



## OPEN ACCESS

## EDITED BY

Kasia M. Bieszczađ,  
Rutgers, The State University of New Jersey,  
United States

## REVIEWED BY

Jose Pablo Ossandon,  
University of Hamburg, Germany  
Joseph Charles Schmidt,  
University of Central Florida, United States

## \*CORRESPONDENCE

Dagmar Zeithamova  
✉ [dasa@uoregon.edu](mailto:dasa@uoregon.edu)

RECEIVED 31 July 2023

ACCEPTED 18 October 2023

PUBLISHED 15 November 2023

## CITATION

Ashby SR, Chaloupka B and Zeithamova D  
(2023) Category bias in similarity ratings: the  
influence of perceptual and strategic biases in  
similarity judgments of faces.  
*Front. Cognit.* 2:1270519.  
doi: 10.3389/fcogn.2023.1270519

## COPYRIGHT

© 2023 Ashby, Chaloupka and Zeithamova.  
This is an open-access article distributed under  
the terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Category bias in similarity ratings: the influence of perceptual and strategic biases in similarity judgments of faces

Stefania R. Ashby<sup>1,2</sup>, Benjamin Chaloupka<sup>3</sup> and  
Dagmar Zeithamova<sup>3\*</sup>

<sup>1</sup>Department of Psychology, Brigham Young University, Provo, UT, United States, <sup>2</sup>Neuroscience Center, Brigham Young University, Provo, UT, United States, <sup>3</sup>Department of Psychology, University of Oregon, Eugene, OR, United States

**Introduction:** After category learning, same-category items tend to be rated as more similar than items from different categories. Whether this category bias in similarity ratings reflects true changes in perception or a strategic judgment bias to rate same-category items more similarly has been debated. The current study investigated the influence of perceptual and strategic judgment biases on perceived similarity ratings of face stimuli.

**Method:** To explore the influence of perceptual and strategic biases, post-learning category bias was measured after learning one of two category structures. In a similarity-consistent structure, faces within a category shared physical features and category bias could reflect a combination of strategic bias and true perceptual changes. In a similarity-inconsistent structure, category membership was orthogonal to physical features and category bias could only be driven by strategic bias to rate same-label faces as more similar.

**Results:** We found a strong category bias after learning, but only when category labels could be aligned to the similarity structure. When category label conflicted with similarity structure, the mere presence of a shared label did not create a bias.

**Discussion:** These findings indicate that category bias in this paradigm is largely driven by a perceptual bias, consistent with proposals that category learning can stretch or shrink perceptual space by biasing attention toward category-relevant and away from category-irrelevant features. More broadly, these findings contribute to our understanding of category-driven biases and may inform bias research in other domains such as social stereotypes.

## KEYWORDS

category learning, perceptual bias, strategic judgment bias, similarity-based categories, arbitrary categories

## Introduction

We encounter a massive amount of information daily. Categorization is one of the ways we organize this plethora of data for future use. Category knowledge allows us to organize that information into meaningful bundles and generalization of category knowledge guides our decisions in novel situations (Markman and Ross, 2003; Ashby and Maddox, 2005, 2011; Seger and Miller, 2010; Goldstone et al., 2018). Category knowledge also affects perception: items within a category are viewed as more similar while items from different categories are viewed as less similar (see Livingston et al., 1998; Goldstone and Steyvers, 2001). This category bias in perception has been documented for well-established categories such as discriminating hues across color boundaries (Bornstein et al., 1976; Sandell et al., 1979; Bornstein and Korda, 1984; Franklin and Davies, 2004; Winawer et al., 2007) and

discriminating speech sounds across phonemic boundaries (Lieberman et al., 1957). Even categories like race induce a category bias in perception (Levin and Angelone, 2002) which may in part explain other phenomena like own-race bias where individuals are better at differentiating people of their same race (Timeo et al., 2017). Additionally, category biases in perception can emerge after individuals learn new categories in the laboratory (Rosch and Mervis, 1975; Goldstone, 1994a,b; Beale and Keil, 1995; Goldstone et al., 2001), and can be induced by the presence of a shared category label even when task instructions emphasize memory for individual items (Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022).

In lab studies examining the impact of newly learned categories, the presence of a category bias in perception has commonly been studied using perceived similarity ratings (Goldstone, 1994a; Beale and Keil, 1995; Kurtz, 1996; Livingston et al., 1998). Typically, participants are shown a set of stimuli and asked to rate how alike all stimuli are to one another. These ratings are taken prior to category learning and then again after participants have learned category information. Evidence for category bias in these similarity ratings is reflected by greater perceived similarity ratings after learning for items that belong to the same category (Goldstone, 1994a; Levin and Beale, 2000; Ashby et al., 2020), and/or lower similarity ratings for items that belong to different categories (Kurtz, 1996; Livingston et al., 1998; Ashby et al., 2020). Our prior work has examined category bias in perceived similarity ratings after participants learned a family category structure of face-blend stimuli varying along many continuous dimensions. We have shown that a category bias in perceived similarity ratings of faces can arise in brain and behavior from the mere presence of a shared family label even when individuals are not explicitly aware of an underlying family category structure and task goals focus on memorizing individual faces rather than category learning (Ashby et al., 2020; Ashby and Zeithamova, 2022). Furthermore, the categorization bias predicts how well individuals can generalize category information to new faces in both young adults (Ashby et al., 2020; Ashby and Zeithamova, 2022) and older adults (Bowman et al., 2022), suggesting that it tracks category knowledge.

What drives changes in perceived similarity ratings following category learning? The mechanism responsible for category bias has been debated. Some have postulated that category learning affects how we perceive the world by warping perceptual space. As attention is biased toward category-relevant features and away from category-irrelevant features, the perceptual space stretches along the relevant dimensions and shrinks along the irrelevant dimensions, making within-category similarity and between-category differences more apparent (Nosofsky, 1984, 1986; Kruschke, 1992, 1996; Lupyan, 2012). Thus, under this *perceptual bias* account, category learning alters representations of individual items to align with category membership. In support of this account, past behavioral work has indicated that information is more discriminable across category-relevant dimensions after category learning (Goldstone, 1994b; Goldstone and Steyvers, 2001; Gureckis and Goldstone, 2008). Further, increased discriminability of items on a category-relevant dimension has been shown to extend to more complex category structures (e.g., object shape and object motion dimensions; Folstein et al., 2014) and result in larger

increases in within-category similarity after learning compared to more simple category structures (Pothos and Reppa, 2014).

Neuroimaging work has provided additional support. For example, Folstein et al. (2013b) indicated that better ability to discriminate among category-relevant features after learning was associated with increased discriminability of their neural representations in visual cortex. Likewise, Juárez et al. (2019) found that an increase in the N1 neural response—typically associated with early perceptual processing in category learning (for example see Curran et al., 2002)—predicted a decrease in similarity ratings for items that belong to different categories after learning. And recent work from our lab found evidence for category-biased neural representations widespread across the cortex that emerged during learning, even when physical similarity was controlled for and category information was not explicitly emphasized (Ashby and Zeithamova, 2022). Taken together, we have proposed that changes in perceived similarity ratings after category learning in our prior work reflect true changes in perception as attention is shifted toward category-relevant and away from category-irrelevant features to better align perceived similarity with category membership.

However, category biases may not just reflect changes in perceptual space. For example, Goldstone and colleagues (2001) suggest an alternative to the perceptual change hypothesis asserting that biases in perceived similarity ratings may also reflect *strategic judgment bias* to take the category label into account when rating similarity. In other words, one may provide higher ratings of items from the same category and lower ratings of items from different categories simply because their same vs. distinct category labels are considered, rather than because they appear more or less similar. Indeed, a contribution of the strategic judgment bias to perceived similarity ratings was concluded in a study in adult category learners by Goldstone and colleagues (2001) which included new items not previously associated with a category label during the similarity rating phase. This also aligns with other work showing that introducing verbal interference disrupts learning of shared linguistic labels which abolishes categorical perception effects (Roberson and Davidoff, 2000; Pilling et al., 2003; Hanley and Roberson, 2011; Simanova et al., 2016) and developmental work showing an effect of shared linguistic labels on similarity judgments in children (Sloutsky and Lo, 1999; Booth, 2014). Thus, differentiating between a perceptual bias and a strategic judgment bias account is important for a mechanistic understanding of the role of concepts in cognition and perception.

