COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e12981

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12981

The Influences of Category Learning on Perceptual Reconstructions

Marina Dubova, Robert L. Goldstone

Psychological and Brain Sciences, Indiana University

Received 7 April 2020; received in revised form 8 October 2020; accepted 2 April 2021

Abstract

We explore different ways in which the human visual system can adapt for perceiving and categorizing the environment. There are various accounts of supervised (categorical) and unsupervised perceptual learning, and different perspectives on the functional relationship between perception and categorization. We suggest that common experimental designs are insufficient to differentiate between hypothesized perceptual learning mechanisms and reveal their possible interplay. We propose a relatively underutilized way of studying potential categorical effects on perception, and we test the predictions of different perceptual learning models using a two-dimensional, interleaved categorization-plus-reconstruction task. We find evidence that the human visual system adapts its encodings to the feature structure of the environment, uses categorical expectations for robust reconstruction, allocates encoding resources with respect to categorization utility, and adapts to prevent miscategorizations.

Keywords: Perceptual learning; Categorization; Perceptual reconstruction; Top-down effects on perception

1. Introduction

Human perception is highly adaptable to the structure of the environment and individual cognitive demands. To serve an individual's information needs, perceptual systems have to both efficiently encode environmental inputs using limited resources and provide robust and refined information about cognitively useful features under different conditions (lighting, occluding, and different angles) in constantly changing environments.

The stimuli for all experiments, the data, and the scripts used to create the experiments, generate the plots, and run all the analyses are available on OSF: https://osf.io/2xr54/

Correspondence should be sent to Marina Dubova, Cognitive Science Program, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405. E-mail: mdubova@iu.edu

Adaptive perceptual changes can be divided into unsupervised (task-irrelevant) and supervised (task-relevant). Unsupervised perceptual learning refers to adaptation to the structure of sensory experiences alone. These effects include, but are not limited to, sensitization and habituation, assimilation to or contrast from the statistical norms of a set of stimuli, and adaptive allocation of encoding resources on the basis of stimuli statistics. Supervised perceptual learning, in turn, refers to adaptation that occurs on the basis of particular behavioral tasks with feedback. Many supervised perceptual learning effects have been reported and replicated, but most of them remain controversial (see Firestone & Scholl, 2016, for a critical review). Supervised effects include selective attention and tuning of perceptual encodings to task-relevant stimuli or dimensions, and assimilation to or contrast from the norms of the stimulus categories relevant for the particular task.

A broad range of evidence suggests that our perceptual systems adaptively respond to the demands of categorization and change on the basis of categories we learn. One phenomenon, acquired categorical perception, refers to situations when, after acquiring a categorical distinction, items belonging to the same category tend to be encoded as being more similar to each other than before the category learning (compression), and/or items from different categories appear more different (expansion) (Goldstone & Hendrickson, 2010). For example, we are experts at discriminating our native language phonemes, yet we face difficulties trying to understand the differences between other languages' sounds that are not reflected in the phonemic structure of our native language (Kuhl, 1991; Kuhl, 1994; Liberman, Harris, Hoffman, & Griffith, 1957). It is difficult for Japanese speakers to distinguish between "r" and "l," which is a trivial task for native speakers of English, where these sounds constitute different phonemic categories (Iverson et al., 2003). Similar effects are observed in human vision. Hues of the same color category ("yellow") are discriminated more slowly and less accurately than hues that receive different language-specific color labels ("yellow" vs. "orange") even when both pairs are equidistant in a psychologically scaled color space (Roberson, Davidoff, Davies, & Shapiro, 2005; Roberson, Davies, & Davidoff, 2000; Winawer et al., 2007). Categorical perception effects have also been reported for multidimensional stimuli, such as face identities (Beale & Keil, 1995) and facial expressions (Etcoff & Magee, 1992; Roberson, Damjanovic, & Pilling, 2007). Results of experiments with participants acquiring new artificial categories indicate that within-category items are identified as more similar than between-category items on the categorization-relevant dimensions (e.g., Goldstone, 1994; Goldstone, 1995; Levin & Beale, 2000; Livingston, Andrews, & Harnad, 1998).

Other studies suggest that the structure of the categories to be learned influences the fidelity of the perceptual encoding of different features in multidimensional stimuli. In particular, dimensions or feature ranges relevant for the categorical discrimination are encoded more accurately than the ones that are irrelevant to the categorization task (Bates, Lerch, Sims, & Jacobs, 2019; de Beeck, Wagemans, & Vogels, 2003; Folstein, Palmeri, & Gauthier, 2013; Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008; Hockema, Blair, & Goldstone, 2005; Livingston et al., 1998; Notman, Sowden, & Özgen, 2005). The effects of learned categorical perception have been shown to be robust across different measurement methods, such as similarity ratings, same-different judgments, and XAB tasks (de Leeuw, Andrews, Livingston, & Chin, 2016).

The empirical evidence for the existence of categorical perception effects has also been criticized as suffering from methodological pitfalls (Firestone & Scholl, 2016). Most of the criticisms question the ability of studies to differentiate between "pure perceptual" effects from the ones related to the mechanisms of memory or decision making. Therefore, the perceptual nature of the reviewed effects (and even the name of the subfield: categorical *perception*) remains controversial. In this study, we aimed to elucidate whether category learning affects perception by using an experimental paradigm which, in our view, is less prone to the alternative interpretations of the results.

In this paper, we suggest that the commonly used perceptual discrimination (such as Goldstone, 1994) and similarity judgment (e.g., Livingston et al., 1998) tasks are not sufficient to reveal the nature of different perceptual learning effects that could manifest themselves in a single experimental situation. They are data-inefficient and provide poor one-dimensional measurements of the dynamic, multidimensional processes of perceptual learning. Considering the variety of potential perceptual learning mechanisms that could affect the results of a single study, we suggest using more data-efficient and rich experimental paradigms that could differentiate between these mechanisms and reveal their possible interplay.

We propose a relatively underutilized way of studying potential categorical perception effects using interleaved categorization and reconstruction tasks. We articulate various supervised and unsupervised perceptual learning mechanisms and test the predictions they make for performance in our reconstruction task. We find that some effects of categorical perception manifest themselves in a stimulus reconstruction task even when a probed stimulus is present on a screen at the time when a perceptual judgment bearing on it is made. Our results suggest that multiple, distinguishable categorical perception effects reflect changes in perceptual reconstructions.

2. Interleaved categorization-reconstruction task

During the experiment, participants perform both categorization and stimulus reconstruction tasks (Fig. 1), where stimulus reconstruction trials follow categorization trials with a constant rate (20% in our case). On categorization trials, participants see a stimulus on the screen, guess which category it belongs to, and then receive immediate feedback on the accuracy of their response. The reconstruction task requires participants to reproduce the "probed" stimulus as accurately as possible by adjusting a second "matching" stimulus which is presented on the screen at the same time (see https://www.indiana.edu/~pcl/robsexperiments/tests&examples/fishCatReproducevisible.html for a demonstration of the experiment). This can be done by using a mouse to move a cursor on the screen where x and y axes correspond to different dimensions of the image to be reconstructed. As participants move the cursor, they see instantaneous changes in the reconstructed image and can submit their response whenever they think that the reconstructed image exactly matches the probed stimulus currently present in another part of the screen.

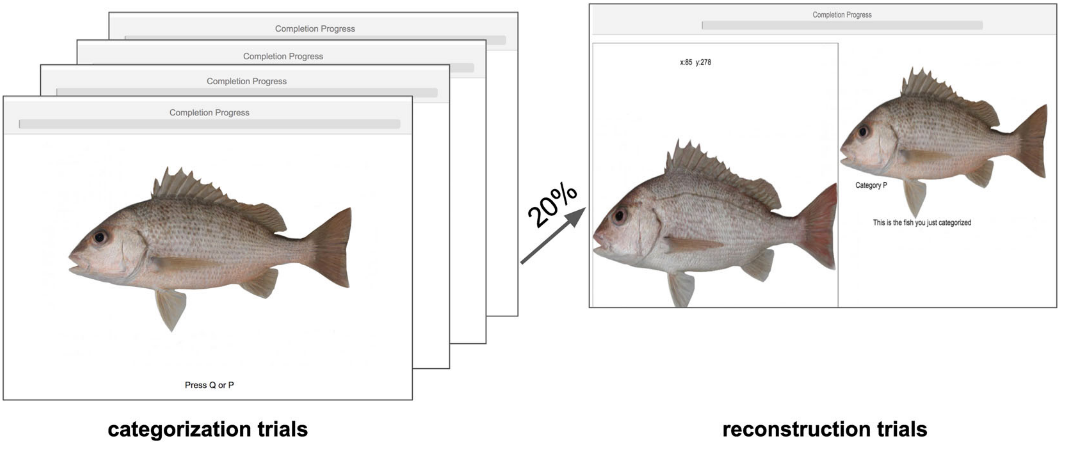

Fig. 1. Interleaved categorization-reconstruction task.

