Cognition 223 (2022) 105025

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Categorical perception meets El Greco: Categories unequally influence color perception of simultaneously present objects

Marina Dubova [a,*,1], Robert L. Goldstone [a,b,1]

[a] Cognitive Science Program, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405, USA
[b] Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405, USA

ARTICLE INFO

Keywords:
Categorical perception
Top-down effects
El Greco fallacy
Color perception
Perceptual learning

ABSTRACT

Broad empirical evidence suggests that higher-level cognitive processes, such as language, categorization, and emotion, shape human visual perception. Do these higher-level processes shape human perception of all the relevant items within an immediately available scene, or do they affect only some of them? Here, we study categorical effects on visual perception by adapting a perceptual matching task so as to minimize potential non-perceptual influences. In three experiments with human adults ($N = 80$; N = 80, $N = 82$), we found that the learned higher-level categories systematically bias human perceptual matchings away from a caricature of their typical color. This effect, however, unequally biased different objects that were simultaneously present within the scene, thus demonstrating a more nuanced picture of top-down influences on perception than has been commonly assumed. In particular, perception of only the object to be matched, not the matching object, was influenced by animal category and it was gazed at less often by participants. These results suggest that category-based associations change perceptual encodings of the items at the periphery of our visual field or the items stored in concurrent memory when a person moves their eyes from one object to another. The main finding of this study calls for a revision of theories of top-down effects on perception and falsify the core assumption behind the El Greco fallacy criticism of them.

1. Introduction

Human perceptual processing adapts to the higher-level goals of an individual, such as categorization, communication, and problem solving. For example, trained fingerprint examiners tend to look at the most diagnostic regions of the fingerprints (Busey & Parada, 2010), professional musicians show greater sensitivity to physically equated acoustic differences between two intervals if they fall into two different categories (e.g. major third and fourth) than if they fall into one interval category (e.g. major third) (Burns & Ward, 1978), and each one of us is tuned to discriminate the phonemes of our native languages (e.g. Kuhl, 1991).

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, or the face of a high school teacher, biases our perception of their exemplars as reflected in perceptual similarity judgments, discriminations, same-different judgments, and matching responses (e.g. de Leeuw, Andrews, Livingston, & Chin, 2016; Dubova & Goldstone, 2021). In particular, category acquisition makes us more sensitive to the sensory stimulations that differentiate one category from another, and biases our perceptual matchings towards or away from the perceptual feature distributions of the item's category. For example, perception of fruit colors tends to be biased towards their typical color (e.g. bananas tend to be perceived as more yellow than they are) as indicated by the finding that participants believe that the fruits are achromatic only when they are filled with a grayish color that is slightly shifted away from the fruit's typical color (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

* Corresponding author.
E-mail addresses: mdubova@iu.edu (M. Dubova), rgoldsto@indiana.edu (R.L. Goldstone).

[1] The stimuli for all experiments, the data, and the scripts used to create the experiments, generate the plots and run all the analyses are publicly available.

https://doi.org/10.1016/j.cognition.2022.105025
Received 10 April 2021; Received in revised form 12 January 2022; Accepted 13 January 2022
Available online 22 January 2022

judgment-based adaptations from the adaptations that directly affect contents of immediate, conscious perceptual experience. This critique, however, relies on the assumption that "pure" immediate perceptual experience exists in isolation from these memory, judgment, and attention processes, and that a "pure" perceptual experience could in principle be measured without being "contaminated" by these other processes. Even though we are skeptical that "pure" perceptual experiences exist and can be studied, it is nonetheless a valuable empirical goal to minimize artifacts due to memory and judgment when studying human perception. Here, we investigate top-down categorical effects on perception by minimizing potential influences from long- and short-term memory, judgment, and specific kinds of attention in a perceptual matching task.

Both proponents and critics of categorical perception commonly assume that the categorical effects on perception, if they exist, should equally affect all of the relevant items in the visual field. For example, if bananas tend to be perceived as more yellow than they actually are, then all of the bananas present in a visual scene are expected to be biased in this way. In the current study, we empirically test this assumption. We find that categorical biases unequally alter the perception of identical items that are simultaneously present on a screen. This result invites the refinement of categorical perception theories and undermines a core assumption behind the El Greco fallacy criticism of categorical perception (see section 1.2).

1.1. Minimizing the influences of non-perceptual processes in perceptual matching paradigm

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) (Fig. 3C).

There are several reasons why our perceptual matching task would be expected to minimize influences from putatively non-perceptual processes on the results. First, both the probe and manipulator items simultaneously appear on the screen while participants are making color 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 presently in front of the participants. We will call this kind of memory ***concurrent***. Concurrent memory is far more rapid than short-term memory and it participates in all human perceptual interactions with the world. 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. Second, the scale of the slider that participants use to make adjustments is identical to the continuous physical scale of the matched dimension (color). This eliminates potential judgment biases that result from translating the physical scale of the stimuli to the scale of responses. Moreover, we randomized the range of the response slider and its direction, so that the participants could not associate the items' categories with particular response positions. 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.

1.2. "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 been criticized by Firestone and Scholl (2014). Namely, the authors questioned categorical perception effects reported 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). To illustrate the alleged methodological fallacy, Firestone and Scholl allude to the "El Greco" phenomenon in art history (Firestone, 2013). 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 distorted or non-distorted figures? The authors suggest that the visual distortions that affect the perceptual experiences should result in non-distorted figures in the painting (Fig. 1A). El Greco would perceive the figures on the canvas as more stretched, too, thus the biases affecting the perception of the real figures and their depictions on the canvas should cancel each other out. Similarly, Firestone and Scholl suggested that categorical influences (for example, a "perceptual magnet effect" which distorts items towards their category prototypes) should equally alter perception of both the probe and the manipulator items when they are drawn from the same category. Thus, categorical perception biases should lead participants to produce unbiased matchings, which contradicts the systematic categorical effects that have been reported.

The "El Greco" criticism relies on a common assumption that the perceptual biases should equally distort all of the relevant items in the observer's environment: e.g. the scene and the canvas for El Greco, or the probe and the manipulator for participants completing the matching task. This assumption will often not be psychologically plausible: humans rarely perceive two physically separated items in the same way at the same time. Rather, we selectively focus on a particular stimulus in the scene, while encoding the other items in concurrent memory between eye-movements. In the discussed examples, one of the compared items is usually present in concurrent memory or peripheral vision while the other one is more directly gazed upon. For example, at a given moment El Greco is directly looking at the canvas while remembering the human figure which is perceptually available for him to study whenever he chooses. We suggest that the biases affecting human active scene perception might follow a nuanced picture. Some biases may selectively distort only the items that are currently being focused on, while other biases may only affect the concurrent memory or peripheral vision 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, as long as there is a difference between the effects altering the foveated and all the other objects at the scene (Fig. 1B). Firestone (2013) notes this possibility, observing that even the originator of the astigmatic El Greco hypothesis argued for its more nuanced versions, e.g. that the visual distortion must vary with viewing distance, and therefore unequally influence the real life figures and their depictions on canvas.

Here, we test whether categorical effects on perception uniformly affect the items in the visual scene. We independently vary 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 neither of them (Fig. 2).

2. Experiment 1

2.1. Method

2.1.1. Participants

We collected data from 80 undergraduate students of Indiana University who received course credit for their participation and were recruited with the Indiana University SONA system. The sample size was decided a priori, based on the other studies with most similar design (e.g. Goldstone, 1995), and pre-registered (https://osf.io/maf93). The participants and the research assistant collecting the data were not aware of the experimental hypothesis. Three participants did not finish the experiment, so their data were not included in further analyses. All participants received an informed consent form and provided their assent. The study was approved by the Indiana University Institutional