Having found strong category biases in brain and behavior in our prior work using a face-blend category learning task (see Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022), the current study sought to determine the degree to which perceptual and strategic judgment bias may contribute to the similarity ratings measures in this paradigm. We reasoned that when faces belonging to the same category share features (as they typically do), category bias in perceived similarity ratings may be fully or partially driven by true changes in perception, such as through attentional bias toward category-relevant features. On the other hand, when category labels are arbitrary and faces within categories do not share features, one cannot alter within-category similarity or between-category differences through learned attentional biases.

As such, any change in perceived similarity ratings would be driven by a strategic judgment bias to rate same category faces similarly. By comparing similarity rating bias observed under these two conditions, we can determine the degree to which strategic judgment bias contributes to category bias in post-learning perceived similarity ratings.

Using a traditional feedback-based category learning task and face-blend stimuli developed in our prior work to control physical similarity of naturalistic stimuli (Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022), we tracked post-learning category bias under two category conditions. Some participants learned categories where stimuli within a category shared features while others learned categories where stimuli within a category did not share features. We reasoned that if category bias in similarity ratings is more reflective of a strategic judgment bias, then category bias in similarity ratings would be expected after learning in both category conditions, even when participants learn category labels that are orthogonal to the physical similarity structure of the stimuli. Alternatively, if a category bias in ratings is more reflective of changes in perception or attention to category-relevant features, then a category bias in perceived similarity ratings would only be expected following learning of categories based on shared features. And finally, if category bias in similarity ratings is a combination of true changes in perception and a strategic judgment bias, comparing the degree of category bias between conditions will allow us to estimate the relative contribution of each process to the total category bias.

## Methods

### Data transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The design and analysis plan for this experiment were preregistered via the Open Science Framework (OSF; <https://osf.io/9serh>) and all data and materials are also publicly available via OSF (<https://osf.io/bhc32/>).

### Participants

Two-hundred and 15 participants were recruited from the University of Oregon via the university SONA research system and completed the study online via Pavlovia ([www.pavlovia.org](http://www.pavlovia.org)) in 2020. Participants received course credit as compensation for their participation. All participants provided informed consent and experimental procedures were approved by Research Compliance Services at the University of Oregon. Six participants were excluded from analyses for failing to respond to 20% or more of the trials in either the similarity ratings or category learning phases. The remaining sample of 209 participants are reported in all analyses. Due to experimenter error in pivoting to online data collection amidst the COVID-19 pandemic, demographic information was only collected for approximately half the sample (93 female, 19 male, 3 unreported; age 18–34 years;  $M_{\text{age}} = 19.46$ ,  $SD_{\text{age}} = 2.05$ ). Since all participants were recruited through

introductory psychology courses at the University of Oregon, we are confident that the limited demographic information collected is representative of the overall sample.

### Rationale for sample size

The present study compared the category bias effect sizes between two groups of subjects, one that learned similarity-based categories (similarity-consistent condition) and one that learned categories that lacked a similarity structure (similarity-inconsistent condition). Sample sizes of at least 100 participants per condition were determined *a priori* via power analyses in G\*Power (Faul et al., 2007, 2009) utilizing effect sizes from previous studies (see Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022). Prior data demonstrated that when categories were similarity-consistent, the effect of a category bias in similarity ratings post-learning was medium-sized (Cohen's  $d = 0.51$ ). We reasoned that this effect size may be reflective of a combination of perceptual change and strategic processes in the similarity-consistent condition, with at least one of these effects contributing at least half ( $d = 0.25$ ) of the total effect size. In contrast, no true perceptual changes would be expected in the similarity-inconsistent condition, and any category bias in similarity ratings would be due solely to strategic processes (with effect size expected to be much smaller than  $d = 0.5$ ). Thus, our target was 80% power to detect a small effect ( $d = 0.25$ , or half of the total  $d = 0.5$  expected in the similarity-consistent condition based on our prior data) at the standard alpha error probability level ( $\alpha = 0.05$ ). This target is sufficient for both detecting differences between conditions and detecting a presence of category bias in the similarity-inconsistent condition (if one exists).

### Stimuli

Stimuli were grayscale images of blended faces created by morphing pairs of “parent” faces (FantaMorph Version 5 by Abrosoft) originating from the Dallas Face Database (Minear and Park, 2004), the Computer Vision Laboratory Database (Peer, 1999), and a Google Image search (Figure 1). Ten parent faces were selected for blending from a larger pool of faces used in prior work (see also Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022). We found in previous data that face blends of very distinctive parent faces were dominated by features of those parents, while features of more average parent faces tended to be lost in the resultant face-blend. Thus, we selected 10 faces that were neither too distinctive nor too average, helping us ensure that all face blends were about equally physically similar to both parent faces rather than randomly biased toward one parent or the other. From the set of 10 parent faces, six were randomly selected for each participant and blending was conducted by morphing three of these parent faces (“A” Parents) with the remaining three parent faces (“B” Parents) individually. This resulted in a set of nine face-blends that shared 50% features from each parent to be used in category learning and similarity ratings tasks (see Figure 1). To measure changes related to category learning and not mere differences in exposure frequency to different face features, it was

important that the training set was derived using an equal number of “A” and “B” parent faces. The 3x3 stimulus structure ensured that the task was not trivially easy (as would be the case with 2x2 structure) while resulting in a manageable number of pair-wise comparisons among all stimuli for the subsequent similarity rating task. To minimize any potential item effects in similarity ratings, participants were yoked within and across conditions, such that for each participant in the similarity-consistent condition with a given set of “A” and “B” parents, there was a participant in the similarity-inconsistent condition with the same parent-face selection, as well as one participant in each condition using the same parent faces but with reversed “A” and “B” parent assignment. The original, non-blended parent faces were never shown to participants nor used during any stage of the experiment.

Using blended faces allowed us to generate more naturalistic, realistic-looking stimuli that vary along many ill-defined features, akin to what we frequently encounter in everyday life, while still controlling their physical similarity. Our prior work demonstrated that participants are highly sensitive to the similarity structure introduced by face blending, such that two faces that share a parent (and thus exhibit that parent’s facial features) are perceived as much more similar than two faces that do not share a parent (e.g., Cohen’s  $d = 1.50$  in pre-learning similarity ratings in Ashby et al., 2020). Furthermore, we used the same stimuli and category structure as our prior work in order to better discern among competing interpretations of our previous work in light of findings from the current study.

Because we were interested in biases driven by the category learning task, it was important that the blended faces were not easily sorted into pre-existing categories prior to training on face families. Because race and gender are their own preexisting categories, all parent faces had to be of a single race and single gender. We chose white faces for the stimulus set because prior work has shown that white Americans exhibit greater own-race effects in face recognition than non-white Americans (Meissner and Brigham, 2001; Herzmann et al., 2017), and we expected our sample to be majority white based on the University of Oregon population demographics (64% white). We chose male for the face gender because females show less own-gender bias than males (Mishra et al., 2019) and because female faces have a greater variability in hair styles that can make the blending process difficult. Exploratory analyses of data from our prior work (see Ashby et al., 2020) indicate that category learning and the resulting category bias in this paradigm do not differ between male and female participants.

## Experimental conditions

Participants were randomly assigned into one of two experimental conditions where they learned to sort the nine face-blends into three categories labeled by family name (Miller, Wilson, Davis). The experimental conditions participants were assigned to determined how the face-blends were assigned into family categories either based on a shared parent (similarity-consistent) or a lack of a shared parent (similarity-inconsistent).