Previous research has employed a reconstruction task for studying categorization and its effects on perception. For example, Busemeyer and Myung (1988) proposed a similar paradigm to study how a learner's category prototype changes over training. Goldstone (1995), Huttenlocher, Hedges, and Vevea (2000), and Duffy, Huttenlocher, Hedges, and Crawford (2010) used one-dimensional reconstruction tasks where the matching stimulus was presented before or at the same time as its reconstruction was performed, to study categorical effects on perception. In contrast to this previous work, in the current study, we use this experimental paradigm to investigate how categorization affects perception of multidimensional naturalistic objects, as opposed to studying the encoding of single dimensions such as size, hue, or brightness.

We suggest that the reconstruction task used in our study weakens the alternative interpretations of potential effects as resulting from judgment and memory-based processes that are irrelevant to everyday perception. In particular, this experimental paradigm minimizes the *scale-dependent* judgment effects by using the reconstruction response scale that corresponds to the physical scale of the stimuli. Moreover, the *immediate* reconstruction task eliminates the long-term memory interpretations of the results. Thus, we suggest that any processes that affect stimulus reconstruction in this multidimensional perceptual reconstruction task are likely to play a role in everyday operation of the human visual system.

We propose that the multidimensional perceptual reconstruction task can provide rich data on unsupervised and supervised (categorical) perceptual learning with far fewer trials than perceptual match or discrimination tasks. The reconstruction trials are beneficial for investigating the potential *trajectories* of perceptual change on the reconstruction space and over the training, which can shed light on how categorization influences visual perception of multidimensional objects and their features, such as overall shape or texture. These data are valuable for studying the dynamics of perceptual changes and the interplay of different learning mechanisms underlying them.

2.1. *Potential effects of perceptual learning*

Different unsupervised and supervised learning effects may manifest themselves in an interleaved categorization-reconstruction study. Unsupervised effects result from experiencing a particular clustering of stimuli (in our case, fish images) over the time of the experiment. Supervised effects correspond to potential changes in perceptual encoding resulting from learning to categorize stimuli with feedback. In our review, we try to focus on computational level models (Marr, 1982) and note that there are many insightful algorithms incorporating task-relevant and task-irrelevant perceptual learning that deserve a separate paper. By comparing computational rather than algorithmic models, we also do not make any additional assumptions on particular mechanisms (e.g., attention) that induce the characterized adaptive changes in perception.

2.1.1. *Unsupervised effects*

In this subsection, we briefly review potential unsupervised effects of perceptual learning that might affect the reconstruction of stimuli in our task (Fig. 2). Unsupervised influences of stimuli on encoding do not involve any feedback about their correct categorization. Instead, these influences come from statistical properties inherent to the distribution of stimuli themselves.

2.1.1.1. Perceptual differentiation. The ecological perception perspective (Gibson, 1969; Gibson & Gibson, 1955) suggests that perceptual learning corresponds to the improvement of an individual's ability to extract features from the environment. It has been suggested that individuals' perceptual experiences refine over time as a result of active perceptual tuning. Perceptual differentiation effects have been claimed to be responsible for differences in expert and novice perceptual recognition and estimation. For example, professional musicians are better than novices in recognizing pitches and sounds of different instruments, while habitual birdwatchers are better than other people in looking for, detecting, and differentiating bird species. In the context of our study, perceptual differentiation effects would correspond to the refinement of stimulus reconstructions with training. If differentiation occurs, the average deviation of an observer's reconstruction of a stimulus from its actual value should decrease with experience for all stimuli dimensions (Fig. 2a).

2.1.1.2. Bayesian perception. Bayesian accounts claim that perception is a rational act of reconstructing (inferring) the source perceptual features. It works by integrating received sensory data (noisy and ambiguous) and prior beliefs on the probabilities of different features' occurrences (Geisler, 2003; Knill, Kersten, & Yuille, 1996). From a Bayesian perspective on unsupervised perceptual learning, each received sensory stimulation updates the expectations (prior beliefs) on the probability of its occurrence. The prior expectations on perceptual experiences might be hierarchical, with different distributions for the perceptual features of different complexity: from pixel-wise brightness to shapes, textures, and patterns manifest on larger spatial scales, and conditional on other information present. After receiving new sensory data, a Bayesian observer integrates the prior expectations of how likely different

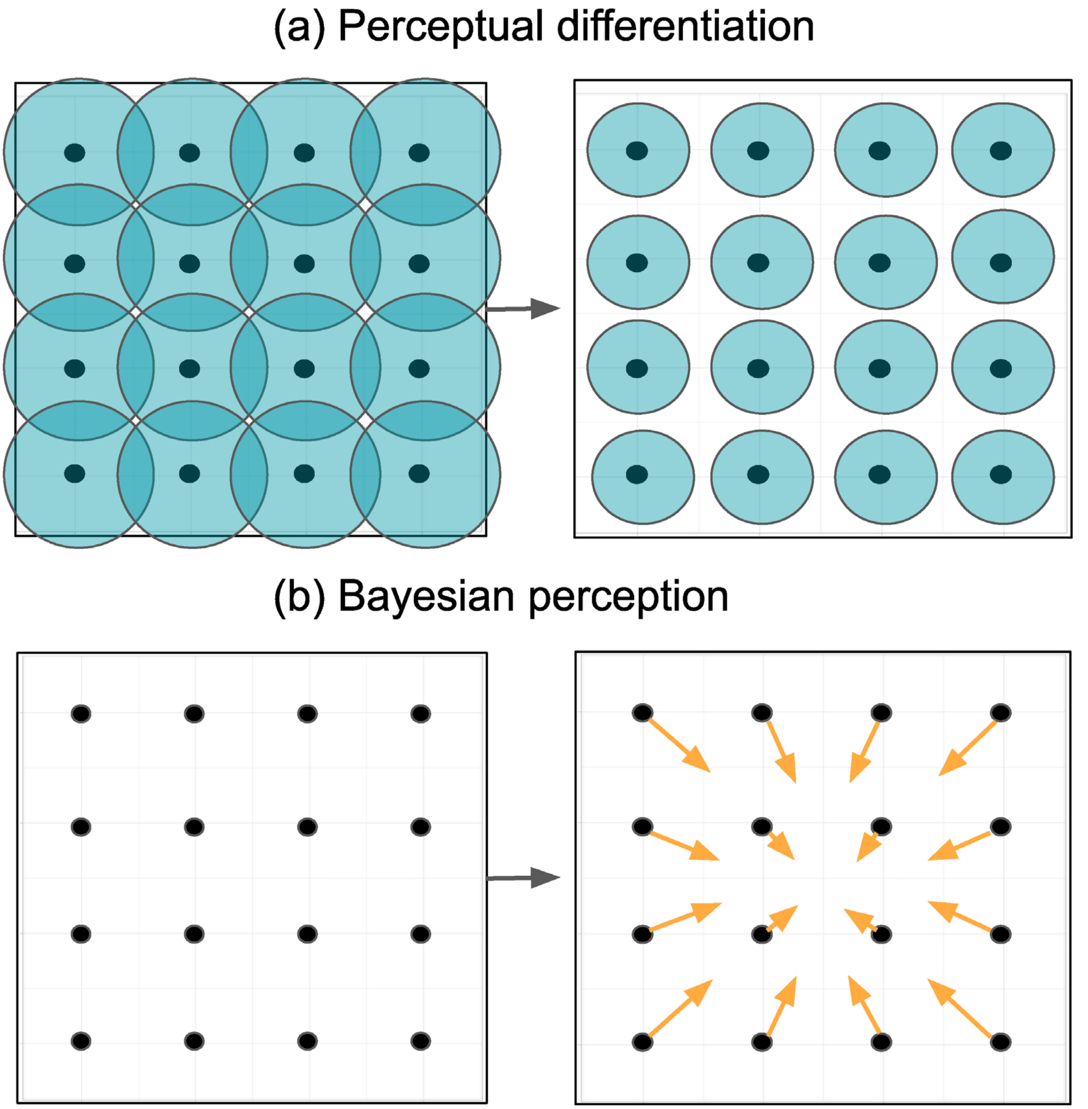

Fig. 2. Potential unsupervised perceptual learning effects. Left: before learning; right: after learning. Black circles correspond to the probed stimulus locations in the morph space. (a) Blue circles represent the variance of the perceptual reconstruction. (b) Yellow arrows represent systematic distortions (biases) in perceptual reconstruction. Note: This is a simplistic illustration of the expected qualitative effects. The exact values and ratios are arbitrary.

