

Label Similarity and Stimulus Similarity Interact in Categorization

Wesley Leong (wesley.leong@uconn.edu)

Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA

Casey L. Roark (casey.roark@unh.edu)

Department of Psychology, University of New Hampshire, Durham, NH, USA

Eiling Yee (eiling.yee@uconn.edu)

Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA

Abstract

When learning to categorize stimuli, do we assume similar things should have similar labels? Are people more likely to respond with closer labels (e.g. 2-1 vs 2-4) when stimuli are more similar to each other? Across five experiments, we report evidence of such a bias and demonstrate that it can surface across a wide range of stimulus modalities and features, and persists regardless of participants' prior knowledge of the dimensions relevant for categorization. We also characterize some of the limits of this effect: it appears sensitive to the specific configuration of label-stimulus mappings, and may depend on overt similarity relations in label space. At minimum, our findings indicate the need to consider label-stimulus configurations when designing categorization experiments. They also hint more broadly at how label-to-stimulus mappings may affect how we structure novel categories.

Keywords: concepts; labels; categorization; similarity; cross-modal

Introduction

What is the relationship between labels and categories? Much work has been done to outline the myriad ways in which labels shape our understanding and perception of categories (e.g. Goldstone, 1994; Lupyan et al., 2007; Lupyan 2012; Smith et al., 1992; Waxman & Markow, 1995; Yamauchi & Markman, 2000). Many of these studies have shown how shared labels guide our perceptions of object similarity, or our assumptions of hidden traits. For example, Livingston, Andrews, and Harnad (1998) found that participants rated images of single-celled organisms as more similar to each other when they shared a label. Fewer studies have looked at the effect on classification behavior when labels are *similar* but not identical (cf. Sloutsky & Fisher, 2012). Here, we focus on this often-overlooked property of categorization experiments—the contribution of similarity structure in label space to biases in categorization behavior.

We start small by presenting experiments using 4-way categorization tasks, in which participants perceive a series of stimuli and classify them into categories by pressing one of four keys. We treat the keys used for the categorization responses as the labels of those categories (e.g. the key “1” mapped to category “1”). In many studies where people learn to categorize novel objects, the experimental design

implicitly assumes that the specific response keys are arbitrary, and thus would be interchangeable (e.g. Maddox & Ashby, 1993; Ashby & Maddox, 1993; Ashby, Queller, & Berretty, 1999). However, we question this assumption of independence. Studies in other domains have found evidence that response keys can interact with seemingly innocuous experimental parameters (e.g. SNARC effect, Dehaene, Bossini, & Giraux, 1993). For example, Lakens et al. (2011) found that the direction and size of the Stroop effect can be affected by how far apart the response keys are. This suggests that in a categorization task, structure among the stimuli to be categorized may in fact interact with structure in the response keys, such that changes in the alignment of those structures can affect behavior.

We propose one way such an interaction might manifest: a bias toward choosing “proximal” labels (e.g., 2-1) for two perceptually similar stimuli and “distal” labels (e.g. 2-4) for dissimilar stimuli¹. In other words, we seek to examine not just whether people categorize stimuli on the basis of similarity to known exemplars (e.g. Estes, 1994; Nosofsky, 1992), but whether this interacts with the similarity of the *labels* themselves. Evidence of such a “proximity bias” would suggest that labels and their referents can be shaped by how the similarity of labels aligns with the similarity of stimuli. We report evidence supporting this “proximity bias” through novel analyses of five experiments, including publicly available data from existing publications (Roark, Smayda, & Chandrasekaran, 2022; Roark & Chandrasekaran, 2023).

Experiment 1

We analyzed open data from an auditory categorization task (Roark & Chandrasekaran, 2023). As the full methods have been described elsewhere, we present an abbreviated version here and highlight the novel analyses. In brief—participants completed 600 trials of a 4-way categorization task in which stimuli varied along two auditory dimensions. Categorization boundaries divided the stimulus space into quadrants, and each category was assigned a numeric label from 1-4. Over the course of the experiment, participants sorted auditory

¹ Because numeric distance and spatial distance are coupled for the numbers 1-4 on a standard keyboard, we do not specify a precise dimension of label “proximity”.

stimuli into categories and received feedback about their responses. More precise analyses of learning rates and individual consistency are detailed in the original paper; here, we present a novel analysis aimed at identifying a proximity bias.