### Similarity-consistent family structure

For similarity-consistent categories, family assignment was aligned with the physical similarity among faces (Figure 1A). Specifically, three of the parent faces (“A” Parents) determined the three family categories (Miller, Wilson, Davis). Any face-blend of that parent face was considered a member of the same category. Because they were blended from the same parent, face-blends from a given category not only shared a family name but also shared physical features (see faces across rows in Figure 1A). Notably, faces that shared a “B” parent also shared physical features but belonged to different families (see faces down columns in Figure 1A). Thus, physical similarity alone did not determine category membership and participants needed to learn which face features were relevant to category membership through training. Category bias was measured by comparing similarity ratings for pairs of faces that shared a parent and belonged to the same family category with similarity ratings for pairs of faces that also shared a parent but belonged to different categories (rows vs. columns in Figure 1A). A tendency to rate faces that share category-relevant “A” parents as more similar than those that share category-irrelevant “B” parents would constitute a category bias in perceived similarity rating. We reasoned that this category bias would be a combination of a true perceptual bias and a strategic judgment bias (total bias = perceptual bias + strategic judgment bias).

### Similarity-inconsistent family structure

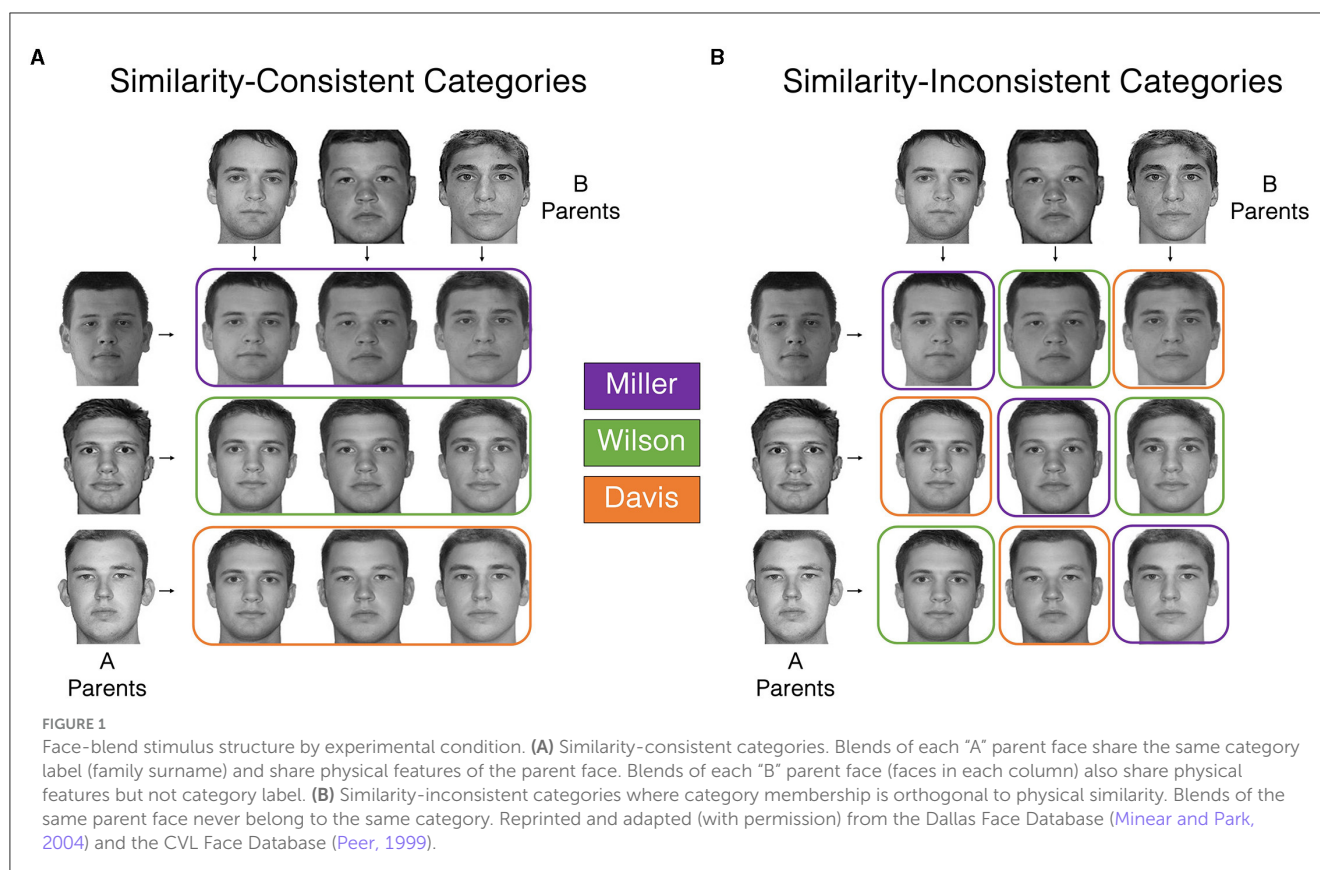
For similarity-inconsistent categories, the same blended faces were used but family assignment was orthogonal to the physical similarity of the face-blends, such that faces from the same category shared a family name but critically did *not* share a parent face (Figure 1B). Thus, there were no relevant or irrelevant physical features to be extracted for the categories and instead learning required that the category label be memorized for each face. This category structure is conceptually similar to the Type 6 category structure from seminal work by Shepard et al. (1961). Category bias can still be measured by comparing ratings for faces that do not share a parent but belong to the same category vs. those who do not share a parent and belong to different categories (i.e., faces that belong vs. not belong to the same category, controlled for physical similarity). Because the stimuli within a category do not share features, learning is unlikely to lead to true perceptual biases and instead any post-learning similarity rating bias would only reflect strategic judgment bias to rate faces as more similar due to the same category label (total bias = strategic judgment bias).

## Procedure

Experimental procedures were identical for participants in both conditions and consisted of four phases: passive viewing, pre-learning similarity ratings, feedback-based category learning, and post-learning similarity ratings. Each phase is detailed below.

### Passive viewing

Prior to any learning, participants were presented with each of the nine face-blends to familiarize them with the stimuli and the



degree of similarity between all faces before collecting perceived similarity ratings. Each of the nine training blends were viewed once (3s), in random order, without any labels and without making any responses. Trials were separated by a fixation inter-stimulus-interval (ISI; 1s).

### Pre-learning similarity ratings

Prior to learning, all possible 36 pairwise comparisons of the nine training blends were presented and participants rated the similarity of the two faces on a scale from one to six (1 = two faces appear very dissimilar, 6 = two faces appear very similar). Trials were presented for 6s with a fixation ISI (1s) after each response.

### Feedback-based category learning of families

On each trial, a single face was presented on the screen for 4s along with three family surnames (Miller, Wilson, Davis) as response options. Participants were instructed to indicate family membership via button press. After each response, corrective feedback was presented (2s), followed by a fixation ISI (1s). Participants completed five blocks of category learning separated by a self-paced break. Each block consisted of three repetitions of each stimulus, for a total of 15 repetitions per stimulus across the entire category learning phase.

### Post-learning similarity ratings

Perceived similarity ratings were again collected immediately after the learning phase with the same procedure as the pre-learning ratings. Pairwise comparisons of the faces were presented in a different random order than the pre-learning ratings.

