings. Like words, drawings are emergent cultural conventions that can be used to communicate about shared meaning. However, unlike words, drawings resemble the shared meaning they reference—a drawing of the concept "chair" *looks like* a chair. Drawings therefore provide an observable, quantifiable index of a person's conceptual representations (e.g., Fan, Yamins, & Turk-Browne, 2018; Long, Fan, & Frank, 2018).

Notably, drawings are not an unmediated window into a person's conceptual representations—a person's drawing of "chair" is not isomorphic to their cognitive representation of a chair (Cohn, 2019). This difference is due in part to cultural conventions about how to represent particular concepts in the drawing modality. For instance, a culture may converge on the convention of drawing the concept "house" with four windows and a chimney. The fact that drawings are conceptual representations mediated by drawing-specific conventions is unproblematic for the current purposes: drawings provide a test bed for understanding the dynamics of shared meaning in one particular modality.

Prior work has examined the emergence of drawing conventions in experimental paradigms (Garrod, Fay, Lee, Oberlander, & MacLeod, 2007). A key finding from this work is that drawings become both more consistent through repeated interactions and also more reliant on memory. For example, when two interlocutors are tasked with communicating the meaning "bunny" through sketch, they might draw a detailed picture with a nose, whiskers, and ears. But, with repeated interactions, this drawing will tend to become more schematic such that "bunny" is represented simply as two ears.

This prior work makes two general predictions about variability in shared meaning. The first is that the degree of interaction among people should be related to the degree of consensus in a concept: More social interaction should lead to higher degree of consensus (less variability between people). This prediction is consonant with findings in the language evolution literature demonstrating that languages with fewer people tend to have more complex (arguably, less variable) language systems (Lupyan & Dale, 2010). Second, psychological variables that influence memory should influence which concepts are more variable across people.