percepts are with the actual data. As a result, the perception of a particular feature or a combination of features is shifted toward the prior probability distribution (Kersten, Mamassian, & Yuille, 2004; Körding, 2007). Bayesian perception effects rely on the formalization of the priors, whereas in most cases, and for our fish images in particular, the perceptual priors are unknown. For the purpose of this paper, we start with the weak assumption of a uniform

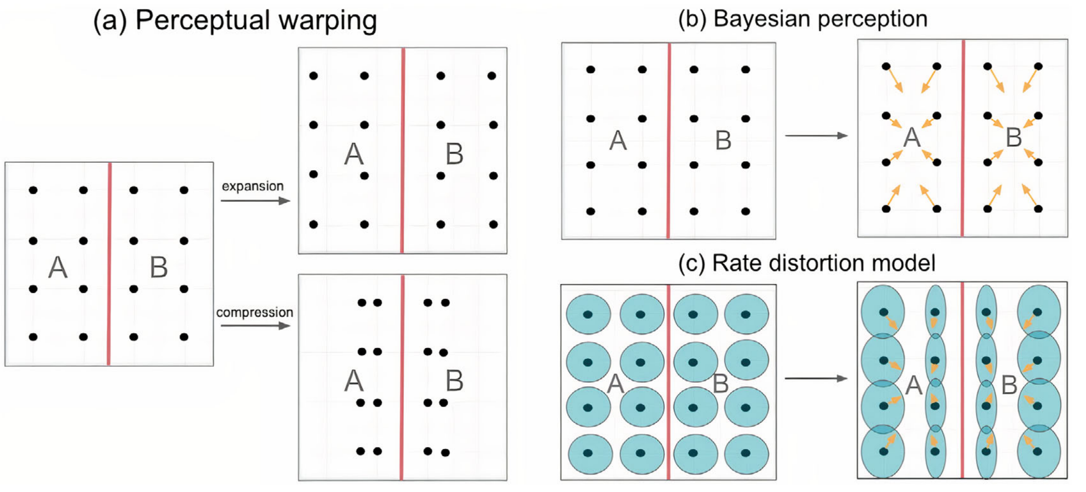

Fig. 3. Potential supervised perceptual learning effects. Left: before learning, right: after learning. Black circles correspond to the probed stimuli locations in the *psychological representation space* (a) or the *morph space* (b, c). The red vertical line represents a horizontal categorical boundary. (b) and (c) Yellow arrows represent the distortion (bias) in the perceptual reconstructions. (c) Blue circles represent the variance of the perceptual reconstruction. Note: This is a simplistic illustration of the expected qualitative effects. The exact values and ratios are arbitrary.

distribution prior for the stimuli's feature ranges used in our experiment. Then, due to the uniform distribution of the dimension values' occurrences in our stimuli, unsupervised learning of a Bayesian observer will result in a *global assimilation* effect (Fig. 2b). The reconstructed stimuli would shift toward the overall mean values on both dimensions, with central stimuli being shifted less than stimuli on the periphery of the dimensional space. The rate distortion model (Sims, 2016; discussed in detail in the next section) makes a similar prediction of the global assimilation effect in a case of unsupervised learning.

2.1.2. Supervised effects

In this subsection, we review potential effects of categorization on perception, drawing from previously proposed explanatory accounts and computational models (Fig. 3).

2.1.2.1. Perceptual warping (expansion or compression). The simplest high-level explanation for categorical perception effects suggests that the perceptual representation space undergoes transformations as a result of categorization training. In particular, two types of warping have been proposed: *compression* leading to items from the same category moving closer together, and *expansion* that leads to items from different categories moving farther away from each other in a perceptual representation space (Livingston et al., 1998; for a similar proposal for auditory perception, see Kuhl, 1994) (Fig. 3a). Simulations of simple neural network models have been used to support this perspective (Goldstone, 2003; Harnad, Hanson, & Lubin, 1991; Harnad, Hanson, & Lubin, 1995). For example, Harnad et al. (1991) trained a shallow neural network with backpropagation to discriminate and then categorize stimuli. Their simulations suggest that the network achieves good performance in the second task by altering distances of internal representations for different category members ("warping"). In our experiment, perceptual warping would be expected to lead to the tendency of the reconstructions not to cross the categorical boundary (*boundary aversion* effect). Compression might also be thought to occur along the dimension that is irrelevant for categorization. However, such compression effects might be the result of greater selective attention to the category-relevant relative to category-irrelevant dimension. Accordingly, for supervised perceptual warping, we restrict ourselves to the comparison of categorization-relevant dimension differences that either do or do not cross category boundaries, reserving discussion of differential learning effects on category-relevant versus category-irrelevant dimension to later (in the rate distortion section).

2.1.2.2. Bayesian perception. From a Bayesian perspective, category learning represents gradual belief accumulation on the perceptual features conditioned on the category. Categories are represented as probability densities in the one- or multidimensional feature space, and these representations are updated after seeing each new category example. Perception of new instances of a category is formalized as Bayesian inference of perceptual features integrating prior beliefs (a probabilistic representation of the category in the feature space) and the received sensory data. Bayesian observer models, such as the Category Adjustment Model (Huttenlocher et al., 2000), predict that the perception of new exemplars shifts toward the category prototype with the closer items being shifted less than the farther ones for our within-category stimuli distributions (see also the adaptations of this model for auditory domain and Sapir–Whorf effects in color perception: Cibelli, Xu, Austerweil, Griffiths, & Regier, 2016; Feldman, Griffiths, & Morgan, 2009). In our experiment, the supervised Bayesian perception effects would lead to the *assimilation* of reconstructions toward the categorical means on both dimensions (Fig. 3b).

2.1.2.3. Rate distortion model. Information theory adds one more component to the Bayesian observer story—the optimal encoding of the sensory signals in a limited capacity perceptual channel. There are infinitely many gradations of the physical sensory signals, while our perceptual systems can represent only a limited number of them. The rate distortion model formalizes an optimal way to allocate limited perceptual coding resources that minimizes the cost of perceptual error, where the latter reflects how deleterious it is to misrepresent a particular sensory stimulation (Bates & Jacobs, 2020; Sims, 2016; Sims, 2018). The supervised categorization task adds a cost of perceptual misrepresentation to the categorization-relevant dimensions or stimulus ranges. Hence, the rate distortion account claims that more encoding resources will be allocated to the categorization-relevant stimulations. For perception of multidimensional stimuli, it predicts the *perceptual* "*tuning*" for categorically diagnostic dimensions, resulting in more accurate encoding along them (Bates & Jacobs, 2020) (Fig. 3c; Bates et al., 2019). The prediction of greater precision of encodings for category-relevant than category-irrelevant dimensions is also made by selective attention accounts in which learning is suggested to increase the salience of dimensions that are relevant, compared to irrelevant, for a category discrimination task (Lawrence, 1949).

Depending on the details of the cost function, the rate distortion model could also suggest that the items close to the category boundary are represented more accurately as the cost of their misperception is higher. The optimal encoding model with a Bayesian decoder naturally predicts perceptual assimilation toward the category prototype, with higher assimilation for the categorization-irrelevant dimensions which are represented less accurately (Fig. 3c). Strength of the assimilation effect toward the category prototype depends on the amount of encoding resources available and relative costs of different tasks perceptual systems deal with (e.g., categorization and reconstruction) (Bates & Jacobs, 2019; Bates & Jacobs, 2020).