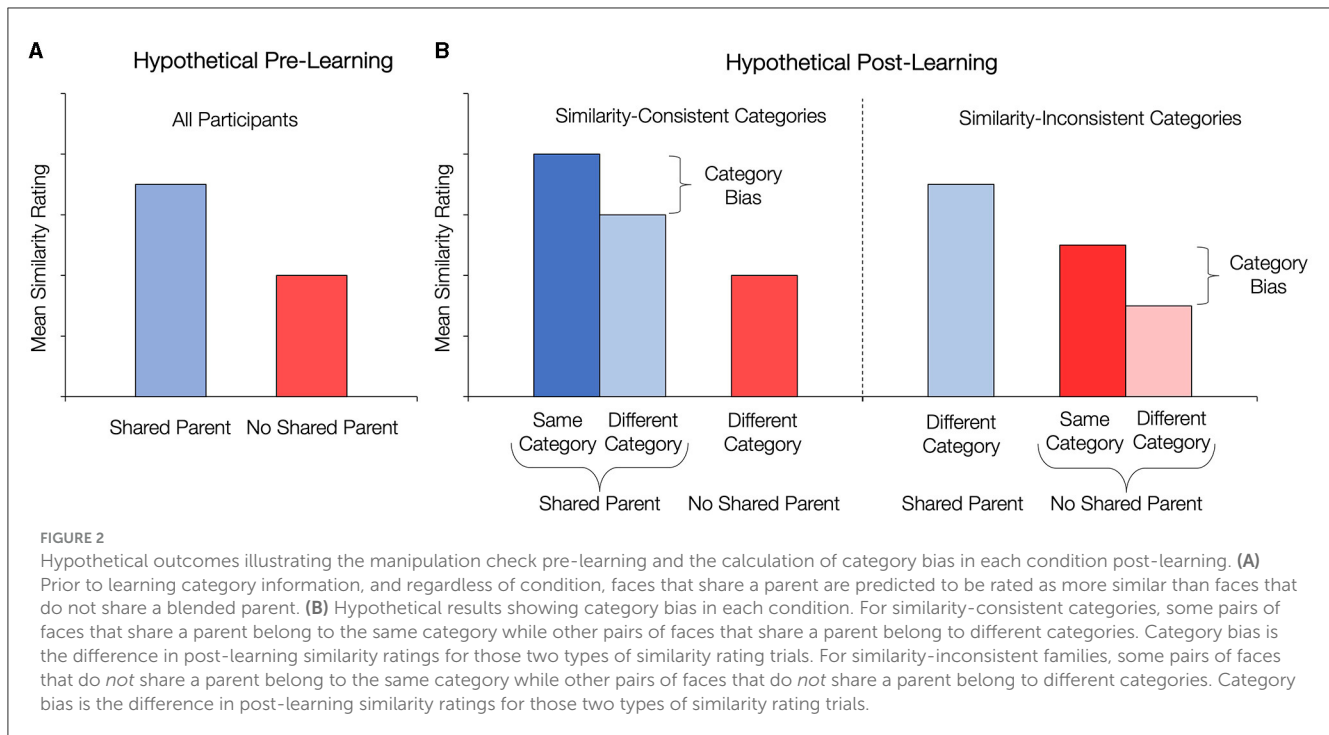
### Statistical analyses

#### Category learning

To determine how well participants learned the category labels in both conditions, we examined learning changes across time using a 2 [category structure condition (similarity-consistent, similarity-inconsistent)] x 5 [learning phase (block 1, block 2, block 3, block 4, block 5)] repeated measures ANOVA with learning phase as a within-subjects factor. Greenhouse-Geisser corrections for any violations of sphericity as detected via Mauchly's test are noted as GG in results.

#### Category bias

To establish that the blending procedure successfully manipulated the physical similarity and perceived similarity of the faces, we first compared the pre-learning similarity ratings for faces that shared a blending parent and faces that did not share a blending parent, collapsed across conditions, using a paired t-test. We predicted that faces that shared a parent would be rated as more similar than faces that did not share a parent (Figure 2A)



indicating that participants are sensitive to the similarity structure among the blended stimuli.

Of primary interest was the measure of category bias in perceived similarity ratings *after* learning (Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022). We sought to compare the degree of category bias when there are category-relevant features that can be extracted with the degree of category bias when there are no category-relevant features that can be extracted. Specifically, one of the pre-learning trial types (shared parent or no shared parent) can be further split after learning based on category membership, with the resulting trial types unique to each experimental condition (Figure 2B). In the similarity-consistent condition, the shared-parent trial type can be further split into two: faces that share a parent and have the same family name and faces that share a parent but have a different family name (Figure 2B, left panel, blue bars). Category bias in this condition was calculated from these shared-parent trial types by taking the difference of the similarity ratings for pairs of faces that were from the same family and similarity ratings for pairs of faces from different families (Figure 2B, left panel, difference between blue bars). Note that in the similarity-consistent condition, the no-shared-parent trials cannot be split into same and different category pairs because in this category structure only faces that shared a blended parent belonged to the same category. Thus, all faces that do not share a parent belong to different categories in this condition (see colored boxes illustrating categories determined across rows in Figure 1A).

In the similarity-inconsistent condition, the no-shared-parent trial type can be further split into two: faces that do not share a parent but belong to the same family and faces that do not share a parent and belong to different families (Figure 2B, right panel, red bars). Category bias in this condition was calculated from these no shared parent trial types by taking the difference of the similarity ratings for pairs of faces that were from the same family

and similarity ratings for pairs of faces from different families (Figure 2B, right panel, difference between red bars). Critically, in the similarity-inconsistent condition the shared-parent trials cannot be split into same and different category pairs because in this category structure only faces that **did not share** a blended parent belong to the same category. Thus, all faces that share a parent belong to different categories in this condition (Figure 1B). Importantly, the category bias in both conditions compares similarity ratings for faces that belong to the same category with ratings for faces that are equated for physical similarity but belong to different categories.

We first examined category bias (same category minus difference category) in each condition separately using a one-sample *t*-test to determine whether category bias is significantly greater than zero. To examine the relative contribution of perceptual change and strategic processes on similarity ratings, we then compared post-learning category bias between conditions using an independent samples *t*-test. The same measures of category bias were also extracted for the same pairs of faces from pre-learning similarity ratings and compared to post-learning category biases using a paired *t*-test. This allowed us to confirm that category biases measured after learning were driven by learning and not capturing potential random pre-existing biases. In addition to test statistics and *p*-values, we include the effect sizes and their confidence intervals throughout the results.

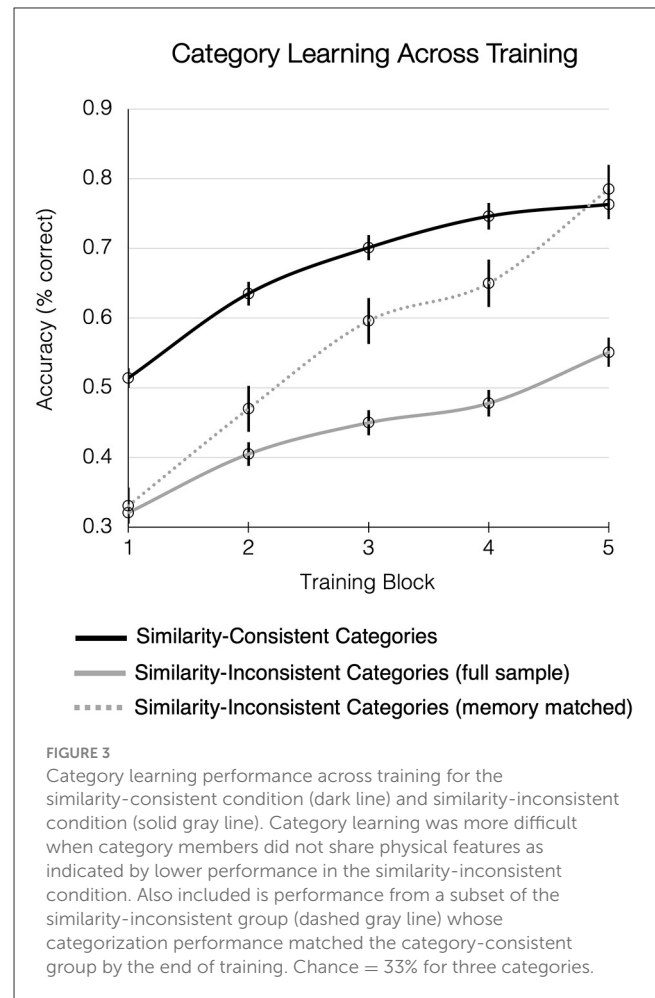
## Results

### Category learning performance

To examine changes in category learning performance across time, mean categorization accuracy was calculated for each of