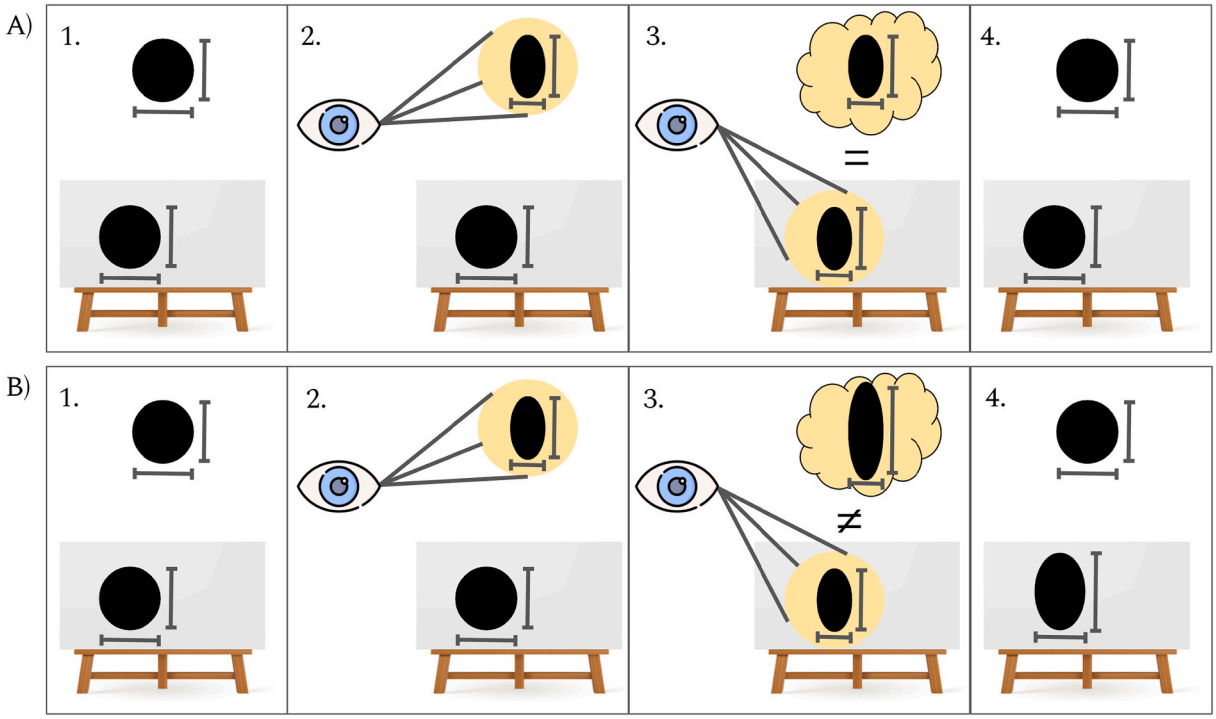

Fig. 1. A) "El Greco effect". Perceptions of the figures outside and inside the canvas are equally distorted by immediate perception (2, 3). The perceptual effects cancel each other out, and the drawings resulting from such distortions are not biased (1, 4). B) a "Real-life El Greco effect" which is not based on the assumption of equivalent immediate perception of the canvas and the scene. The potentially different biases that affect immediate perception and concurrent memory combine (2, 3) and result in a drawing that is distorted in a way that reflects the difference between concurrent memory bias and the immediate perception bias (4). Yellow circles represent the focus of immediate perception, yellow clouds represent the concurrent memory encoding. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
The figure uses materials from flaticon.com

Review Board.

2.1.2. Stimuli

We used 5 dolphin-shaped contours and 5 turtle-shaped contours to represent examples of two animal categories. These shape-based categories are associated with particular color ranges within the experiment. We used one amoeba-shaped contour for the "neutral" category trials: this shape will have no color association induced in the experiment (see Fig. 3A). 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 (see Appendix C for the translation of the stimuli colors to the HSV scale). Thus, across the experiment, one randomly selected animal category was generally more green, while the other category was generally more purple. 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 cannot be attributable to a general distortion in the perception of specific color values (Fig. 3B).

2.1.3. Procedure

The experiment consisted of 240 color matching trials and could be found at https://pcl.sitehost.iu.edu/robsexperiments/tests&examples/colorReproduction/colorReproduction.html. 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 in. computers was enclosed in a sound-dampening cubicle with a constant 5-watt, overhead illumination.

On each trial, participants adjusted the color of the manipulator stimulus on the bottom right corner of the screen to make it identical to the color of the probe stimulus on the top left corner. A colorful diagonal filler appeared between the two stimuli to prevent participants from using the strategy of imagining the colors next to each other and looking for a color discontinuity. Participants were not limited in the time that they could spend matching the colors. Participants pressed the "Continue" button on the bottom of the screen to finalize their matching response and progress to the next trial (Fig. 3C).

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* (an "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 those chosen category's color distribution (Fig. 3B). 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. 3B). The category-to-color assignment was randomly chosen for each participant when they started the experiment.

Participants used a slider on the bottom of the screen to make adjustments to the manipulator. The slider contained 190 color 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). Thus, the color of the manipulator on each frame was determined by the following formula for the normal slider: rgb(start+x, 255 + start-x, 180); and the following formula for the reverse slider: rgb(255 + start-x, start+x, 180), where x is the current position of the slider and *start* is the starting position of the slider on

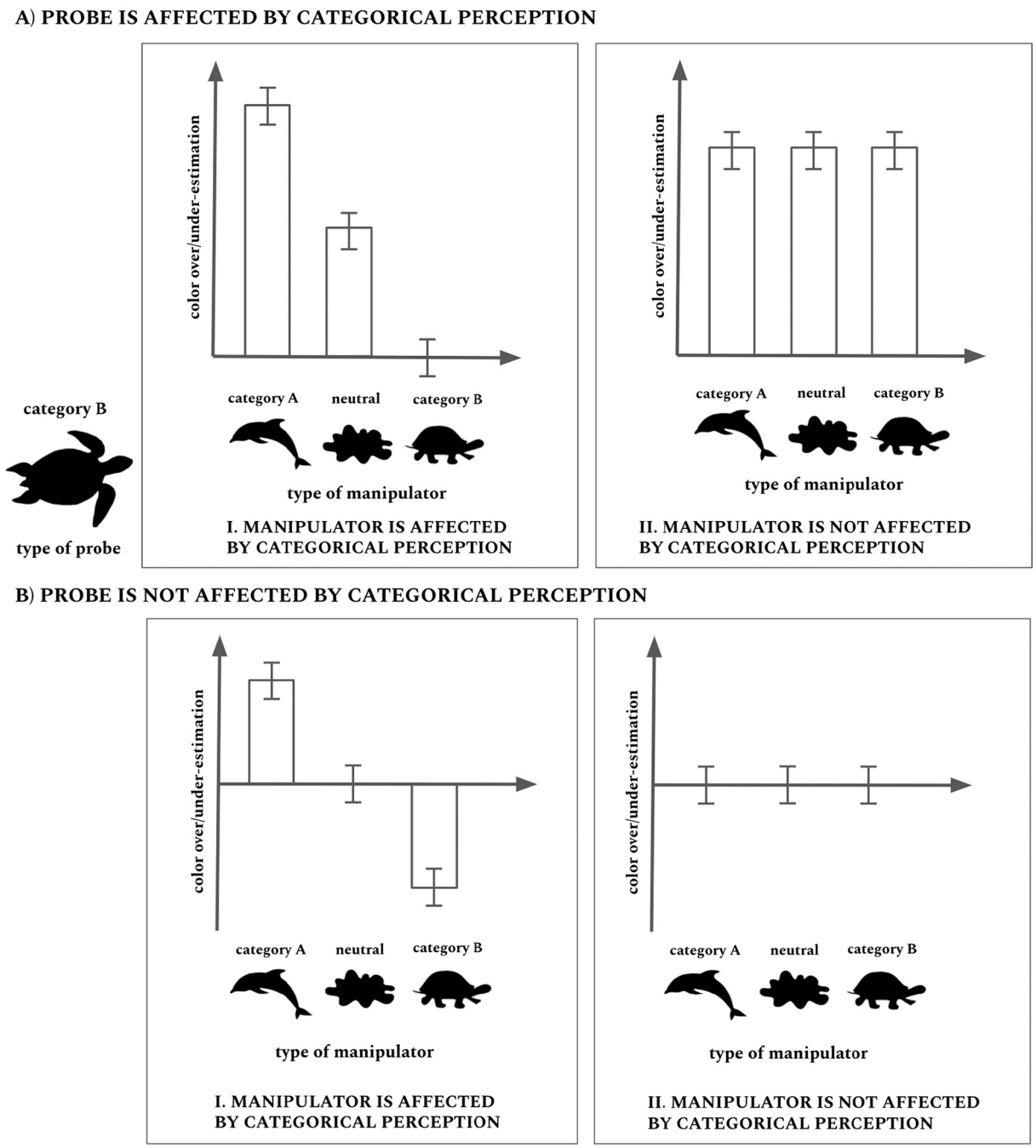

Fig. 2. Qualitative predictions of the category-based matching bias patterns for each of the conceptual alternatives: A) I. Probe and manipulator perception are both subject to category-based biases; II. Probe perception is affected by a category-based bias, but the manipulator perception is undistorted; B) I. Manipulator perception acquires a category-based bias, but probe perception does not; II. Neither probe nor manipulator is affected by the associated category. Note that the predictions are not committed to the direction (assimilative or contrastive) of the categorical effect, as long as both category A and category B distort the items in the same way.

a given trial.