Methods

One hundred and thirty-three participants were recruited using Prolific (www.prolific.com) and completed the experiment through Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). Although Roark and Chandrasekaran (2023) ran two experiments, they are effectively similar for our purposes, thus we present them here in aggregate.

Stimuli 300 auditory stimuli (artificial nonspeech ripples) were synthesized, varying along two dimensions: spectral modulation (cycles/octave) and temporal modulation (Hz). These stimuli populated quadrants in two-dimensional stimulus space (Figure 1A). Each category was centered in one of these quadrants, resulting in 75 stimuli per category. Importantly for the present analyses, each category had two equidistant category neighbors in stimulus space. However, one of these neighbor categories had a numerically proximal label (e.g. 2-1), and the other had a numerically distal label (e.g. 2-4) (Figure 1B). One can thus quantify, within a range of inter-stimulus distances, any bias for one neighboring category label compared to its counterpart.

Procedure Participants completed 500 or 600 trials in a randomized order. Each trial started with a fixation cross for 1s, followed by an auditory stimulus for 1s, then participants made an untimed response by pressing 1/2/3/4 on their keyboard. During the response period, these numbers were displayed horizontally on the participants' screen, corresponding to the category labels. After making a response, participants received feedback for 1s about the accuracy of their response (correct or incorrect). There was no inter-stimulus interval.

Analysis We analyzed participants' response behavior to determine if they show the predicted proximity bias. The logic behind our analysis is as follows: On trials where they receive "correct" feedback, participants can be confident of the category of that stimulus. For trials *following* these correct trials, we can calculate a potential proximity bias by examining participants' likelihood of making a numerically proximal response when the stimulus is *perceptually* similar to the one they just correctly categorized².

For each participant, we first selected all trial pairs for which the first trial had been correctly categorized. Then, these pairs were binned by normed Euclidean distance between the stimuli based on their *perceptual* features,

² Although we use the term "perceptual" similarity, distance is calculated based on acoustic similarity as a convenient proxy. We acknowledge, however, that acoustic and perceptual similarity may not be identical.

resulting in 10 bins of inter-stimulus distance windows. Given a correct response, i , we operationalized bias for a subsequent response j within this distance window as:

$$Bias_{i,j} = \frac{R_{i,j} - A_{i,j}}{\text{Correct responses}_i}$$

where $R_{i,j}$ is the number of j responses following a correct response i , and $A_{i,j}$ is the number of times category j *actually* followed a correct response i . Because this scales with the total number of correct responses, we normalized it by the number of correct i trials, thus obtaining a measure of bias that is agnostic of overall accuracy. If participants choose a particular response more frequently than it actually occurs, this value is positive. Likewise, underresponding results in a negative bias.

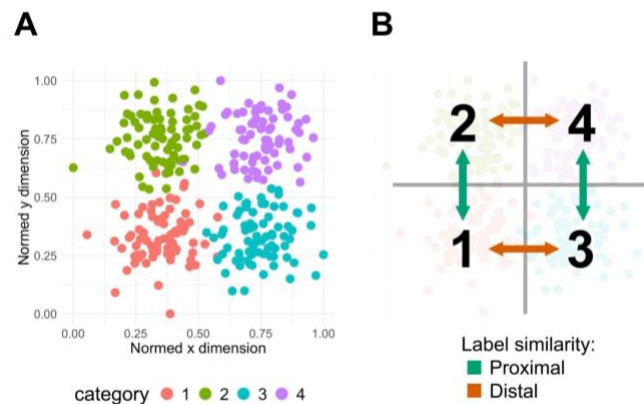


Figure 1: Example of a stimulus distribution across two dimensions (A: individual stimuli, B: overlaid with category labels illustrating proximity structure). All experiments in this paper followed the same general schematic of stimuli separated into quadrants and category labels running 1/2/4/3 clockwise, although precise stimulus and boundary placements differed.

We thus obtained, for each category, bias measures for the two neighboring categories in stimulus space (one of which has a proximal and the other has a distal numeric label; Figure 1B) and quantified them as a function of increasing inter-stimulus distance. We predicted that participants would have a preference for the proximal label (e.g. 2-1 compared to 2-4) when stimuli were more perceptually similar to each other, and that this pattern would shift as inter-stimulus perceptual distance increased, resulting in an interaction between label proximity and perceptual distance on response bias. To capture this interaction, we subtracted distal bias from proximal bias at each distance bin and ran a linear mixed-effects model on the difference.