The aforementioned Bayesian perception models and the rate distortion model both formalize categorical perception effects as resulting from rational computations. However, they conceptually differ in their assumptions regarding the goal of these adaptations. The Bayesian models view categorical effects as shaping perception to make it more robust to noise, ambiguity, and changing conditions. The rate distortion model additionally claims that these effects reflect the adaptiveness of perception to better serve categorization needs. The latter view is compatible with ecological approaches to perception that claim that we learn to extract information useful for the tasks we deal with (Gibson, 1979) and a Bayesian ideal observer framework incorporating utility functions related to the tasks (not necessarily perceptual) (Geisler, 2003; Geisler & Diehl, 2003).

The processes described by the Bayesian and rate distortion models of categorical perception can be motivated by predictive coding, the idea that human cognition aims to minimize its surprise about its environment (Clark, 2013; Friston & Kiebel, 2009), viewing perception as a constructive inferential process affected by those expectations (Barlow, 1990). Such a system uses any resource that helps it to predict the upcoming information, including perceptual features of an object or its categorical label (Lupyan, 2015; Lupyan & Clark, 2015; Lupyan et al., 2020). Optimizing one's prediction naturally leads to biases in these perceptual predictions, well described by the aforementioned computational models, depending on the assumed task structure and resource limitations. By an alternative account, the supervised (e.g., category-dependent) distortions are motivated by reinforcement learning, which often requires concise, robust, and task-motivated representations (Gershman & Niv, 2010; Niv et al., 2015).

3. Method

3.1. Participants

Sixty-three undergraduate students from Indiana University participated in the experiment to fulfill course requirements. Participants were pseudo-randomly assigned to the horizontal and vertical categorization rule conditions when they started the experiment on a computer.

3.2. Stimuli

Two hundred and twenty-five morph images of fish were created in Abrosoft FantaMorph to constitute the stimuli for the experiment. We started with four actual fish images to create two different axes in the feature space (Fig. 4a). The horizontal dimension was conceptualized as morphing from an image of Pagrus auratus (a species of the fish commonly known

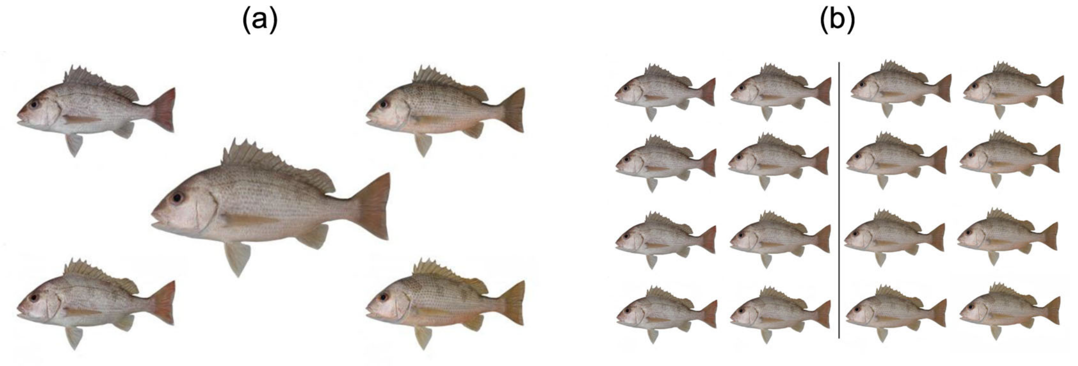

Fig. 4. Morph stimuli used in our experiment. (a) Four source fish images used to form x and y axes, and the morph image in the middle. (b) Morph stimuli for the categorization trials with the horizontal categorical boundary being shown.

as snapper) to Lutjanus johnii (another species of snapper). The vertical dimension morphed between Pomadasys kaakan (a species of javelin fish) to Pomadasys maculatus (javelin). Every fish in a 15×15 matrix of fish was constructed by integrating the images corresponding to its horizontal and vertical dimensions. Thus, the upper left-hand fish of the 15×15 matrix consisted of a morph halfway between Pagrus auratus and Pomadasys kaakan. The lower-left fish was halfway between Pagrus auratus and Pomadasys maculatus. The upper-right fish was halfway between Lutjanus johnii and Pomadasys kaakan. The lower-right fish was halfway between Lutjanus johnii and Pomadasys maculatus. The rest of the matrix of fish was populated by creating equally spaced, linearly interpolated images in both shape and texture components. For example, the fish located at the coordinate {10,5} was two-thirds of the way from Pagrus auratus to Lutjanus johnii on the horizontal dimension, and one-third of the way from Pomadasys kaakan to Pomadasys maculatus on the vertical dimension. Shape morphs were created by first regularizing the four source fish images so that they had roughly equal outer shape outlines. Eighty-eight aligned control points were identified for each of the four images based on relatively clear image landmarks, such as the leftmost point on the eye, the tip of the dorsal fin, and tip of the lower lip. None of the stimuli used in the experiment correspond to the source images.

All 225 stimuli could be shown during the reconstruction phase of the experiment, but only a subset of the fish were used during categorization. In particular, categorizations involved the 4×4 subset of the 225 fish located at positions 2, 5, 8, and 11 out of the 15 possible values along each of the two dimensions (Fig. 4b). Fish images were assigned to different categories (A and B) relative to their location in the two-dimensional morph space. For the horizontal categorization rule condition, the stimuli were evenly distributed to the categories on the basis of their location on the x-axis. For the vertical categorization rule condition, categorical labels were assigned on the basis of stimuli locations on the y-axis. Thus, values of 2 and 5 were assigned Category "P" and values 8 and 11 were assigned Category "Q" for roughly half of the participants, and vice versa for the other half.

3.3. Task and procedure

Every trial consisted of a categorization trial with the image randomly selected from the 4×4 matrix of categorization fish. On random 20% of categorization trials, a reconstruction trial immediately followed categorization, and on the remaining trials, one categorization trial immediately followed another.

During categorization trials, participants saw a fish morph image and guessed its category (P or Q). They received immediate feedback after their response in the form of either "Incorrect. This fish is [P/Q]" or "Correct. This fish is [P/Q]."

Reconstruction trials required participants to match the fish image on a left side reproduction window to the probed exemplar which was present on the right side of the screen. The category label ("P" or "Q") appeared next to the probed fish image on the screen, and below the fish there was a reminder message: "This is the fish you just categorized." Participants adjusted the reconstructed fish image by moving the mouse within the reconstruction window. The initial mouse positions were determined by the last position on the previous trial, and the image on the reproduction window corresponded to its location in the 15×15 2D morph space. The image on the reconstruction window gradually changed as participants moved the mouse so that from the participants' perspective, the reconstructed image looked like it was dynamically morphing as participants moved their mouse.

We recorded the categorization judgments (for categorization trials) and the coordinates of the reproduced images on the 2D morph plane (for reconstruction trials).