2.1.4. 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. Given the physical distance of the two items and the presence of the colorful dividing filler, it is difficult to solve the matching task without keeping at least one of the items in peripheral vision or repeatedly moving the eyes between manipulator and probe. Moreover, the relation between the manipulator's color 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 color. Thus, we assume 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 or hold it in peripheral vision (see Experiment 3 for the empirical evidence supporting this assumption).

The potential categorical perception effect (either assimilative or contrastive) is supposed to bias the middle-color items from the two categories in opposite directions. This allows us to differentiate between the categorical effects that bias the probe (concurrent memory or peripheral vision), manipulator (foveated perception), both probe and manipulator (concurrent memory/peripheral vision and foveated perception), and neither of them. If both probe and manipulator are affected by category-based color associations, adjusting a manipulator

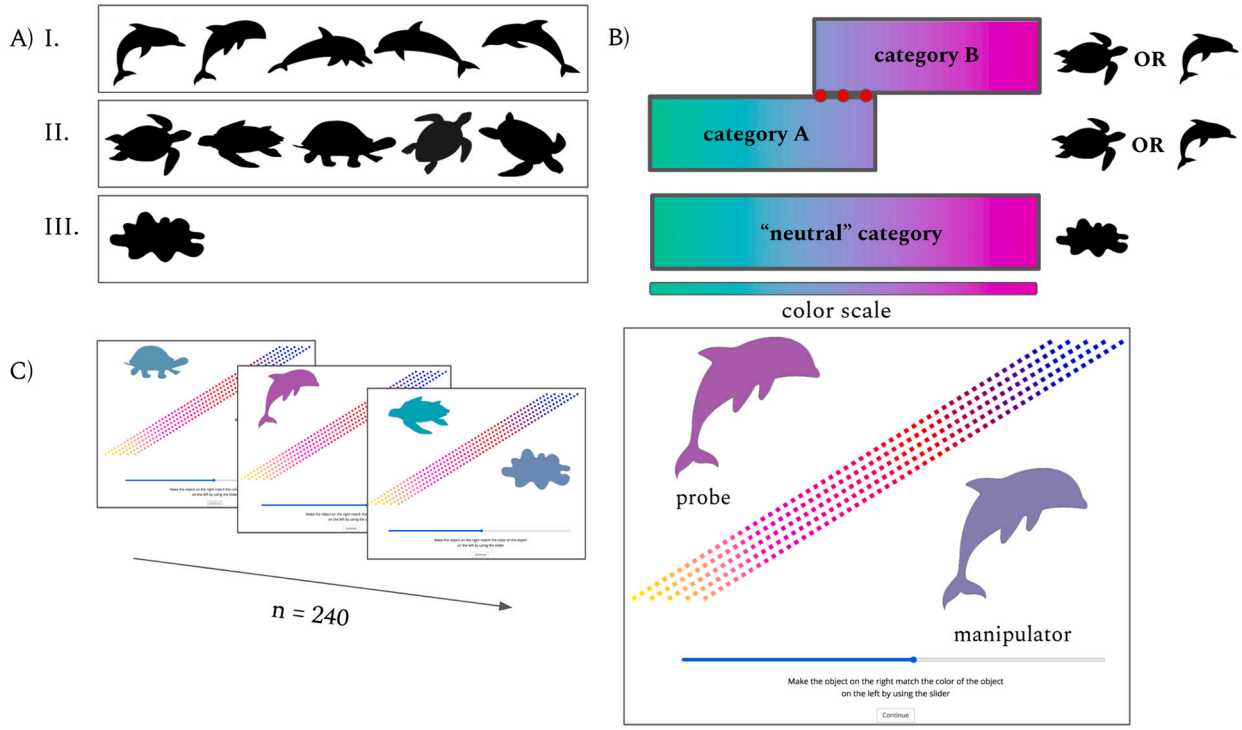

Fig. 3. A) Stimuli contours used in the experiments: I. Category 1 ("dolphins"), II. Category 2 ("turtles"), III. "Neutral" category not associated with any color distribution. B) Color distributions of the probe items' colors for two categories. Red dots in the middle of the color scale represent the colors that were used for both categories. C) Experimental procedure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 compensate for the opposite-direction biases. 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" manipulators' effect will lay somewhere in between. These distortion patterns would indicate that categorization affects both manipulator and the probe (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).

Our testing of whether two items are equally affected by the categorical bias is agnostic about the direction of this bias: whether the categorical bias makes the middle colors appear more similar to the categorical shape-based color distribution (assimilation) or more different from it (contrast). Based on the results of previous studies of effects of category learning on color matching (e.g. Cibelli, Xu, Austerweil, Griffiths, & Regier, 2016; Goldstone, 1995), we might expect that the shape-based categories bias the perception of the colors towards the learned color distributions corresponding to these categories (assimilation). In this case, participants will estimate middle colors as more green than they are if the shape comes from the "green" category, and more purple than they are if the shape comes from the "purple" category. However, category-based contrast effects have also been observed (e.g. Dubova & Moskvichev, 2019), and would result in the opposite tendencies.

2.2. Analysis

De-identified data and scripts for reproducing analyses and figures are available at https://osf.io/uprch/.

2.2.1. 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 was larger than 13 (1 participant).
2. We also excluded data from individual trials where a participant was too far off from the actual stimulus color – in particular, when the absolute deviation was larger than 18 (128 trials).
3. For hypothesis testing, we only used trials where the probe was filled with one of overlapping colors that were present in both categories (e.g. the 3 middle color values that appeared in each category).

After the data exclusion, the dataset of trials for the hypothesis testing consisted of 3634 color matching trials.

The confirmatory 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 (1 subject), as well as trials with the absolute matching error higher than 3 standard deviations from the average matching error (64 trials). After the exclusion, this additional dataset for hypothesis testing consisted of 3641 matching trials.

2.2.2. 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 Stan (Carpenter et al., 2017) for performing all Bayesian analyses in this paper and followed the reporting guidelines and the ROPE-test from the bayestestR package (Makowski, Ben-Shachar, Chen, & Lüdecke, 2019; Makowski, Ben-Shachar, & Lüdecke, 2019). We used the following formula for the

mixed linear regression models:

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

We expected a significant effect of probe and manipulator category 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 et al., 2016; Bae, Olkkonen, Allred, & Flombaum, 2015).

2.3. 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 color matchings were biased away and not towards their respective category color distributions (Fig. 5A & Fig. 6A). The effect of the manipulator category (3 levels: "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 predictors for the response slider direction (normal or reverse) and slider starting position to the model. The significance of the reported effects and 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 were consistent with the results of frequentist mixed linear regression analysis (see Appendix A for the details).

3. Experiment 2

One could argue that the effect that we found in Experiment 1 could have resulted from associative learning of colors with particular shapes, rather than the higher-level semantic categories (dolphins and turtles) that the shapes belong to. In this next experiment, we aimed to replicate our main result in an online-based study and to control for the effect of particular shapes by reserving some animal contours to be used only with the test, overlapping, colors.

3.1. Method

3.1.1. Participants

We collected data from 80 undergraduate students of Indiana University who received course credit for their participation and were recruited with the Indiana University SONA system. The sample size was the same as the sample size planned for Experiment 1. Two participants completed the experiment two times and the order of their attempts was not recorded, so the data from these participants were not included in further analyses. We collected data from two more participants to yield a total number of 80 participants. All participants received an informed consent form and provided their assent. The study was approved by the Indiana University Institutional Review Board.

3.1.2. Stimuli

We used exactly the same stimuli and color distributions as we used in Experiment 1. For each participant, one turtle-shaped and one dolphin-shaped contour were randomly chosen as *transfer* stimuli. These transfer stimuli were only presented with the test (overlapping) colors and, hence, can be used to test whether the categorical effect results from associating color distributions to particular shapes or to the higher-level categories (e.g. dolphins and turtles) to which the shapes belong.

3.1.3. Procedure

We used the experimental procedure from our first experiment. In contrast to the first experiment, only four out of five contours were presented with all the colors of their category. The remaining transfer contours were only used with the test colors in which the distributions for the two categories completely overlapped. These transfer contours were used for half of the trials with these testing colors as probe items, and as manipulator items only in the identical shape condition when the transfer contour was chosen as a probe (~12.5% of matching trials with test colors).