the five blocks of the feedback-based category learning task (Figure 3). Categorization accuracy scores were submitted to a 2 x 5 repeated measures ANOVA, with experimental condition (similarity-consistent, similarity-inconsistent) as a between-subject factor and learning phase (block 1, block 2, block 3, block 4, block 5) as a within-subject factor. A significant main effect of condition [ $F(1, 207) = 113.48, p < 0.001, \eta_p^2 = 0.354$ ] indicated participants had higher category-learning performance when faces shared physical features [ $M_{\text{accuracy}} = 67\%$ , 95% CI (64%, 70%), chance = 33%] compared to when faces did not share any physical features [ $M_{\text{accuracy}} = 44\%$ , 95% CI (41%, 47%), chance = 33%]. A significant main effect of learning phase [ $F(2.85, 590.33) = 151.67, p < 0.001, \eta_p^2 = 0.42, GG$ ] with a significant linear trend [ $F(1, 207) = 288.70, p < 0.001, \eta_p^2 = 0.582$ ] indicated increased accuracy as learning progressed. Lastly, a significant interaction between experimental condition and learning phase [ $F(2.85, 590.33) = 3.92, p = 0.010, \eta_p^2 = 0.019, GG$ ] indicated differences in the learning rate between conditions. Follow-up comparisons of the two conditions on each block showed that while the differences in performance were large and significant throughout (all  $t > 7.0, p < 0.001, d > 0.97$  in each block), the effect size fluctuated among blocks. The largest difference in performance between conditions was observed in Block 1 of training ( $d = 1.39$ ) and the smallest difference in Block 5 ( $d = 0.97$ ). As suggested by Figure 3, performance in the similarity-inconsistent condition continued to improve through the fifth block of training while improvements slowed down earlier in the similarity-consistent condition.

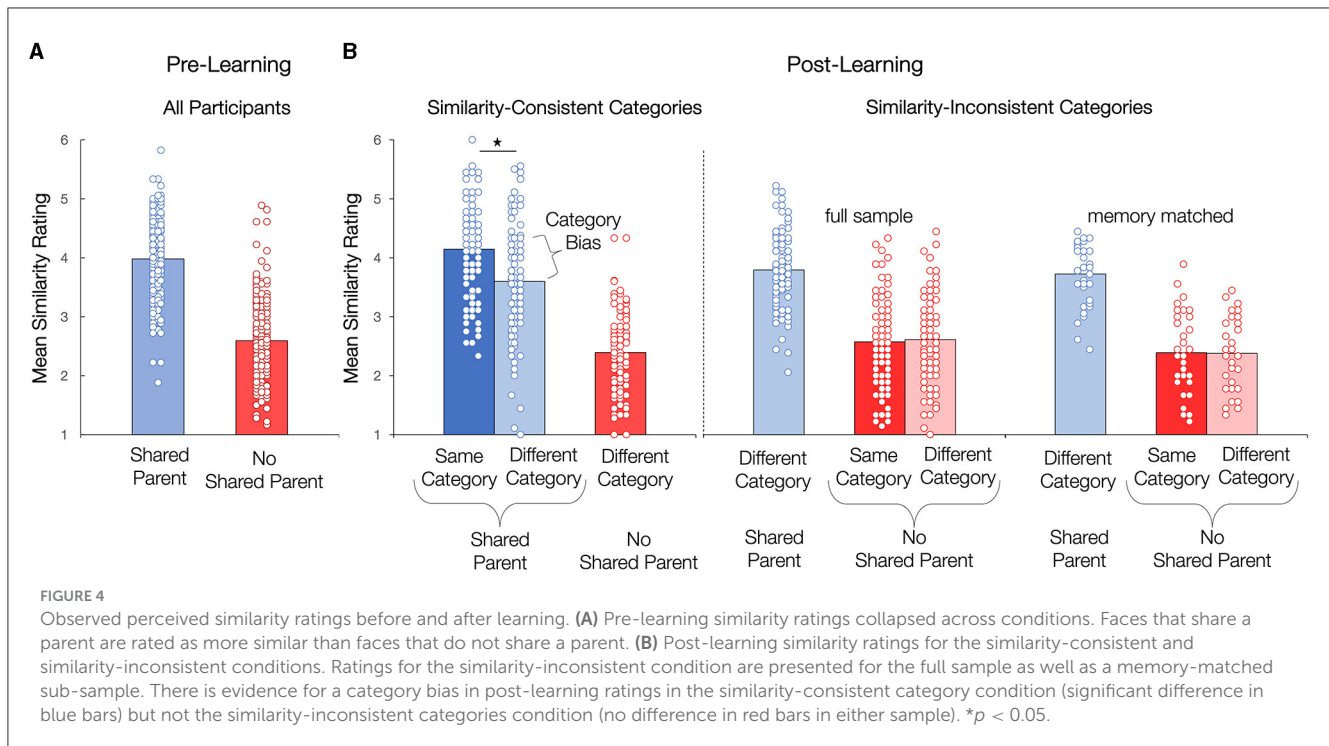
Overall, results from the training phase indicated that participants learned similarity-consistent categories to a higher degree than similarity-inconsistent categories. Namely, a lack of shared physical features between same-family faces made the learning task more difficult for participants in the similarity-inconsistent condition. Due to the low learning performance, measuring and interpreting category bias is a challenge in this condition. For example, if a category bias is not found in the similarity-inconsistent condition, it may argue against a contribution of strategic judgment bias or instead reflect a lack of opportunity to show a strategic bias because participants were not able to adequately learn the family labels. Thus, we conducted all the following analyses twice. In addition to analyses with the full category-inconsistent sample, we also focused on a subset of participants from the similarity-inconsistent condition whose final training block performance was above 60%. This resulted in a subset ( $n = 35$ ) whose average final training performance matched that from the similarity-consistent condition (see the dashed line in Figure 3). We refer to this subset as “similarity-inconsistent (memory-matched)”. Notably, using the memory-matched subset cannot eliminate the inherent difficulty differences across conditions, as evidenced by longer reaction times during the last block of training in the similarity-inconsistent (memory-matched) group ( $M_{\text{correctRT}} = 1.67$ ) compared to the similarity-consistent group [ $M_{\text{correctRT}} = 1.41; t(137) = 4.29, p < 0.001, d = 0.84$ ]. However, analyzing the subset of good learners in the similarity-inconsistent condition allows us to make sure that any potential differences in category bias are not driven simply by a lack of memory for the category labels in that condition.



## Category bias in similarity ratings

First, we examined similarity ratings prior to learning category information collapsed across conditions, to verify that participants were sensitive to the similarity manipulation. Results presented in Figure 4A indicated that faces which shared a parent were rated as more similar than faces that did not share a parent [ $M_{\text{diff}} = 1.39; t(208) = 31.56, p < 0.001, d = 2.18, 95\% \text{ CI } (1.93, 2.43)$ ], confirming that our blending procedure worked as intended. As a secondary check, we examined whether there was evidence of a pre-existing bias toward category-membership prior to learning. As anticipated, there was no evidence for a category bias in perceptual similarity ratings prior to category learning in either condition [similarity-consistent:  $t(103) = 0.07, p = 0.47, d = 0.007, 95\% \text{ CI } (-0.19, 0.20)$ ; similarity-inconsistent:  $t(104) = -0.13, p = 0.45, d = -0.013, 95\% \text{ CI } (-0.20, 0.18)$ ].

To measure the emergence of a category bias after learning, we first examined category bias in post-learning similarity ratings in each condition. In line with prior work, we found evidence for a post-learning category bias in the similarity-consistent condition (Figure 4B, left panel, significant difference between blue bars;  $t(103) = 4.47, p < 0.001, d = 0.44, 95\% \text{ CI } (0.24, 0.64)$ ). A paired-samples t-test confirmed that the category-bias observed post-learning was not a result of



pre-existing differences in perceived similarity ratings prior to learning [post-bias vs. pre-bias  $M_{diff} = 0.54$ ;  $t(103) = 4.72$ ,  $p < 0.001$ ,  $d = 0.46$ , 95% CI (0.26, 0.66)]. In comparison to pre-learning, participants increased their similarity rating for faces that shared a blended parent but belonged to the same family category [ $t(105) = 1.86$ ,  $p = 0.066$ ,  $d = 0.18$ ] and decreased their similarity rating for faces that shared a blended parent but belonged to different family categories [ $t(105) = -4.30$ ,  $p < 0.001$ ,  $d = -0.42$ ]. Participants who showed better category learning during training showed stronger post-learning category bias [ $R^2 = 0.27$ ,  $F(1, 102) = 37.71$ ,  $p < 0.001$ ], consistent with the notion that the bias is driven by category learning.