4. Results

4.1. Analysis

Data were excluded prior to the analysis on the following bases:

1. Participants who did not finish the study (one participant).
2. Participants who accidentally participated in both experimental conditions and it was impossible to determine which one was taken first (one participant).
3. Participants whose overall categorization performance was worse than or equal to chance (50% correct) (three participants).
4. Participants whose average reproduction error (Euclidean distance from the probe) exceeded three standard deviations from average of all the participants (one participant).
5. Individual reconstruction trials that correspond to moving the mouse one of four corners (i.e., with coordinates 0,0) (364 reconstruction trials which constitute 8% of the reconstruction data).
6. Individual reconstruction trials that are more likely to have been generated by random guesses. We first simulated random reconstruction data by sampling reconstruction coordinates from a uniform distribution and then computing their Euclidean distances from coordinates of a randomly chosen probe stimulus for every trial. We then compared the empirical distribution of reconstruction distances with a random

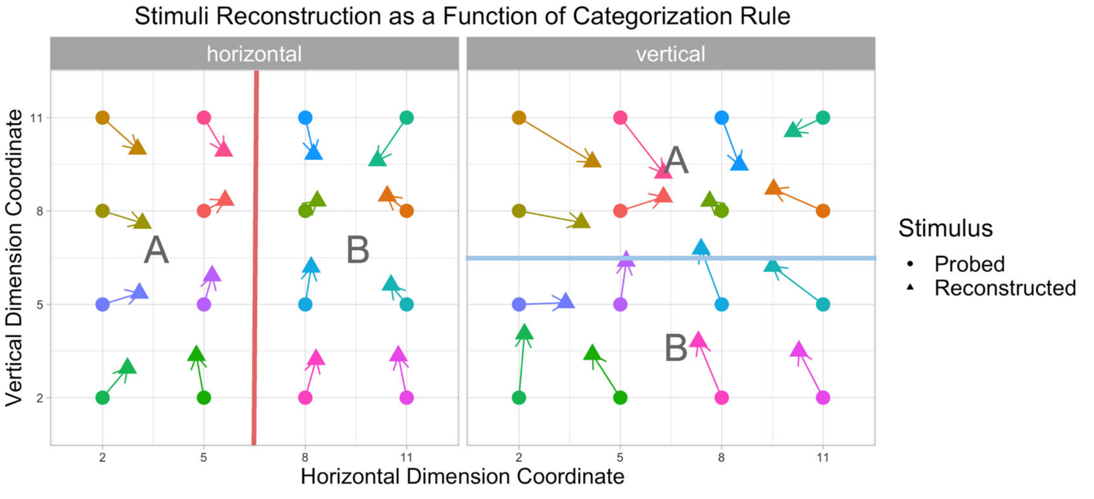

Fig. 5. The average reconstructions for different categorization rule conditions. Left: horizontal categorization rule; right: vertical categorization rule. The circles correspond to the probed example's positions on the 2D morph space. Triangles correspond to the average of per-participant mean reproduction coordinates. Notice that the reconstructed images tended to migrate to the global average values on both dimensions.

distribution and eliminated reconstruction trials with distances that have higher probability under the random guess distribution (see Appendix A) (1,077 trials which constitute 27% of the remaining reconstruction data after step 5). This criterion reflected the best way that we found to eliminate all the data potentially resulting from inattentive reconstruction by filtering out all the reconstruction trials that are more likely to have been generated by randomly guessing a coordinate on the reconstruction space than by reproducing the actual position with noise. We also report results of alternative analysis with milder criterion, where reconstruction trials with distances higher than two standard deviations from mean were eliminated (95 trials which constitute 2% of the remaining reconstruction data after step 5) (see Appendices A, B, and C). The dominant and the alternative criteria were determined prior to the analyses.

Two-way ANOVA and mixed linear regression analyses were performed to test the experimental hypotheses on the data from the reconstruction trials.

4.2. *Visualization and exploratory analysis*

We observed a strong overall assimilation effect toward the average values on both dimensions (Figs. 5 and 6).

The categorization performance of most of the participants was high (Fig. 7) and stayed relatively stable during the course of learning, suggesting that the task was quite easy. Moreover, participants in the horizontal categorization group were generally more accurate in reconstructing horizontal than vertical dimension (Fig. 7).

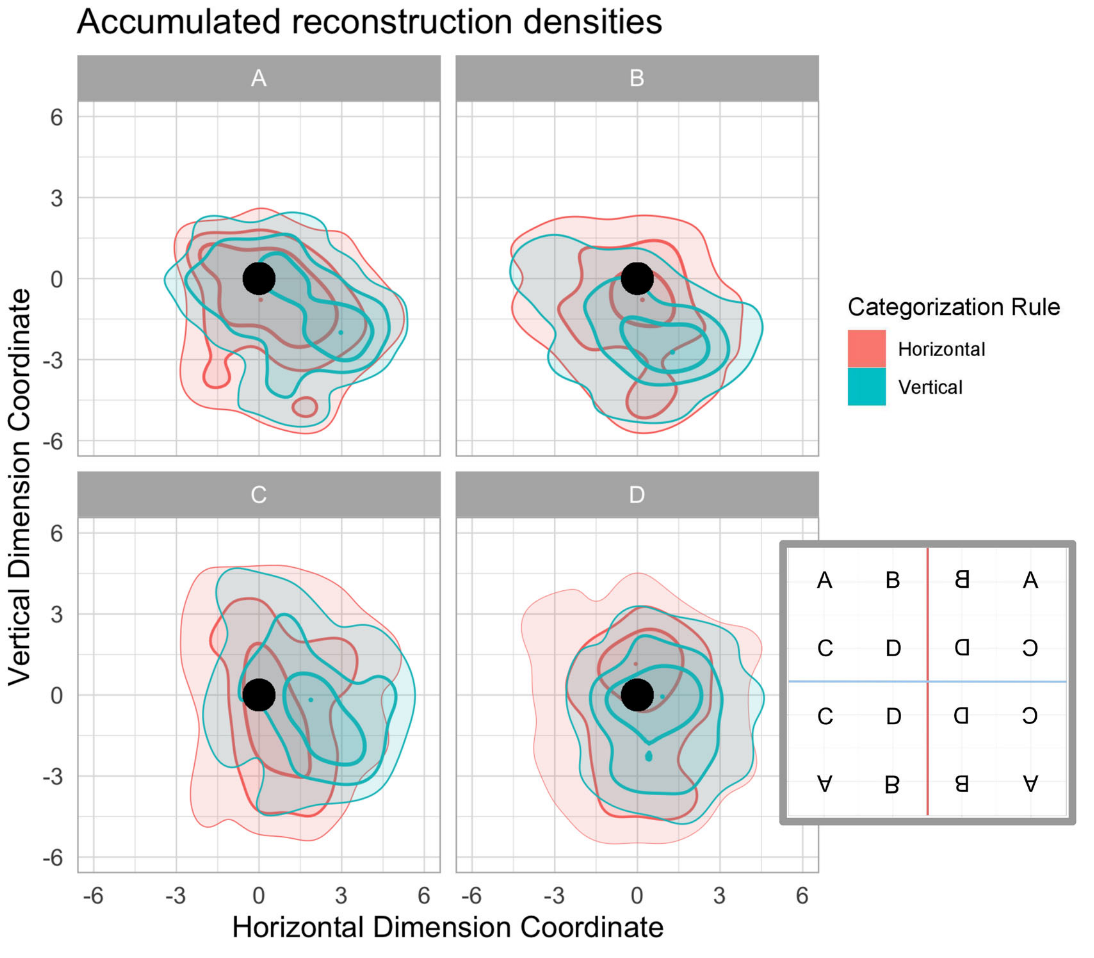

Fig. 6. Accumulated density plot of the reconstructed stimuli distances from probed stimuli for different categorization rule conditions. Sixteen probed stimuli are divided into four groups (A, B, C, and D) on the basis of their relative position with respect to the categorical boundaries (see bottom right corner). The data for these groups were combined: reconstruction deviations were projected to correspond to the same direction of movement relative to categorical boundaries (see bottom right corner). Notice that the horizontal categorization rule participants demonstrated greater variability in the vertical dimension reconstruction than the vertical categorization rule participants, and vice versa for the horizontal dimension. Moreover, the vertical categorization rule participants' reconstructions are less likely to be close to the vertical categorical boundary than the reconstructions of the horizontal categorization rule participants, and vice versa for the horizontal categorical boundary.