We conducted this experiment online and the procedure could be found at https://pcl.sitehost.iu.edu/mdubova/colorReproduction/colorReproduction_transfer.html.

3.2. Analysis

De-identified data and scripts for reproducing analyses and figures are available at https://osf.io/uprch/.

3.2.1. Data exclusion criteria

We used the data exclusion criteria from Experiment 1. Two participants and 163 trials were excluded with the pre-registered exclusion criteria, resulting in 3730 matching trials for the confirmatory analysis. Two participants and 41 trials were excluded with the data-driven exclusion criteria, resulting in 3740 matching trials for the confirmatory analysis.

3.2.2. Hypothesis testing

Relative to Experiment 1, the mixed linear model and an alternative, Bayesian mixed linear model, had an additional predictor that indicated whether a given trial involved a transfer shape and its interaction with a probe category:

Signed deviation from the actual stimulus' color ~ probe category + manipulator category + transfer contour + probe category * manipulator category + transfer contour * probe category + trial number + (1|subject)

We expected to find a main contrast effect of probe category and no significant effect of manipulator category.

3.3. Results

We replicated a significant negative effect of the probe category on the matching of the three overlapping colors across the two categories ("purple" category: est. = −1.71, 95% CI [−2.26,-1.17], $t = -6.14$, $p < 0.001$). The effect of the manipulator category was not significant ("purple" category: est. = 0.09, 95% CI [−0.4,0.58], $t = 0.37$, $p = 0.712$; "green" category: est. = −0.27, 95% CI [−0.69,0.15], $t = -1.25$, $p = 0.212$) (Fig. 5B & Fig. 6B). The effect of the transfer contour and its interaction with the probe category were significant (transfer contour: est. = −0.47, 95% CI [−0.82,-0.12], $t = -2.63$, $p = 0.009$; "purple" probe category *X* transfer contour: est. = 1.06, 95% CI [0.57,1.55], $t = 4.2$, $p < 0.001$), but they did not affect the main effect of the probe category. Thus, the influence of categories on color matching was found even for particular shapes that were only associated with the overlapping, middle colors, even though participants' matchings were less biased by the probe category in this case.

Adding a random slope for the probe category (probe category|

subject) to the model did not qualitatively change the results. Adding the predictors of the response slider direction (normal or reverse) and slider starting position on each trial to the model, again, did not affect the reported results or their significance, but there was a significant effect of the slider direction ("reverse": est. = 0.35, 95% CI [0.11,0.6], $t = 2.82$, $p = 0.005$). Data exclusion method (pre-registered or data-driven) did not change the results.

Results of the Bayesian mixed linear regression analysis were consistent with the results of frequentist mixed linear regression analysis (see Appendix A for details).

4. Experiment 3

After establishing that probe and manipulator perception are unequally affected by shape categories, we next aimed to uncover the factors that determine unequal influence of often identical stimuli by categorical biases. We hypothesized that due to the spacing between probe and manipulator items on the screen, participants could not foveate on two items at the same time. Therefore, they must either look between the two items to perceive both of them in the periphery, perceive one of the items in the periphery while focusing on the other item, or hold one item in a concurrent memory while comparing it to the other stimulus during the reconstruction. To determine the patterns of participants' perceptual interactions with the probe and manipulator stimuli, we measured participants' eye movements while they were performing the same reconstruction task as in Experiments 1 and 2.

4.1. Method

4.1.1. Participants

We collected data from 82 undergraduate students of Indiana University who received course credit for their participation. This experiment was conducted in a laboratory, as Experiment 1, and the participants were recruited with the Indiana University SONA system. The planned sample size was 80 as in Experiments 1 and 2, but two more participants showed up for the experiment session than anticipated. The participants were not aware of the experimental hypothesis. All participants received an informed consent form and provided their assent. The study was approved by the Indiana University Institutional Review Board.

4.1.2. Stimuli

We used exactly the same stimuli and color distributions as we used in Experiment 1. For each participant, one turtle-shaped and one dolphin-shaped contour were randomly chosen as *transfer* stimuli as in Experiment 2.

4.1.3. Procedure

We used the same experimental procedure that was used in the second experiment. We used the Webgazer.js tool (Papoutsaki et al., 2016) to record participants' eye movements through a web-camera during the experiment. Participants completed two webgazer calibration sessions at the beginning and in the middle of the experiment. Webgazer recorded the participants' gaze locations every 60 ms.

The experiment was executed in the Chrome browser in a laboratory context where each of 10 Apple iMac 24 in. computers was enclosed in a sound-dampening cubicle with a constant 5-watt, overhead illumination.

The procedure can be found at https://pcl.sitehost.iu.edu/mdubova/colorReproduction/color_reproduction_transfer_eye_tracking/colorReproduction_transfer_with_eye_tracking.html.

4.2. Analysis

De-identified data and scripts for reproducing analyses and figures are available at https://osf.io/uprch/.

4.2.1. Data exclusion criteria

We used the data exclusion criteria from Experiment 1. Three participants and 46 trials were excluded with the pre-registered exclusion criteria, resulting in 3788 matching trials for the confirmatory analysis. Three participants and 18 trials were excluded with the data-driven exclusion criteria, resulting in 3791 matching trials for the confirmatory analysis.

When analyzing and visualizing the eye-tracking data, we only excluded the trials where the tool did not record a single fixation around the stimuli (probe or manipulator) during the whole trial. The area around the stimuli that participants had to fixate on at least once during the trial was defined as a rectangle with the upper-left corner being 50px further (on both X and Y dimensions) than the upper-left corner of the probe, and 50px further than the lower-right corner of the manipulator (Fig. 4A). This resulted in excluding 267 out of 19,680 matching trials.

4.2.2. Hypothesis testing

We used a mixed linear model and an alternative, Bayesian mixed linear model, with the same formula as Experiment 2:

Signed deviation from the actual stimulus' color ~ probe category + manipulator category + transfer contour + probe category * manipulator category + transfer contour * probe category + trial number + (1|subject)

We expected to find a main contrast effect of probe category and no significant effect of manipulator category.

4.3. Results

4.3.1. Hypothesis testing

We replicated a significant contrast effect of the probe category on the matching of the three overlapping colors across the two categories ("purple" category: est. = −1.51, 95% CI [−1.99, −1.03], $t = -6.18$, $p < 0.001$). The effect of the manipulator category was not significant ("purple" category: est. = 0.13, 95% CI [−0.3,0.56], $t = 0.59$, $p = 0.55$; "green" category: est. = −0.19, 95% CI [−0.57,0.18], $t = -1.03$, $p = 0.3$) (Fig. 5C & Fig. 6C). The effect of the transfer contour and its interaction with the probe category were significant (transfer contour: est. = −0.54, 95% CI [−0.85,-0.24], $t = -3.49$, $p < 0.001$; "purple" probe category *X* transfer contour: est. = 0.76, 95% CI [0.32,1.19], $t = 3.44$, $p < 0.001$), but they did not affect the main effect of the probe category. Again, participants' matchings were generally biased by the probe category, but were less distorted when the probe shape was not associated with any color range (transfer contour).

Adding a random slope of the probe category (probe category|subject) to the model did not qualitatively change the results. Adding the predictors of the response slider direction (normal or reverse) and slider starting position on each trial to the model, again, did not affect the reported results and their significance, but there was a significant effect of the slider direction ("reverse": est. = 0.25, 95% CI [0.04,0.47], $t = 2.32$, $p < 0.05$). The data exclusion method (pre-registered or data-driven) did not change the results.

Both frequentist and Bayesian mixed linear regression analyses led to similar results (see Appendix A for details).