No evidence for a category bias was found in the similarity-inconsistent condition, neither in the full sample [Figure 4B, middle panel, no difference between red bars;  $t(104) = -0.82$ ,  $p = 0.207$ ,  $d = -0.08$ , 95% CI (-0.27, 0.11)] nor in the memory-matched sub-sample [Figure 4B, right panel, no difference between red bars;  $t(34) = 0.16$ ,  $p = 0.44$ ,  $d = 0.027$ , 95% CI (-0.30, 0.36)]. Consistent with the lack of bias among the top learners, there was also no relationship between how well participants learned the category labels during training and their subsequent category bias [ $R^2 = 0.01$ ,  $F(1, 103) = 1.13$ ,  $p = 0.29$ ], as there was little bias among both good and poor learners. Thus, even when controlling for memory for category labels, a category-bias in post-learning similarity ratings was only evident for categories where members shared physical features. No evidence for category bias was observed when stimuli within a category did not share features, indicating that without a physical similarity structure, a shared category label is not sufficient to induce a bias in perceived similarity ratings.

## Discussion

Categorization tasks utilizing artificial or photography-based faces have decades of tradition (Etcovv and Magee, 1992; Calder et al., 1996; Young et al., 1997; Stevenage, 1998; Goldstone and Steyvers, 2001; Goldstone et al., 2001; Roberson et al., 2007; Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022). Here, we utilized blended faces because it allowed us to control the similarity structure of the categories while offering stimuli that are realistic looking and holistically perceived. We sought to determine the degree to which perceived similarity ratings in the face-blend categorization task used in our prior work (Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022) are driven by perceptual and strategic judgment biases. We tracked post-learning category bias after a traditional feedback-based category learning task under two learning conditions. In the similarity-consistent condition, category membership was consistent with the similarity structure among stimuli where category bias can reflect a combination of a perceptual bias and a strategic judgment bias. In the similarity-inconsistent condition, category membership was orthogonal from physical features and any similarity rating bias could be attributed to a strategic judgment bias alone. Results showed evidence for a category bias in post-learning similarity ratings, but only when items shared physical features within categories. Thus, we replicated the findings of categorization-driven biases, but also identified a boundary condition on the existence of a similarity structure within categories. Because no category bias was observed in the condition where items shared labels but did not share physical characteristics, the emergence of a category bias in similarity ratings appears to largely reflect a change in perceived



similarity and not merely a strategic decision to rate items with shared labels more similarly.

Our new approach demonstrating a role of perceptual changes in category-biased similarity ratings provides further support for an attentional theory of category learning and category-induced biases. Past work has postulated that learning category information prompts individuals to shift their attention toward category-relevant features and away from category-irrelevant features (Nosofsky, 1986, 1987), which leads to changes in perception to highlight within-category similarity and between category differences (Goldstone, 1994a; Beale and Keil, 1995; Kurtz, 1996; Livingston et al., 1998; Goldstone and Steyvers, 2001; Gureckis and Goldstone, 2008; Folstein et al., 2013a,b; Soto, 2019). For example, when one acquires a new category where category membership depends on brightness but not size, increased attention to brightness “stretches” that dimension, enhancing perception of differences in brightness. In contrast, because size is irrelevant for category membership, decreased attention to the size dimension “shrinks” the perceptual space along the size dimension, decreasing perception of size differences (Goldstone, 1994a). Within the context of the current task, imagine that one of the category-relevant parents might have prominent eyebrows but the three relevant parents do not differ much in their mouth shape. The increased attention to eyebrows and decreased attention to mouth shape after category learning would lead participants to be more likely to perceive eyebrow differences (highlighting between-category differences and within-category similarity of eyebrows) and less strongly perceive mouth differences (making any within-category differences of mouth shape driven by irrelevant parent features less salient). Because faces are processed differently by the brain compared to other types of objects (for reviews see Ruiz-Soler and Beltran, 2006; Tsao and Livingstone, 2008; Behrmann et al., 2016) and face perception may be “special” in that faces are processed on a more holistic level than other types of stimuli (Farah et al., 1998; for review see Piepers and Robbins, 2012), it will be important to test generalization of current findings to other types of stimuli. Nevertheless, as category-driven attentional biases have been previously shown for many types of stimuli (Bornstein et al., 1976; Bornstein and Korda, 1984; Livingston et al., 1998; Goldstone and Steyvers, 2001; Franklin and Davies, 2004), the principles of how category learning biases perception through stretching and shrinking perceptual space along different dimensions may be relatively universal.

Our behavioral data suggesting a bias in perception, consistent with shifts of attention, further align with evidence for an attentional shift as well as reports of category bias reflected in the brain. For example, Mack et al. (2013) found neural representations for category features distributed widespread across the cortex but only when attention weights to individual stimulus features were considered. Ashby and colleagues (2020) found neural biases toward category-relevant information widespread throughout the cortex when items shared physical features. The current study contributes further support for the attentional theory of category bias by demonstrating that category bias only emerges under conditions where items within categories share physical features, giving individuals category-relevant features to which they may attend.

Although our findings are consistent with the proposals that category learning results in attentional shifts from category irrelevant to relevant features, creating perceptual biases, we cannot completely rule out the possibility that strategic bias may still contribute to the total bias observed for similarity-based categories. For example, fast category learning in the similarity-consistent condition may have allowed participants to engage in attentional tuning to category-diagnostic features more extensively (Blair et al., 2009; Chen et al., 2013; McColeman et al., 2014), which in turn may have facilitated faster access to the associated category labels during the rating task and greater impact of labels in ratings. The differences in reaction times between conditions, even when limited to the top learners in the similarity-inconsistent condition, indeed indicate that the label became available to participants much faster in the similarity-consistent than the similarity inconsistent condition. Furthermore, the similarity-inconsistent condition required rote memorization of face-label pairs, which may not be sufficient to evoke a strategic bias. For example, Fotiadis and Protopapas (2022) found facilitatory effects of verbal labels on category learning but **not** for a paired-associate task that did not involve similarity-based categories which may more closely resemble the similarity-inconsistent condition in the current study. As category labels may not have appeared as meaningful information in the similarity-inconsistent condition, the effect of the shared label may have been minimal on both perceptual and strategic judgments when it was not strongly activated by the stimulus (Lupyan, 2012). Conceptually consistent with the Fotiadis and Protopapas (2022) study, our findings demonstrate that shared category labels, *per se*, are not sufficient to create category biases in similarity ratings when categories do not align with similarity structure, providing a new boundary condition on the effect of language on perception.