4.3. *Perceptual differentiation*

According to unsupervised differentiation theory, simple repeated exposure to the stimuli should improve their reconstruction accuracy. To test this, we computed Euclidean distances of the reconstructed images from the respective probed images on the morph space. Then, we applied a mixed linear model with the formula: $distance \sim trial\_number + (1|subject)$. The

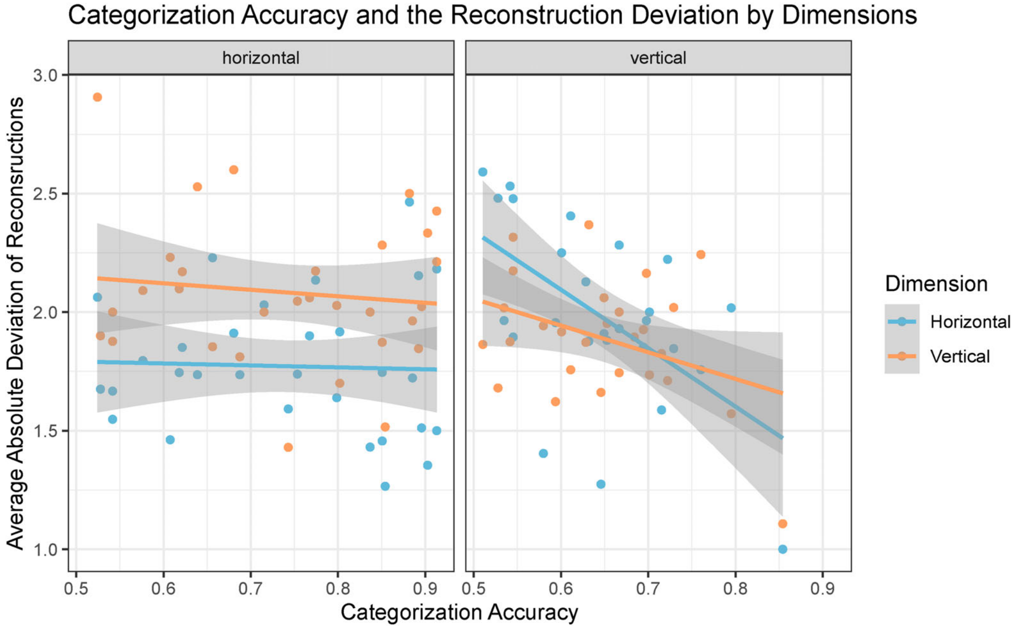

Fig. 7. The mean deviations of stimuli reconstructions and average categorization accuracies per participant for different categorization rule conditions. Left: horizontal categorization rule; right: vertical categorization rule.

effect of the trial number was insignificant [$t = –0.54$, $p = .59$, est = –0.00]. For the test on reconstruction variance, we divided all the trials into four successive blocks and computed standard deviations of reconstructions for trial blocks per each participant. Then, we applied mixed linear model with the formula: $reconstruction\_std \sim block + (1|subject)$. The effect of trial block was insignificant [$t = –0.99$, $p = .32$, est = –0.02]. These results replicated in the alternative analysis (Appendix B).

4.4. Reconstruction refinement on categorically relevant dimensions

Models of perceptual encoding/attention resources allocation predict the refinement of the reconstructions on categorically relevant dimensions. For this analysis, we computed the absolute distances between the x and y coordinates of the source images and the x and y coordinates of the reconstructed images, respectively. There was a significant interaction between the effects of categorization rule (horizontal or vertical) and the axis of computed deviation from the actual stimuli (horizontal or vertical) [two-way ANOVA: $F = 15.46$, $p<.001$, ges = 0.1]: Participants learning the horizontal categorization rule were more accurate in reconstructing the horizontal dimension than the vertical categorization rule participants, whereas participants assigned to the vertical categorization rule were more accurate in reconstructing the vertical dimension than the horizontal categorization rule participants. The vertical dimension (y-axis) was significantly harder to reconstruct in general [$F = 5.85$, $p<.05$, ges = 0.04]. The effect of categorization rule on the reconstruction deviations was

Fig. 8. Left: Reconstruction refinement on category-relevant dimensions. Middle: Reconstruction bias with respect to the categorical boundary (boundary aversion effect). Right: Assimilation toward category prototype.

insignificant [$F = 0.02$, $p = .88$, ges $= 0.0002$] (Fig. 8a). All the results replicated in the alternative analysis (Appendix B).

4.5. *Boundary aversion*

Perceptual warping predicts that perceptual reconstructions are less likely to move toward the categorical boundary and cross it, rather than move in the opposite direction. We computed the signed distances between the categorical boundary and the reconstructed images' coordinates. The signs of the deviations from boundary were reversed for the data points on one side of it so as to correspond to the same relative directions of movement (positive deviation = toward the boundary, negative = away from the boundary). All the groups demonstrated strong reconstruction bias toward the boundary, consistent with global assimilation. The horizontal categorization rule participants demonstrated smaller reconstruction bias toward the horizontal boundary than the vertical categorization rule participants, and vice versa for the vertical categorization rule participants (Fig. 8). This effect, however, was insignificant in our main analysis [$F = 2.58$, $p = .11$, ges $= 0.01$]. The effects of reconstruction dimension [$F = 1.42$, $p = .24$, ges $= 0.006$] and categorization rule were also insignificant [$F = 0.22$, $p = .64$, ges $= 0.003$] (Fig. 8b). In the alternative analysis, the effects of reconstruction dimension and interaction of reconstruction dimension and categorization rule were significant: Participants in both groups generally demonstrated higher global assimilation bias in the vertical dimension, and the assimilation bias for each dimension was smaller when it was diagnostic for categorization (Appendix B).

4.6. *Perceptual compression (assimilation) to the prototype*

Assimilation of reconstructions toward the category prototype is a prediction of the Bayesian models of categorical perception. We looked at the absolute deviations of stimuli reconstructions from the categorical means in both dimensions. The distances from both vertical and horizontal categorical means were computed for each participant (with no regard to their actual categorization rules) to compare two categorization groups. We performed a two-way ANOVA to test the effects of categorization rule (horizontal or vertical), and the prototype dimension (horizontal or vertical) on the absolute reconstruction distances. All three effects were insignificant in our main analysis [reconstruction dimension: $F = 0.0004$, $p = .98$, ges $= 0.00$; categorization rule: $F = 0.05$, $p = .82$, ges $= 0.0004$; categorization rule and prototype dimension interaction: $F = 3.39$, $p = .07$, ges $= 0.03$] (Fig. 8c). In the alternative analysis, the effects of reconstruction dimension and interaction of reconstruction dimension and categorization rule were significant: Participants in both groups demonstrated stronger assimilation to the prototype in horizontal dimension. More importantly, the horizontal categorization rule participants demonstrated stronger assimilation to the prototype in horizontal dimension than the vertical categorization rule participants, and vice versa for the vertical categorization rule group (Appendix B).

5. Discussion

Our study demonstrates that a variety of potential perceptual learning mechanisms could play a role in a single experiment measuring reconstruction error and systematic bias. However, the categorical perception literature and models often neglect other unsupervised and supervised effects that can affect the results and their interpretation. We suggest that perceptual learning mechanisms cannot be studied in isolation and argue for a systems approach to understanding perceptual learning.

In this paper, we propose an experimental method that can provide rich data on how different perceptual learning mechanisms work and interact. We reviewed computational unsupervised and supervised (categorical) perceptual learning models and articulated their predictions for our task. We tested these effects using an interleaved categorization-reconstruction experimental design with immediate reconstruction trials.

First, we observed a strong unsupervised assimilation effect. Specifically, participants consistently biased their reconstruction responses toward the global average values on both dimensions. This effect is compatible with many Bayesian perception and rate distortion models that formalize perception as an optimal stimulus reconstruction process. It is not clear, however, if the global assimilation bias found in our experiment is perceptual: Our experimental design does not completely eliminate potential decision making or mouse movement strategies that could lead to this effect.

The interpretation of some of the supervised effects that we found evidence for is weakened by the potential nonperceptual explanations as well. In particular, our experimental design is susceptible to the El Greco methodological fallacy (Firestone & Scholl, 2014; Firestone & Scholl, 2016), named after a case from art history. In the original story, art experts suggested that the elongated human figures in paintings by El Greco resulted from his astigmatic "stretched" perception of the environment. However, the logic of this argument is inconsistent. The astigmatism should have also stretched the perception of the canvas itself, leading to these perceptual effects cancelling each other out, so the painted bodies would look normal. The same logic applied to categorical perception studies challenges experiments in which the examples of the *same* category are compared or matched to reveal the potential categorical effects (e.g., Goldstone, 1995). Given the hypothesized perceptual nature of these effects, the items from the same category should be distorted in the same way, so the effects should cancel each other out and lead to the undistorted matchings in such experimental designs. In the reconstruction trials of our experiment, the probed and the accurately matched stimuli belong to the same category. Therefore, the categorical biases should affect the perception of both of them in the same direction. In such a case, Firestone and Scholl argued that the experiment is unable to distinguish between situations with and without perceptual changes and the reported distortions cannot be explained in terms of perceptual processes. We address this criticism in the discussion of particular categorical effects that we found evidence for.