4.3.2. Eye-tracking analysis

We looked at the frequencies of trials with qualitatively different eye-movement patterns. On 4830 out of 19,413 trials (24.9%), participants fixed their gaze at the probe for more time than at the manipulator (Fig. 4B), and 1736 of them (8.9% of the total number of trials) they did not fixate on the manipulator at all (Fig. 4E). On 13,593 trials (70.0%), participants fixated on the manipulator for more time than the probe (Fig. 4C), and on 6337 of these trials (32.6% of the total number of trials), participants did not fixate on the probe at all (Fig. 4F). On 659 trials (3.4%), participants did not directly fixate on either probe or manipulator (Fig. 4D). Overall, participants were far more likely to look at the manipulator stimulus than the probe stimulus during a trial

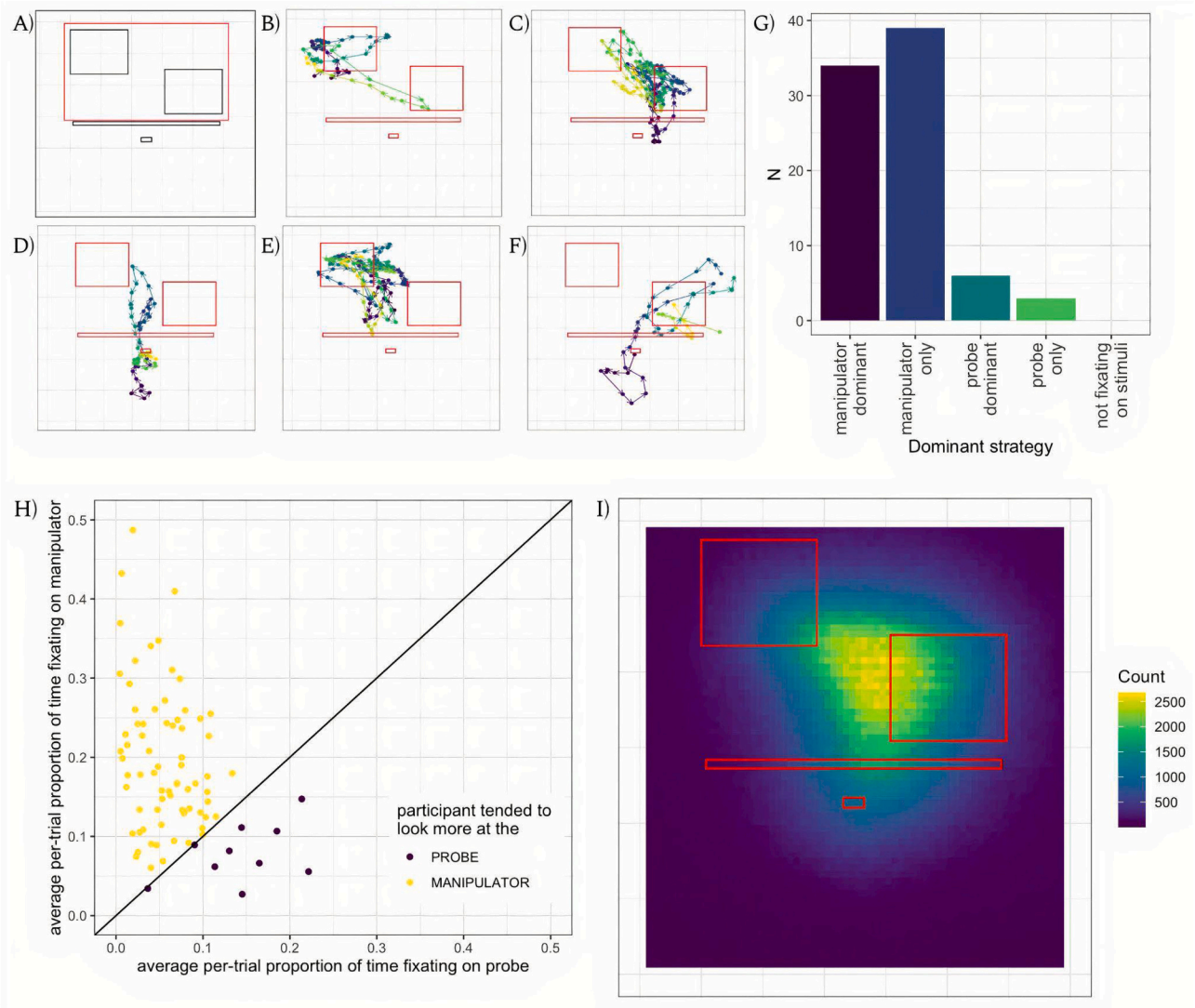

Fig. 4. Participants' eye-movements when they were completing the perceptual reconstruction task. The red rectangles (black on Fig. 4A) illustrate the probe, manipulator, slider, and the "next" button locations on the screen. A) The trial-by-trial exclusion criterion: trials were excluded from eye-movement analysis when not even a single fixation was recorded in the area around the probe and manipulator stimuli (red rectangle) during a trial. B) An example trajectory of eye movements from a trial where the participant looked at the probe more than the manipulator (color shows the time of the fixation: more yellow = later in the trial) ("probe dominant" strategy). C) An example trajectory of eye movements from a trial where the participant looked at the manipulator more than the probe ("manipulator dominant" strategy). D) An example trajectory of eye movements where the participant did not make any direct fixations on the probe or manipulator ("not fixating on stimuli" strategy). E) An example trajectory of eye movements where the participant fixated on the probe item, but did not fixate on the manipulator item at all ("probe only" strategy). F) An example trajectory of eye movements where the participant fixated on the manipulator item, but did not fixate on the probe item at all ("manipulator only" strategy). G) Distribution of participants with respect to their dominant looking strategy. H) Distribution of participants with respect to their average per-trial proportion of time fixating on the probe and manipulator items. I) Cumulative frequencies of fixations on all parts of the screen for all participants over all the trials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 4I).

We computed the most frequent (dominant) per-trial eye-movement pattern for each participant (Fig. 4G). We found that 34 out of 82 participants (41.5%) showed the "manipulator dominant" pattern of eye movements (Fig. 4C) more frequently than any other pattern across the trials: these participants tended to spend more time fixating on the manipulator than at the probe, but still looked at the probe at least once during the trial. Thirty nine out of 82 participants (47.6%) tended to fixate only on the manipulator stimulus on most of their trials ("manipulator only" strategy: Fig. 4F). Six out of 82 participants (7.3%) preferred to look at the probe stimulus while still moving the eyes back to the manipulator ("probe dominant" strategy: Fig. 4B) on the majority of trials. Lastly, 3 out of 82 participants (3.6%) preferred to only look at the probe stimulus when completing the matching trials ("probe only" strategy: Fig. 4E). There were no participants who preferred to look around the probe and manipulator stimuli without directly fixating on them on most of their trials (Fig. 4D).

Next, we looked at the per-participant average per-trial proportions of time fixating at the probe and at the manipulator. We found that 72 out of 82 participants, on average, looked at manipulator stimulus more than at the probe when they were completing the trials (Fig. 4H).

Finally, we analyzed which stimulus participants were likely to look at when they were making the slider movement. We used a mixed linear regression with a formula:

Actively moving a slider ~ looking at manipulator + looking at probe + looking at the slider + looking at the "next" button + (1|subject_id)

where the "looking at manipulator/probe/slider/'next' button"

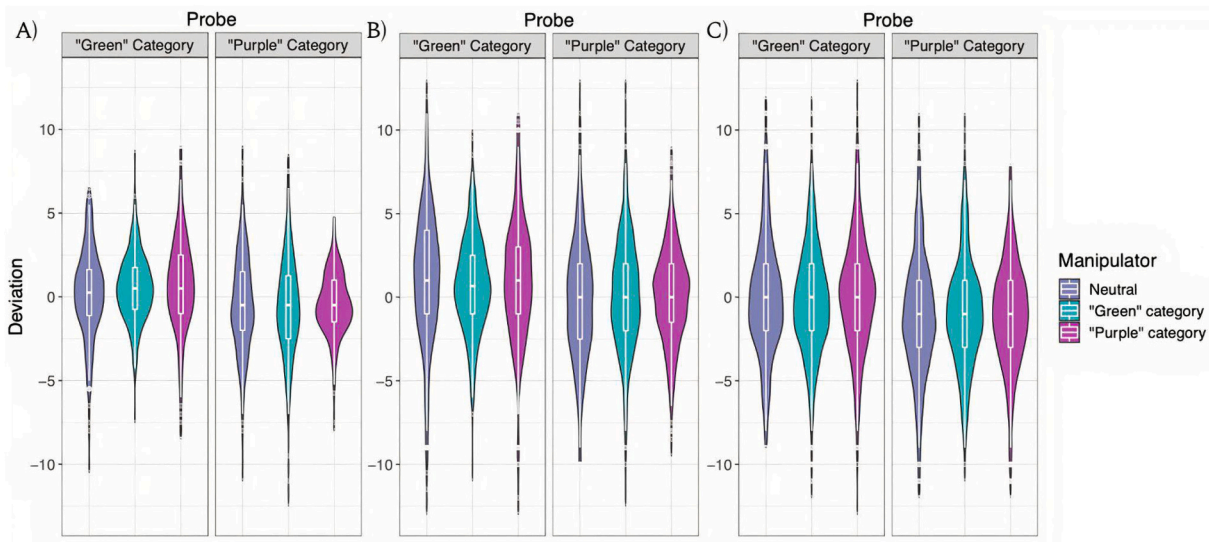

Fig. 5. Effects of the shape-based category and manipulator type on matching errors for the three testing colors that were present in both categories for the first (A), second (B), and third (C) experiment. Negative matching error represents distortion to the left (green) side of the color dimension, positive matching error reflects distortion to the right (purple) side of the color dimension. Notice that the probe items from the "green" category were more likely to be reconstructed as more purple than they were, no matter what category the manipulator item belonged to. Likewise, the probe items from the "purple" category were more likely to be reconstructed as more green than they were, no matter what category the manipulator item belonged to. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

variables correspond to exclusive binary indicators of whether participants are looking at the location of the respective stimulus at each moment of eye movement recording during the trial. The "actively moving the slider" target variable corresponds to a binary indicator of whether the participant actively moved a slider within the corresponding time window of eye movement recording.