Our findings indicated that the role of strategic judgment in perceived similarity ratings in the current task is likely limited. In contrast, Goldstone and colleagues (2001) concluded that category bias in similarity ratings measures not only the perceptual similarity of the items themselves but also the similarity of their associated labels and categories. One reason for the differences in findings between Goldstone and colleagues (2001) and the current study may be differences in experimental design. First, a category bias (greater within- vs. between-category similarity) was already present prior to learning in the Goldstone et al. (2001) study because items within a category tended to be more similar than items from different categories. Here, we only compared items equated for physical similarity within and between categories to ensure that the category bias, and subsequent changes in perceived similarity, are due to category learning alone and cannot be explained by pre-existing similarity differences. In fact, participants in our similarity-inconsistent condition had to work *against* the similarity structure among the stimuli to learn the category assignments. Second, category bias that reflected a strategic judgment bias in the previous study was measured indirectly through similarity ratings to a separate, non-categorizable, neutral face. Furthermore, participants were trained to learn category information under conditions where shared similarity within and between categories was present in varying degrees. In contrast, we trained participants to learn category information using a

category similarity structure that held similarity constant within and between categories (i.e., similarity-consistent group) or held similarity orthogonal to within and between category designation (i.e., similarity-inconsistent group), allowing for a direct and clear measurement of category bias. Finally, the current study is the first to our knowledge that measures the degree of strategic judgment bias that may be inherent to similarity ratings by including a comparison learning condition where physical similarity is orthogonal to category structure and associated category labels. Nevertheless, it will be important to explore the strategic bias question in the future with other paradigms, including arbitrary or “functional” categories where stimuli with no inherent similarity structure are arbitrarily assigned to categories (Pan and Sakagami, 2012; Taylor et al., 2021), to fully determine the extent to which verbal labels influence similarity judgments across a broader range of category structures.

In our prior behavioral and neuroimaging work, we found evidence for category biases in face similarity ratings even when controlling for physical similarity (Ashby et al., 2020; Ashby and Zeithamova, 2022; Bowman et al., 2022). These findings suggest that when a degree of shared similarity exists and a verbal label aligning with some of those shared features exists, the brain may be hardwired to allocate a disproportionate amount of attention to shared features. This effect may be one of the mechanisms behind perpetuating social categories and stereotypes (Amodio, 2014). For example, racial biases may be driven by some physical similarities within groups and differences between groups that are then overexaggerated in people’s perceptions (Byatt and Rhodes, 2004). Objectively, the degree of variability across facial features within various races is small (Goldstein, 1979a,b). However, when individuals differ along some physical feature (such as skin color) and a category label (such as race) becomes associated with that feature, the feature can become emphasized in people’s perceptions. In contrast, any features not aligned with category membership become less perceptually salient or ignored. This would alter our perceptions, highlighting within-category similarity and between-category differences, and biasing us to perceive the categories as more differentiated than they really are.

Interestingly, we also found that shared features were necessary to induce a category bias in perception in the current study. Yet, not all biases in the real world, such as religious bias or political affiliation bias, can be tied to salient physical features. Therefore, these results may indicate that different mechanisms could be responsible for types of in-group/out-group biases that are not grounded in physical similarity. Rather than relying on attentional biases to category relevant features, biases that are not similarity-based may work through associative inference, where value or affect associated with one member of a category spreads to other members of the category through associations (Pan and Sakagami, 2012; Wimmer and Shohamy, 2012; Martinez et al., 2016). More broadly, our results indicate that category learning involving similarity-based structure and learning involving arbitrary category labels likely involve distinct mechanisms.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

## Ethics statement

The studies involving humans were approved by University of Oregon Research Compliance Services. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

SA: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing—original draft, Writing—review & editing. BC: Project administration, Software, Writing—review & editing. DZ: Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing—original draft, Writing—review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Institute for Neurological Disorders and Stroke Grant R01-NS112366 awarded to DZ.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nat. Rev. Neurosci.* 15, 670–682. doi: 10.1038/nrn3800
- Ashby, F. G., and Maddox, W. T. (2005). Human category learning. *Annual Rev. Psychol.* 56, 149–178. doi: 10.1146/annurev.psych.56.091103.070217
- Ashby, F. G., and Maddox, W. T. (2011). Human category learning 2.0. *Annals N. Y. Acad. Sci.* 1224, 147–161. doi: 10.1111/j.1749-6632.2010.05874.x
- Ashby, S. R., Bowman, C. R., and Zeithamova, D. (2020). Perceived similarity ratings predict generalization success after traditional category learning and a new paired-associate learning task. *Psychon. Bull. Rev.* 27, 791–800. doi: 10.3758/s13423-020-01754-3
- Ashby, S. R., and Zeithamova, D. (2022). Category-biased neural representations form spontaneously during learning that emphasizes memory for specific instances. *J. Neurosci.* 42, 865–876. doi: 10.1523/JNEUROSCI.1396-21.2021
- Beale, J. M., and Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition* 57, 217–239. doi: 10.1016/0010-0277(95)00669-X
- Behrmann, M., Scherf, K. S., and Avidan, G. (2016). Neural mechanisms of face perception, their emergence over development, and their breakdown. *Wiley Interdiscip. Rev. Cogn. Sci.* 7, 247–263. doi: 10.1002/wcs.1388
- Blair, M. R., Watson, M. R., and Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition* 112, 330–336. doi: 10.1016/j.cognition.2009.04.008
- Booth, A. E. (2014). Conceptually coherent categories support label-based inductive generalization in preschoolers. *J. Exp. Child Psychol.* 123, 1–14. doi: 10.1016/j.jecp.2014.01.007
- Bornstein, M. H., Kessen, W., and Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 115–129. doi: 10.1037/0096-1523.2.1.115
- Bornstein, M. H., and Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychol. Res.* 46, 207–222. doi: 10.1007/BF00308884
- Bowman, C. R., Ashby, S. R., and Zeithamova, D. (2022). Age effects on category learning, categorical perception, and generalization. *Memory* 30, 1–18. doi: 10.1080/09658211.2021.2003818
- Byatt, G., and Rhodes, G. (2004). Identification of own-race and other-race faces: implications for the representation of race in face space. *Psychon. Bull. Rev.* 11, 735–741. doi: 10.3758/BF03196628
- Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., and Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cogn.* 3, 81–118. doi: 10.1080/713756735
- Chen, L., Meier, K. M., Blair, M. R., Watson, M. R., and Wood, M. J. (2013). Temporal characteristics of overt attentional behavior during category learning. *Atten. Percept. Psychophys.* 75, 244–256. doi: 10.3758/s13414-012-0395-8
- Curran, T., Tanaka, J. W., and Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2, 1–18. doi: 10.3758/CABN.2.1.1
- Etcovv, N. L., and Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition* 44, 227–240. doi: 10.1016/0010-0277(92)90002-Y
- Farah, M. J., Wilson, K. D., Drain, M., and Tanaka, J. N. (1998). What is “Special” about face perception? *Psychol. Rev.* 105, 482–498. doi: 10.1037/0033-295X.105.3.482
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G\*Power: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Folstein, J. R., Gauthier, I., and Palmeri, T. J. (2013a). How category learning affects object representations: not all morphspaces stretch alike. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1–7. doi: 10.1037/a0025836
- Folstein, J. R., Palmeri, T. J., and Gauthier, I. (2013b). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex* 23, 814–823. doi: 10.1093/cercor/bhs067
- Folstein, J. R., Palmeri, T. J., and Gauthier, I. (2014). Perceptual advantage for category-relevant perceptual dimensions: the case of shape and motion. *Front. Psychol.* 5, 1–7. doi: 10.3389/fpsyg.2014.01394
- Fotiadis, F. A., and Protopapas, A. (2022). Immediate and sustained effects of verbal labels for newly-learned categories. *Q. J. Exp. Psychol.* 76, 1773–1789. doi: 10.1177/17470218221126659
- Franklin, A., and Davies, I. R. L. (2004). New evidence for infant colour categories. *Br. J. Dev. Psychol.* 22, 349–377. doi: 10.1348/0261510041552738
- Goldstein, A. G. (1979a). Facial feature variation: anthropometric data II. *Bull. Psychon. Soc.* 13, 191–193. doi: 10.3758/BF03335056
- Goldstein, A. G. (1979b). Race-related variation of facial features: anthropometric data I. *Bull. Psychon. Soc.* 13, 187–190. doi: 10.3758/BF03335055
- Goldstone, R. L. (1994a). Influences of categorization on perceptual discrimination. *J. Exp. Psychol. Gen.* 123, 178–200. doi: 10.1037/0096-3445.123.2.178
- Goldstone, R. L. (1994b). The role of similarity in categorization: providing a groundwork. *Cognition* 52, 125–157. doi: 10.1016/0010-0277(94)90065-5
- Goldstone, R. L., Kersten, A., and Carvalho, P. F. (2018). “Categorization and concepts” In *Stevens’ Handbook of Experimental Psychology and Cognitive Neuroscience, Language and Thought: Developmental and Social Psychology*. 4th ed. Eds J. T. Wixted, S. L. Thompson-Schill (Hoboken, NJ: John Wiley and Sons, Incorporated), 275–317.
- Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition* 78, 27–43. doi: 10.1016/S0010-0277(00)00099-8
- Goldstone, R. L., and Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *J. Exp. Psychol. Gen.* 130, 116–139. doi: 10.1037/0096-3445.130.1.116
- Gureckis, T. M., and 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*. 10.
- Hanley, J. R., and Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychon. Bull. Rev.* 18, 355–363. doi: 10.3758/s13423-010-0043-z
- Herzmann, G., Minor, G., and Adkins, M. (2017). Neural correlates of memory encoding and recognition for own-race and other-race faces in an associative-memory task. *Brain Res.* 1655, 194–203. doi: 10.1016/j.brainres.2016.10.028
- Juárez, F. P. G., Sicotte, T., Thériault, C., and Harnad, S. (2019). Category learning can alter perception and its neural correlates. *PLoS ONE* 14, 1–30. doi: 10.1371/journal.pone.0226000
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* 99, 22–44. doi: 10.1037/0033-295X.99.1.22
- Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connect. Sci.* 8, 225–247. doi: 10.1080/095400996116893
- Kurtz, K. J. (1996). Category-based similarity. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (Routledge), 290.
- Levin, D. T., and Angelone, B. L. (2002). Categorical perception of race. *Perception* 31, 567–578. doi: 10.1068/p3315
- Levin, D. T., and Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Percept. Psychophys.* 62, 386–401. doi: 10.3758/BF03205558
- Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417
- Livingston, K. R., Andrews, J. K., and Harnad, S. (1998). Categorical perception effects induced by category learning. *J. Exp. Psychol. Learn. Memory Cogn.* 24, 732–753. doi: 10.1037/0278-7393.24.3.732
- Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychol.* 3, 54. doi: 10.3389/fpsyg.2012.00054
- Mack, M. L., Preston, A. R., and Love, B. C. (2013). Decoding the brain’s algorithm for categorization from its neural implementation. *Current Biol.* 23, 2023–2027. doi: 10.1016/j.cub.2013.08.035
- Markman, A. B., and Ross, B. H. (2003). Category use and category learning. *Psychol. Bull.* 129, 592–613. doi: 10.1037/0033-2909.129.4.592
- Martinez, J. E., Mack, M. L., Gelman, B. D., and Preston, A. R. (2016). Knowledge of social affiliations biases economic decisions. *PLoS ONE*, 11. doi: 10.1371/journal.pone.0159918
- McColeman, C. M., Barnes, J. I., Chen, L., Meier, K. M., Walshe, R. C., and Blair, M. R. (2014). Learning-induced changes in attentional allocation during categorization: a sizable catalog of attention change as measured by eye movements. *PLoS ONE*, 9, e83302. doi: 10.1371/journal.pone.0083302
- Meissner, C. A., and Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychol. Public Pol. Law* 7, 3–35. doi: 10.1037/1076-8971.7.1.3
- Miner, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods Inst. Comp.* 36, 630–633. doi: 10.3758/BF03206543