The overall reconstruction bias and the variance did not decrease over learning blocks as suggested by the perceptual differentiation theory. However, participants became more accurate in reconstructing the category-relevant dimension. This suggests that the allocation of perceptual encoding or attention resources is affected by the task structure, with categorization being an example of such a task. To our knowledge, this is the first time the selective refinement of category-relevant dimensions on an immediate (rather than delayed) reconstruction has been demonstrated. Moreover, this effect excludes potential "El Greco-effect" explanations—this encoding effect is not related to *biases* in perceptual representations but rather only relates to *how refined* these representations are.

Participants' reconstructions were less biased toward the global mean on the category-relevant dimension (boundary aversion effect). Perceptual warping, Bayesian, and rate distortion models of categorical perception are consistent with this result. We also found evidence for assimilation toward the category prototypes (along the categorization-diagnostic dimensions) as suggested by the Bayesian perception and rate distortion models. These two effects, however, were supported by one of our two alternative analyses. We explain it by lower statistical power of the dominant analysis, caused by exclusion of more data; in fact, patterns in the data determined by two alternative exclusion criteria were very similar (see Fig. 8 and Appendix B). Moreover, we cannot eliminate the alternative "El Greco" explanation of these two effects which remains to be addressed in future studies. However, given that the alternative, nonperceptual, interpretations are minimized in the immediate perceptual reconstruction task, we suggest that a potential nonperceptual interpretation for these effects is clarified.

We did not have sufficient data to reveal the trajectories of perceptual change during category learning. Most of the participants achieved good performance during the first 25% of their categorization trials and maintained it until the end of the experiment. We explain the high categorization accuracy by the relative simplicity of the categorization rules, despite their perceptual subtlety, and the interpretability of stimuli dimensions.

The supervised effects found in our experiment suggest that perception is not an independent encoding and preprocessing of sensory data—it serves particular higher-level cognitive demands such as categorization, and it is adapted with respect to them. First, visual perception allocates its limited encoding resources with respect to categorization utility. Second, it biases the stimuli representations away from the category boundary, preventing miscategorization of the stimulus as a result. This supervised, categorization-influenced effect operates additively with the strong unsupervised assimilation toward the global average stimulus. As a result, the reproduced stimuli are not overall distorted away from category boundaries, but are just less distorted toward them. Finally, categorical labels support robust perceptual reconstruction by providing additional information on what perceptual stimulations are more likely, leading to the reconstruction bias toward categorical prototypes.

6. Conclusion

Visual perception is highly adaptive both to the structure of the environment and to higher-level cognitive demands. In particular, vision adapts its attentional or encoding resources with respect to their categorization utility, making the representations of category-relevant dimensions more accurate. It adaptively uses categorical information to improve reconstruction under uncertainty and with limited representational resources, biasing perceptual representations toward category prototypes. Moreover, the human visual system also biases its perceptual encodings away from the category boundary to prevent miscategorizations, partially compensating for the countervailing unsupervised assimilation to the global mean of the entire set of stimuli. We suggest that perceptual learning mechanisms can be effectively studied as a system, using experimental tasks that provide rich data on absolute and directional perceptual changes.

Acknowledgments

The authors thank Willa Mannering for preparing the stimuli for the experiment, Reina Munoz for assisting with data collection, Arseny Moskvichev, Thomas Gorman, Phillip Wolff, Patricia Bauer, and Robert Nosofsky for helpful suggestions on this study and the analysis, and Roman Tikhonov for his useful comments on the earlier version of this manuscript. The authors also thank two anonymous reviewers for their helpful feedback on the manuscript.

Open Research Badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/2xr54/.

References

Barlow, H. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research*, *30*(11), 1561–1571.

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review*, *127*(5), 891–917. https://doi.org/10.1037/rev0000197

Bates, C. J., Lerch, R. A., Sims, C. R., & Jacobs, R. A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of Vision*, *19*(2), 1–23.

Bates, C., & Jacobs, R. (2019). Efficient data compression leads to categorical bias in perception and perceptual memory. *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 1369–1375.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217–239. https://doi.org/10.1016/0010-0277(95)00669-X

Busemeyer, J. R., & Myung, I. J. (1988). A new method for investigating prototype learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 3–11. https://doi.org/10.1037/0278-7393.14.1.3

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. https://doi.org/10.1371/journal.pone.0158725

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

de Beeck, H. O., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: Dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General*, *132*(4), 491–511. https://doi.org/10.1037/0096-3445.132.4.491

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), 9. Retrieved from http://doi.org/10.1525/collabra.32

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, E. L. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review*, *17*(2), 224–230. https://doi.org/10.3758/PBR.17.2.224

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227–240. https://doi.org/10.1016/0010-0277(92)90002-Y

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

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*, e229. https://doi.org/10.1017/S0140525X15000965

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*(4), 814–823. https://doi.org/10.1093/cercor/bhs067

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221.

Geisler, W. S. (2003). Ideal observer analysis. *Visual Neurosciences*, *10*(7), 825–837.

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, *27*(3), 379–402.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice Hall.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32–41. https://doi.org/10.1037/h0048826

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200. https://doi.org/10.1037/0096-3445.123.2.178

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. (2003). Learning to perceive while perceiving to learn. In R. Kimchi, M. Behrmann & C. Olson. *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 233–280). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116–139. https://doi.org/10.1037/0096-3445.130.1.116

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

Harnad, S., Hanson, S. J., & Lubin, J. (1991). *Categorical perception and the evolution of supervised learning in neural nets*. Retrieved from https://eprints.soton.ac.uk/253378/

Harnad, S., Hanson, S. J., & Lubin, J. (1995). Learned categorical perception in neural nets: Implications for symbol grounding. In V. Honavar & L. Uhr (Eds.), *Symbol processors and connectionist network models in artificial intelligence and cognitive modelling: Steps toward principled integration* (pp. 191–206).

Hockema, S. A., Blair, M. R., & Goldstone, R. L. (2005). Differentiation for novel dimensions. In B. Bara, L. Barsalou & M. Bucciarelli. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 953–958). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Knill, D. C., Kersten, D., & Yuille, A. (1996). Introduction: A Bayesian formulation of visual perception. *Perception as Bayesian Inference*, *1*, 1–21.

Körding, K. (2007). Decision theory: What "should" the nervous system do? *Science*, *318*(5850), 606–610.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

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

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*(6), 812–822. https://doi.org/10.1016/0959-4388(94)90128-7

Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, *39*, 770–784.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics*, *62*(2), 386–401. https://doi.org/10.3758/BF03205558

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732–753. https://doi.org/10.1037/0278-7393.24.3.732

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, *6*(4), 547–569.

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.

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience*, *35*(21), 8145–8157.

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition*, *95*(2), B1–B14. https://doi.org/10.1016/j.cognition.2004.07.002
Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition*, *35*(7), 1814–1829. https://doi.org/10.3758/BF03193512
Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001
Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*(3), 369–398. https://doi.org/10.1037/0096-3445.129.3.369
Sims, C. R. (2016). Rate–distortion theory and human perception. *Cognition*, *152*, 181–198. https://doi.org/10.1016/j.cognition.2016.03.020
Sims, C. R. (2018). Efficient coding explains the universal law of generalization in human perception. *Science*, *360*(6389), 652–656.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

APPENDIX A: Exclusion criteria

Here, we show where both exclusion criteria belong with respect to empirical distribution of reconstruction deviations and simulated distribution of random guesses (Fig. A1). Note that exclusion criterion which eliminates reconstruction trials that are more than two standard deviations away from the average is much milder than the exclusion criterion eliminating all the trials that are more likely to have been generated by randomly guessing a coordinate on the reconstruction space than by reproducing the actual position with noise.