We found a positive relationship between actively making a slider adjustment and looking at either the manipulator or probe, but the effect of looking at the manipulator was much stronger and explained more variance than the effect of looking at probe (looking at manipulator: est. = 0.11, 95% CI [0.11,0.11], $t = 142.63$, $p < 0.001$; looking at probe: est. = 0.03, 95% CI [0.02,0.03], $t = 24.8$, $p < 0.001$). Somewhat surprisingly, participants were unlikely to look at the slider itself or at the "next" button when they were moving the slider (looking at slider: est. = −0.08, 95% CI [−0.09,-0.08], $t = -42.58$, $p < 0.001$, looking at the "next" button: est. = −0.2, 95% CI [−0.21,-0.18], $t = -31.35$, $p < 0.001$).

5. Combined meta-analysis

Finally, we combined the data from all three experiments to robustly estimate the effects of interest. We followed the same exclusion criteria and analysis methods as we used for analyzing the data from individual experiments (see 3.2.2. and 4.2.2. for the linear mixed models formula).

We replicated a significant contrast effect of the probe category on the matching of the three overlapping colors across the two categories ("purple" category: est. = −1.241, 95% CI [−1.53, −0.95], $t = -8.44$, $p < 0.001$). The effect of the manipulator category was not significant ("purple" category: est. = 0.24, 95% CI [−0.03,0.51], $t = 1.75$, $p = 0.08$; "green" category: est. = −0.06, 95% CI [−0.30,0.17], $t = -0.51$, $p = 0.6$). The effect of the transfer contour and its interaction with the probe category were significant (transfer contour: est. = −0.40, 95% CI [−0.62,-0.19], $t = -3.68$, $p < 0.001$; "purple" probe category *X* transfer contour: est. = 0.70, 95% CI [0.42,0.99], $t = 4.78$, $p < 0.001$), but they did not modulate the main effect of the probe category. The model with a random slope for the probe category (probe category|subject) could not converge, so we cannot report the results. Adding the predictors of the response slider direction (normal or reverse) and slider starting position on each trial to the model, again, did not modulate the reported results or their significance, but there was a significant effect of the slider direction ("reverse": est. = 0.29, 95% CI [0.16,0.43], $t = 4.23$, $p < 0.001$). Data exclusion method (pre-registered or data-driven) did not change the results.

Both frequentist and Bayesian mixed linear regression analyses led to similar results. According to the Bayesian version of the linear mixed-effects model, the effect of "purple" probe category has close to 100% probability of being negative (median = −1.05, 95% CI [−1.28, −0.85] and can be considered as significant (ROPE test: 0% in [−0.1,0.1]). Analysis of the aggregate data did not show any significant effect of the transfer contour or interaction of the transfer contour and the probe category (transfer contour: median = −0.004, 95% CI [−19.06, 18.97], 0.82% in [−0.1,0.1] (not interpretable due to the wide range of posterior distribution); probe category *X* transfer contour: median = −0.17, 95% CI [−20.06, 19.69], significant: 0.86% in [−0.1,0.1] (not interpretable due to the wide range of posterior distribution)). The lack of evidence for the effects of the manipulator category and interaction of the probe and manipulator category were also replicated in the Bayesian analysis ("purple" manipulator category: median = 0.26, 95% CI [0.05, 0.48], undecided significance: 4.26% in [−0.1,0.1]; "green" manipulator category: median = 0.11, 95% CI [−11.28, 11.04], 1.47% in [−0.1,0.1] (not interpretable due to the wide range of posterior distribution); "purple" probe category X "green" manipulator category: median = 0.02, 95% CI [−10.95, 11.36], 1.55% in [−0.1,0.1] (not interpretable due to the wide range of posterior distribution); "purple" probe category X "purple" manipulator category: median = 0.02, 95% CI [−0.26, 0.30], undecided significance: 52.39% in [−0.1,0.1]; "green" probe category X "green" manipulator category: median = −0.25, 95% CI [−11.20, 11.13], 1.54% in [−0.1,0.1] (not interpretable due to the wide range of posterior distribution)).

6. Discussion

We conducted three experiments (Exp. 1 and Exp. 3: in laboratory; Exp. 2: online) to determine whether categorical associations equally affect color perception of all the objects in a visually immediate scene, as typically assumed by both proponents and critics of categorical

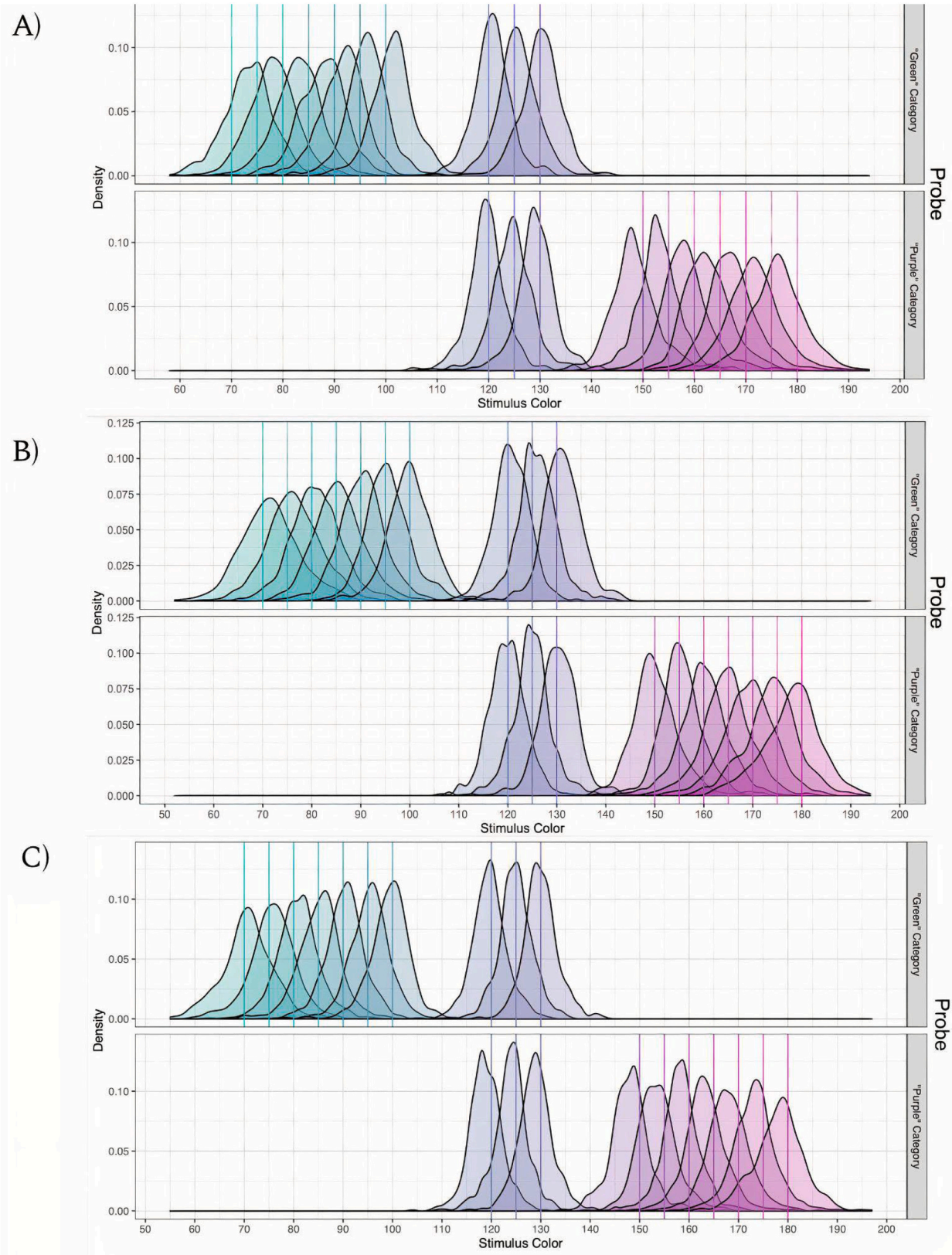

Fig. 6. Matching distributions for Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C). Vertical lines represent the actual colors of the probe on the color dimension, and the densities reflect the matching distributions for these colors. The physical colors of the densities and vertical lines reflect the actual colors that were presented to participants for each of these color conditions. Notice that the matching distributions are asymmetrically skewed, with the skewness increasing for more extreme items on the color dimension. Also, notice that the matching distributions are biased away from the extremes of the item's category color.

perception effects. In the experiments, we adapted a perceptual matching paradigm to minimize potential artifacts from putatively non-perceptual effects, such as memory and decision making.