- Mishra, M. V., Likitlersuang, J., B., Wilmer, J., Cohan, S., Germine, L., et al. (2019). Gender differences in familiar face recognition and the influence of sociocultural gender inequality. *Sci. Rep.* 9, 1–12. doi: 10.1038/s41598-019-54074-5
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 104–114. doi: 10.1037/0278-7393.10.1.104
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–57. doi: 10.1037/0096-3445.115.1.39
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 87–108. doi: 10.1037/0278-7393.13.1.87
- Pan, X., and Sakagami, M. (2012). Category representation and generalization in the prefrontal cortex. *Eur. J. Neurosci.* 35, 1083–1091. doi: 10.1111/j.1460-9568.2011.07981.x
- Peer, P. (1999). *CVL Face Database*. Computer Vision Lab, Faculty of Computer and Information Science, University of Ljubljana, Slovenia.; Computer Vision Lab, Faculty of Computer and Information Science, University of Ljubljana, Slovenia. Available online at: <http://www.lrv.fri.uni-lj.si/facedb.html> (accessed, 2018).
- Piepers, D. W., and Robbins, R. A. (2012). A review and clarification of the terms “holistic,” “configural,” and “relational” in the face perception literature. *Front. Psychol.* 3, 559. doi: 10.3389/fpsyg.2012.00559
- Pilling, M., Wiggett, A., Ozgen, E., and Davies, I. R. (2003). Is color “categorical perception” really perceptual? *Mem. Cogn.* 31, 538–551. doi: 10.3758/BF03196095
- Pothos, E. M., and Reppa, I. (2014). The fickle nature of similarity change as a result of categorization. *Q. J. Exp. Psychol.* 67, 2425–2438. doi: 10.1080/17470218.2014.931977
- Roberson, D., Damjanovic, L., and Pilling, M. (2007). Categorical perception of facial expressions: evidence for a “category adjustment” model. *Mem. Cogn.* 35, 1814–1829. doi: 10.3758/BF03193512
- Roberson, D., and Davidoff, J. (2000). The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cogn.* 28, 977–986. doi: 10.3758/BF03209345
- Rosch, E., and Mervis, C. B. (1975). Family resemblances. *Cogn. Psychol.* 7, 573–605. doi: 10.1016/0010-0285(75)90024-9
- Ruiz-Soler, M., and Beltran, F. S. (2006). Face perception: an integrative review of the role of spatial frequencies. *Psychol. Res.* 70, 273–292. doi: 10.1007/s00426-005-0215-z
- Sandell, J. H., Gross, C. G., and Bornstein, M. H. (1979). Color categories in macaques. *J. Comp. Physiol. Psychol.* 93, 626–635. doi: 10.1037/h0077594
- Seeger, C. A., and Miller, E. K. (2010). Category learning in the brain. *Annual Rev. Neurosci.* 33, 203–219. doi: 10.1146/annurev.neuro.051508.135546
- Shepard, R. N., Hovland, C. I., and Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychol. Monographs Gen. Appl.* 75, 1–42. doi: 10.1037/h0093825
- Simanova, I., Francken, J. C., de Lange, F. P., and Bekkering, H. (2016). Linguistic priors shape categorical perception. *Lang. Cogn. Neurosci.* 31, 159–165. doi: 10.1080/23273798.2015.1072638
- Sloutsky, V. M., and Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1. Linguistic labels and the development of similarity judgment. *Dev. Psychol.* 35, 1478–1492. doi: 10.1037/0012-1649.35.6.1478
- Soto, F. A. (2019). Categorization training changes the visual representation of face identity. *Atten. Percept. Psychophys.* 81, 1220–1227. doi: 10.3758/s13414-019-01765-w
- Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *Br. J. Psychol.* 89, 39–57. doi: 10.1111/j.2044-8295.1998.tb02672.x
- Taylor, J. E., Cortese, A., Barron, H. C., Pan, X., Sakagami, M., and Zeithamova, D. (2021). How do we generalize? *Neuron. Behav. Data Anal. Theory* 1, 1–39. doi: 10.51628/001c.27687
- Timeo, S., Farroni, T., and Maass, A. (2017). Race and color: two sides of one story? Development of biases in categorical perception. *Child Dev.* 88, 83–102. doi: 10.1111/cdev.12564
- Tsao, D. Y., and Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Rev. Neurosci.* 31, 411–437. doi: 10.1146/annurev.neuro.30.051606.094238
- Wimmer, G. E., and Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science* 338, 270–273. doi: 10.1126/science.1223252
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *PNAS*, 104, 7780–7785. doi: 10.1073/pnas.0701644104
- Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., and Perrett, D. I. (1997). Facial expression megamix: tests of dimensional and category accounts of emotion recognition. *Cognition* 63, 271–313. doi: 10.1016/S0010-0277(97)00003-6