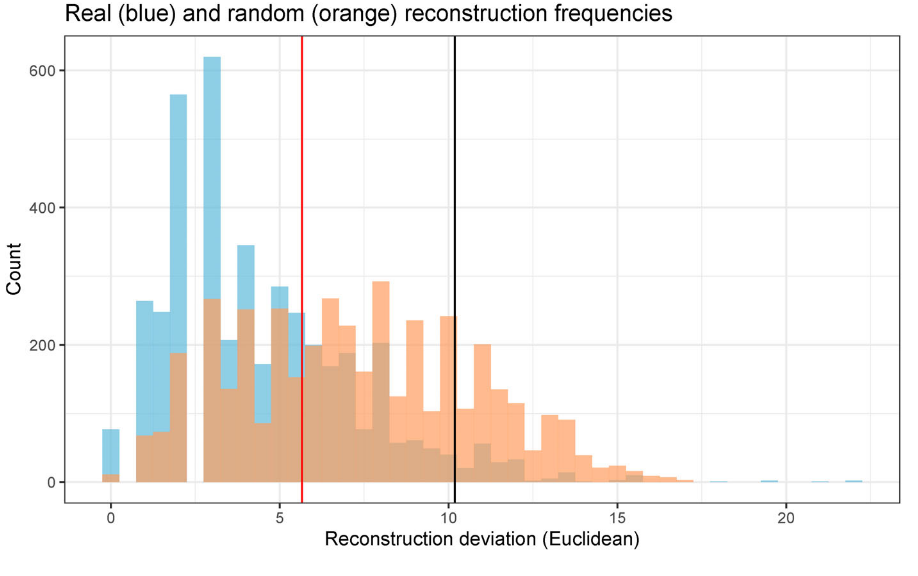

Fig. A1. Exclusion criteria based on the comparison with distribution of random guesses (red) and two standard deviations from the mean reconstruction (black). Histograms of the empirical (blue) and simulated (orange) reconstruction data.

APPENDIX B: Results of alternative analysis

Here, we report detailed results of the alternative analysis performed after the milder exclusion criterion (Fig. B1) was applied to the data.

1.1. Perceptual differentiation

We applied a mixed linear model with the formula: $distance \sim trial\_number + (1|subject)$. The effect of the trial number was insignificant [$t = 0.33$, $p = .74$, est = 0.0001]. For the test on reconstruction variance, we divided all the trials into four successive blocks and computed standard deviations of reconstructions for trial blocks per each participant. Then, we applied mixed linear model with formula: $reconstruction\_std \sim block + (1|subject)$. The effect of trial block was insignificant [$t = 0.02$, $p = .99$, est = 0.0004].

1.2. Reconstruction refinement on categorically relevant dimensions

We found a significant interaction between the effects of categorization rule (horizontal or vertical) and the dimension of computed deviation from the actual stimuli (horizontal or vertical) [$F = 6.22$, $p<.05$, ges = 0.03]: Participants learning the horizontal categorization rule were more accurate in reconstructing the horizontal dimension than the vertical categorization rule participants, whereas participants assigned to the vertical categorization rule were more accurate in reconstructing the vertical dimension than the horizontal categorization rule participants. Vertical dimension (y-axis) was significantly harder to reconstruct in general [$F = 26.9$, $p<.001$, ges = 0.11]. The effect of categorization rule on the reconstruction deviations was insignificant [$F = 0.002$, $p = .97$, ges = 0.00].

1.3. Boundary aversion

Horizontal categorization rule group demonstrated significantly smaller reconstruction bias toward the horizontal boundary than the vertical categorization rule group, and vice versa for the vertical categorization rule group [$F = 5.69$, $p<.05$, ges = 0.03]. Participants in both groups demonstrated significantly higher global assimilation effect in vertical than the

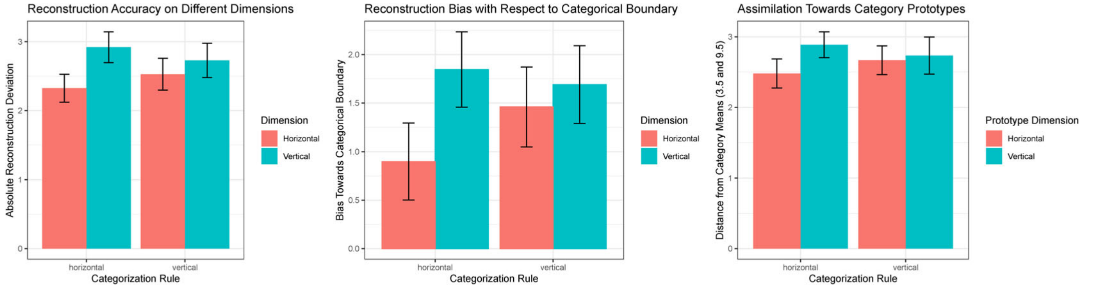

Fig. B1. Hypotheses testing performed on the data with alternative exclusion criteria. Left: reconstruction refinement on category-relevant dimensions. Middle: reconstruction bias with respect to the categorical boundary (boundary aversion effect). Right: assimilation toward category prototype.

horizontal dimension [$F = 16.4$, $p < .001$, ges $= 0.08$]. The effect of categorization rule was insignificant [$F = 0.77$, $p = .38$, ges $= 0.01$].

1.4. Perceptual compression (assimilation) to the prototype

Participants demonstrated higher assimilation to the prototype in a dimension when it was diagnostic for categorization [$F = 4.1$, $p < .05$, ges $= 0.02$]. Participants in both groups showed significantly higher assimilation to the prototype in horizontal than vertical dimension [$F = 8.56$, $p < .01$, ges $= 0.05$]. The effect of categorization rule was insignificant [$F = 0.02$, $p = .89$, ges $= 0.0003$].

APPENDIX C: Original analysis

We performed one data analysis prior to the manuscript submission. For the original analysis, we excluded all the reconstruction trials with Manhattan distance higher than 6 and used Manhattan distance when it was applicable. Other procedures are identical in the original and new analyses. Due to the reviewers' objection to the exclusion criterion and distance metric that we used, we decided to reanalyze the data using new exclusion criteria and Euclidean distance. Here, we report our original results.

1.1. Perceptual differentiation

We first applied a mixed linear model with the formula: *distance* $\sim$ *trial_number* + *(1|subject)*. The effect of the trial number was insignificant [$t = -0.76$, $p = .45$, est $= -0.0002$]. For the test on reconstruction variance, we divided all the trials into four successive blocks, and computed standard deviations of reconstructions for trial blocks per each participant. Then, we applied mixed linear model with formula: reconstruction_std $\sim$ block + (1|subject). The effect of trial block was insignificant [$t = -0.237$, $p = .81$, est $= -0.004$].

1.2. Reconstruction refinement on categorically relevant dimensions

We computed the absolute distances between the x and y coordinates of the source images and the x and y coordinates of the reconstructed images, respectively. There was a significant

Fig. C1. Hypotheses testing performed on the data with original exclusion criteria. Left: reconstruction refinement on category-relevant dimensions. Middle: reconstruction bias with respect to the categorical boundary (boundary aversion effect). Right: assimilation toward category prototype.

interaction between the effects of categorization rule (horizontal or vertical) and the axis of computed deviation from the actual stimuli (horizontal or vertical) [two-way ANOVA: $F = 14.03$, $p < .001$, ges = 0.11]. The vertical dimension (y-axis) was significantly harder to reconstruct in general [$F = 6.52$, $p < .05$, ges = 0.06]. The effect of categorization rule on the reconstruction deviations was insignificant [$F = 0.1$, $p = .75$, ges = 0.0009].

1.3. Boundary aversion

All the groups demonstrated strong reconstruction bias toward the boundary, consistent with global assimilation. There was a significant interaction between the effects of participants' categorization rule (horizontal or vertical) and the dimension of deviation (horizontal or vertical) in the predicted direction [$F = 4.64$, $p < .05$, ges = 0.02]. In particular, the horizontal categorization rule participants demonstrated smaller reconstruction bias toward the horizontal boundary than the vertical categorization rule participants, and vice versa for the vertical categorization rule participants. The effects of the reconstruction dimension and categorization rule were insignificant [reconstruction dimension: $F = 1.9$, $p = .17$, ges = 0.01; categorization rule: $F = 0.53$, $p = .47$, ges = 0.007].

1.4. Perceptual compression (assimilation) to the prototype

All three effects were insignificant in our main analysis [reconstruction dimension: $F = 0.03$, $p = .86$, ges = 0.0003; categorization rule: $F = 0.23$, $p = .63$, ges = 0.002; categorization rule and prototype dimension interaction: $F = 3.86$, $p = .05$, ges = 0.03].