For the outer, non-central items, we found that participants' perceptual color matchings were generally biased towards the average color in the middle (Fig. 6). This result is consistent with many perceptual and decision making accounts. First of all, the bias towards the global mean of the feature range is a strategy that minimizes the overall reconstruction error (decision making and predictive coding: e.g. Lupyan & Clark, 2015) and is consistent with altering reconstructions towards the prior color distribution of all the stimuli (Bayesian account; see Dubova & Goldstone, 2021 for a review of the bias towards an overall mean in perceptual reconstructions). Moreover, the colors of the probe item that was primarily kept in peripheral vision could have been perceived as less saturated (Gordon & Abramov, 1977), thus, biasing the reconstructions towards less saturated colors in the middle (see Appendix C for the saturation values for all the stimuli).

To study the potential *categorical* effects, we analyzed only reconstructions for the three colors in the middle of the color scale that were shared by both shape-based categories. In this way, we eliminated all the color stimulus-based artifacts as alternative interpretations of the category-based difference in the reconstructions. We found that participants' perceptual matchings of colors were biased by the shape-based category of the probe, but not the manipulator stimulus when both were simultaneously present on the screen. The specific probe shapes that were associated with the experimentally assigned categorical color distributions induced higher bias in perceptual matchings, however, the "transfer" probe shapes that were never associated with any color distribution still biased participants' matchings because these shapes belonged to categories that did have color associations. For example, if turtles tended to be purple for a participant, then a particular turtle shape that was only associated with a "neutral" color halfway between purple and green tended to be reconstructed as greener than it actually was. This provides evidence for an influence of laboratory-induced associations between colors and animal categories on color reconstructions.

There are several differences between manipulator and probe items that could cause their differential susceptibility to a top-down categorical bias. First of all, the probe color remains constant while the manipulator color is actively changed by the participants. Second, the positioning of two items on the screen is different: the probe item is presented at the top-left corner of the screen, while the manipulator is located at the bottom-right. Third, we found that participants demonstrated consistent and differential eye movements with respect to the probe and manipulator stimuli when they were completing the matching trials. The majority of participants either foveated on the manipulator item while perceiving the probe at the periphery (peripheral vision strategy), or foveated on manipulator item for the majority of time and occasionally moved their eyes to the probe item (concurrent memory and peripheral vision strategy). In both of these cases, the participants' perception was primarily focused at the manipulator item. Even though all three interpretations are possible, we prefer the third explanation as the simplest account: the categorical bias most likely affects the stimuli that the eyes are not focused on at the direct moment of the judgment (they are perceived at the periphery or reside in a concurrent memory encoding that operates at the timescale of moving the eyes from one part of the screen to another). This result is consistent with predictive processing and rational inference accounts of perception: they propose that the stimuli that are perceived with more noise or uncertainty are subject to stronger biases than the stimuli perceived more clearly (e.g. Knill & Richards, 1996; Lupyan & Clark, 2015; Witzel, Olkkonen, & Gegenfurtner, 2018).

All these possibilities for why the probe and manipulator are unequally affected by their categories are likely to be present during human perceptual interactions with the world. In particular, in most natural scenes, there will be stimulations at the top left and the bottom right of the visual field; some items are directly manipulated by us (e.g. when El Greco is drawing his figures) while the others remain static or are non-interactive; some objects are kept in peripheral vision or concurrent memory while others are directly focused on. We suspect that concurrent memory could be similar to transsaccadic memory which has been proposed to store the contents of objects of a scene during active perception. Again, we note that this kind of memory 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 (Edelman, 1989; O'Regan & Noë, 2001). Therefore, we suggest that the El Greco-fallacy assumption that all the items in a visual scene should be equally affected by top-down effects will not be met in many realistic scenarios. We hope that further studies clarify the properties that make some objects more susceptible to categorical biases than others.

Valenti and Firestone (2019) recently reported a similar observation. In this study, participants adjusted the color of a background (BG) to match the color of a foreground (FG) object. The authors measured the consistency of the matching bias depending on whether the foreground and background came from a category with a consistent color association ("heart" shape) or from a category with no particular color association ("square" shape), testing all their combinations ("heart" BG & "heart" FG; "heart" BG & "square" FG; "square" BG & "heart" FG; "square" BG & "square" FG). They found categorical influences on color matchings for different combinations of foreground and background categories that were inconsistent with the "El Greco logic" of how perceptual effects should simultaneously operate on the foreground and background objects (as in Fig. 2A.I). The authors interpreted their result as indicating the non-perceptual nature of the influence of shape of color perception and argued for the validity of "El Greco logic" predictions as a test of the perceptual nature of psychological effects. 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 minimizing clearly non-perceptual artifacts due to, for example, memory or decisional biases. Moreover, perceptual biases can easily be inconsistent with "El Greco" predictions, as these predictions are based on the false assumption of the perceptual equivalence of all items in a scene. While the "El Greco logic" can be used for testing whether a given bias equally affects different types of objects in a task, as we explored in this study, it cannot be used for differentiating between perceptual and non-perceptual effects.

We found that participants' probe color encodings were contrasted away from the probe category's caricature on the color scale, instead of being assimilated to the mean of the category-specific color distribution as we expected (Figs. 5 & 6). Imagine that dolphins are generally greener than turtles. A 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 color. Accordingly, the caricature of dolphins would be greener than the prototypical dolphin, perhaps as extreme as color = 60 (Fig. 6). Then, there was a tendency for all dolphins' reconstructions to be biased in the purple direction from their veridical values. This contrast-from-caricature effect is consistent with previous results on interrelated category representations. In particular, when simultaneously learned categories are frequently alternated, people often form caricature-based category representations that emphasize the diagnostic features that distinguish one category from others (Ameel & Storms, 2006; Davis & Love, 2010; Goldstone, 1996; Goldstone, Steyvers, & Rogosky, 2003; Levering & Kurtz, 2006; Palmeri & Nosofsky, 2001). Our results indicate that probe stimuli are encoded relative to such caricatured representations.

The contrast-from-caricature effect on perceptual concurrent memory encodings is inconsistent with many rational models of categorical effects on perception (e.g. Huttenlocher, Hedges, & Duncan, 1991; Feldman, Griffiths, & Morgan, 2009; Witzel et al., 2018; Bates & Jacobs, 2020). In particular, a rational perceptual encoding is commonly modeled as assimilation towards the relevant priors (e.g. the respective

category color distribution). Our data suggest that the concurrent memory encodings of some exemplars are contrasted away from, not biased towards, the perceptual prior distribution. Two recent studies used a similar matching paradigm to study the effects of language labels on color reconstruction (Bae et al., 2015; Cibelli et al., 2016). These studies showed a significant effect of prototypes associated with English color labels on perceptual matching, and this evidence was construed as supporting a rational model of adjusting color perception towards the label-associated category prototype (see also Huttenlocher, Hedges, & Duncan, 1991; Feldman, Griffiths, & Morgan, 2009; Bates & Jacobs, 2020). 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 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 representations (either prototypes or caricatures) is a commonly observed, if neglected, phenomenon. There are, however, at least two rational models of visual perception that predict contrastive, rather than assimilative, effects for stimuli that are far away from the observer's prior expectations, which is consistent with the category distributions and testing colors in our experiments (see Fig. 6) (Dubova & Moskvichev, 2019; Wei & Stocker, 2015). Further studies could helpfully clarify the role of category distributions in biasing visual encodings in different directions.

Declaration of Competing Interest

None.

Acknowledgement

Authors thank Reina Munoz for the help with data collection for Experiment 1. Authors are grateful to Arseny Moskvichev, Gunar Epping, Jerome Busemeyer, and Indiana University Statistical Consulting Center for their suggestions on modeling and data analysis and Eleanor Schille-Hudson for helpful suggestions on the earlier version of this manuscript. Authors also thank 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. Finally, authors thank six anonymous reviewers (from Cognitive Science Society Annual Meeting and Cognition journal), as well as Chaz Firestone and Jonathan Folstein, whose suggestions greatly improved the manuscript.

Appendix A. Stimuli colors on HSV scale

Here, we provide a table of the stimuli colors in both RGB and HSV scales (Table A.1, Table A.2).

Table A.1
Stimuli color values in RGB and HSV scales for category A.

Red	Green	Blue	Hue	Saturation	Value
70	185	180	177	62.2	72.5
75	180	180	180	58.3	70.6
80	175	180	183	55.6	70.6
85	170	180	186	52.8	70.6
90	165	180	190	50.0	70.6
95	160	180	194	47.2	70.6
100	155	180	199	44.4	70.6
120	135	180	225	33.3	70.6
125	130	180	235	30.6	70.6
130	125	180	245	30.6	70.6

Table A.2
Stimuli color values in RGB and HSV scales for category B.

Red	Green	Blue	Hue	Saturation	Value
120	135	180	225	33.3	70.6
125	130	180	235	30.6	70.6
130	125	180	245	30.6	70.6
150	105	180	276	41.7	70.6
155	100	180	281	44.4	70.6
160	95	180	286	47.2	70.6
165	90	180	290	50.0	70.6
170	85	180	294	52.8	70.6
175	80	180	297	55.6	70.6
180	75	180	300	58.3	70.6

Appendix B. Bayesian analyses results

Here, we present detailed results of the Bayesian confirmatory hypothesis testing for experiment 1, 2, and 3.

B.1. Experiment 1

According to the Bayesian version of the linear mixed effects model, the effect of “purple” probe category had a close to 100% probability of being negative (median = -0.89, 95% CI [-1.28, -0.49] and can be considered as significant (ROPE test: 0% in [-0.1,0.1]). Bayesian analysis also indicated lack of evidence for the effect of manipulator category and its interaction with the probe category (“purple” manipulator category: median = 0.38,

95% CI [-0.02, 0.78], undecided significance: 6.14% in [-0.1,0.1]; "green" manipulator category: median = -0.13, 95% CI [-11.45, 10.98], 1.32% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution); "purple" probe category X "green" manipulator category: median = -0.008, 95% CI [-11.21, 11.19], 1.47% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution); "purple" probe category X "purple" manipulator category: median = -0.29, 95% CI [-0.81, 0.24], undecided significance: 17.48% in [-0.1,0.1]; "green" probe category X "green" manipulator category: median = 0.12, 95% CI [-11.26, 11.16], 1.42% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution)).

B.2. Experiment 2

According to the Bayesian version of the linear mixed effects model, the effect of "purple" probe category had a close to 100% probability of being negative (median = -1.70, 95% CI [-2.17, -1.26] and can be considered as significant (ROPE test: 0% in [-0.1,0.1]). The effects of the transfer contour and its interaction with the probe category also replicated their direction and significance (transfer contour: median = -0.50, 95% CI [-0.77, -0.20], significant: 0% in [-0.1,0.1]; probe category X transfer contour: median = 0.92, 95% CI [0.52, 1.35], significant: 0% in [-0.1,0.1]). The lack of evidence for the effects of the manipulator category and interaction of the probe and manipulator category were also replicated in the Bayesian analysis ("purple" manipulator category: median = 0.09, 95% CI [-0.29, 0.50], undecided significance: 35.89% in [-0.1,0.1]; "green" manipulator category: median = 0.06, 95% CI [-11.05, 11.61], 1.38% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution); "purple" probe category X "purple" manipulator category: median = 0.35, 95% CI [-0.19, 0.88], undecided significance: 13.77% in [-0.1,0.1]; "purple" probe category X "green" manipulator category: median = 0.18, 95% CI [-11.42, 11.26], 1.79% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution); "green" probe category X "green" manipulator category: median = -0.32, 95% CI [-12.16, 10.60], 1.61% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution)).

B.3. Experiment 3

According to the Bayesian version of the linear mixed-effects model, the effect of "purple" probe category has a close to 100% probability of being *negative* (median = -1.40, 95% CI [-1.79, -1.03] and can be considered as *significant* (ROPE test: 0% in [-0.1,0.1]). The effects of the transfer contour and its interaction with the probe category also replicated their direction and significance (transfer contour: median = -0.39, 95% CI [-0.65, -0.15], significant: 0% in [-0.1,0.1]; probe category *X* transfer contour: median = 0.72, 95% CI [0.38, 1.07], significant: 0% in [-0.1,0.1]). The lack of evidence for the effects of the manipulator category and interaction of the probe and manipulator category were also replicated in the Bayesian analysis ("purple" manipulator category: median = 0.33, 95% CI [-0.01, 0.68], undecided significance: 7.47% in [-0.1,0.1]; "green" manipulator category: median = 0.05, 95% CI [-11.27, 11.38], 1.43% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution); "purple" probe category X "green" manipulator category: median = 0.18, 95% CI [-11.30, 11.37], 1.42% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution); "purple" probe category X "purple" manipulator category: median = -0.0008, 95% CI [-0.46, 0.46], undecided significance: 34.89% in [-0.1,0.1]; "green" probe category X "green" manipulator category: median = -0.17, 95% CI [-11.52, 11.18], 1.48% in [-0.1,0.1] (not interpretable due to the wide range of posterior distribution)).

Appendix C. Supplementary data

Supplementary de-identified data, scripts for reproducing analyses and figures are available at https://osf.io/uprch/. Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105025.

References

Ameel, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language, 55*(3), 402–421. https://doi.org/10.1016/j.jml.2006.05.005

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), 744–763. https://doi.org/10.1037/xge0000076

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological review, 127*(5), 891.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America, 63*(2), 456–468. https://doi.org/10.1121/1.381737

Busey, T. A., & Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin & Review, 17*(2), 155–160. https://doi.org/10.3758/PBR.17.2.155

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 076*(i01). https://ideas.repec.org/a/jss/jstsof/v076i01.html.

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), Article e0158725. https://doi.org/10.1371/journal.pone.0158725

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science, 21*(2), 234–242. https://doi.org/10.1177/0956797609357712

Dubova, M., & Goldstone, R. L. (2021). The influences of category learning on perceptual reconstructions. *Cognitive Science, 45*(5), Article e12981. https://doi.org/10.1111/cogs.12981

Dubova, M., & Moskvichev, A. (2019). Adaptation Aftereffects as a Result of Bayesian Categorization. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1669–1675).

Edelman, G. M. (1989). *The remembered present: A biological theory of consciousness*. Basic Books.

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. https://doi.org/10.1068/p7488

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. https://doi.org/10.1177/0956797613485092

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/S0140525X15000965

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science, 6*(5), 298–304. https://doi.org/10.1111/j.1467-9280.1995.tb00514.x

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24*(5), 608–628. https://doi.org/10.3758/BF03201087

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs. Cognitive Science, 1*(1), 69–78. https://doi.org/10.1002/wcs.26

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition, 31*(2), 169–180. https://doi.org/10.3758/BF03194377

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. Hue and saturation. *JOSA, 67*(2), 202–207.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368. https://doi.org/10.1038/nn1794

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

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

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*(2), 93–107. https://doi.org/10.3758/BF03212211

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

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*(9). https://doi.org/10.1525/collabra.32

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).

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science, 24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software, 4*(40), 1541. https://doi.org/10.21105/joss.01541

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology, 10*, 2767. https://doi.org/10.3389/fpsyg.2019.02767

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 939–973. https://doi.org/10.1017/S0140525X01000115

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. https://doi.org/10.1080/02724980042000084

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). *WebGazer: Scalable Webcam Eye Tracking Using User Interactions*. IJCAI. https://openreview.net/forum?id=rkZPfVMuWS.

Valenti, J. J., & Firestone, C. (2019). Finding the "odd one out": Memory color effects and the logic of appearance. *Cognition, 191*, Article 103934. https://doi.org/10.1016/j.cognition.2019.04.003

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. *Nature Neuroscience, 18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Witzel, C., Olkkonen, M., & Gegenfurtner, K. R. (2018). A Bayesian model of the memory colour effect. *I-Perception, 9*(3). https://doi.org/10.1177/2041669518771715, 2041669518771715.