Results

As Roark and Chandrasekaran (2023) reported, participants were able to learn this task to some degree: mean accuracy at the end of the experiment was 45-50% (with chance being 25%). As noted, in-depth analyses of participant performance are outlined in their paper and we do not discuss them further.

We were interested in whether people show a bias toward a numerically proximal label (e.g. 2-1) when successive stimuli are perceptually similar to each other, compared to a numerically distal label (e.g. 2-4), and vice-versa when the stimuli are perceptually far apart. Thus we examined the interaction between label bias (proximal vs. distal) and inter-stimulus distance. The results showed a statistically significant effect of this proximity bias ($\beta = -0.17, p < .001$; Figure 2B)—the negative slope reflects that as stimuli were closer together, people showed a preference for responding with the numerically proximal label, but this preference reversed as the stimuli got further apart.

Follow-up analyses revealed that this effect was present irrespective of experiment half (first half of trials: $\beta = -0.13, p < .001$; second half of trials: $\beta = -0.15, p < .001$) and did not vary by participants' final accuracy ($B = -0.30, p = 0.17$), indicating that it was not the product of initial unfamiliarity with the task or based on overall performance.

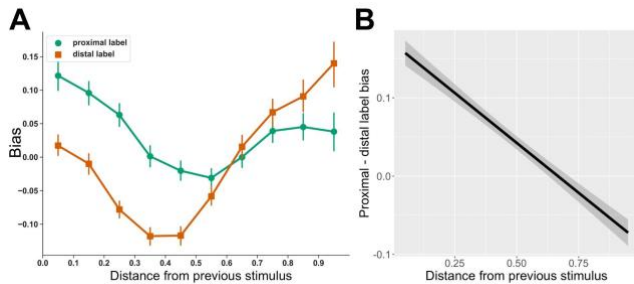


Figure 2: Calculated bias for particular response keys as a function of inter-stimulus distance. **A:** Bias for the “proximal” label (e.g. 2-1) shown in red; bias for the “distal label” (e.g. 2-4) shown in green. **B:** A linear regression fit to the difference between proximal and distal label biases. A negative slope supports our hypothesis that people have a preference to respond with the numerically proximal label (e.g. 2-1) when stimuli are more similar to each other, and vice versa (e.g. 2-4) when they are further apart.

Discussion

Our results showed a clear proximity bias effect. However, because this analysis was opportunistic (using publicly available data), there were some parameters that had not been controlled for in the original design, raising the question of potential confounds. One possibility is that the observed effect was purely a coincidence of how the stimuli are arranged in stimulus space. That is, learning the 4-way categorization task would have produced the behavior we observed regardless of the labels or response keys. Alternatively, an asymmetry in perceptual sensitivity to the stimulus dimensions (e.g. greater acuity to detect changes in

temporal compared to spectral modulation) could lead to a natural, perceptually-driven separation of these categories in the mind that just so happens to align with label proximity. If the observed effect was the result of these confounds, remapping the labels to orthogonalize the proximity structure while keeping the stimulus space constant should produce a reversal of this effect, thus washing it out in the aggregate. We therefore ran Experiment 2 to determine whether the observed proximity effect was an artifact of these aspects of the design.

Experiment 2

Experiment 2 aimed to test whether the same proximity bias observed in Experiment 1 would replicate in a new participant population (original labels condition) and to control for potential confounds from the specific label-to-category mapping used in Experiment 1 by counterbalancing the assignment of labels to regions of the stimulus space (remapped labels condition).

Methods

One hundred and fifty-eight undergraduate psychology students at the University of Connecticut participated in the experiment for course credit. We used the original Gorilla experiment files from Roark and Chandrasekaran (2023) and added a “remapped labels” condition. For this, we orthogonalized the labels’ proximity structure in relation to stimulus space by remapping the original label-to-category assignments (via rotating the labels 90° clockwise in stimulus space, Figure 3A). This change affected only the labels; the stimulus distribution remained the same. Participants learned the categories with labels either in the original configuration (original labels condition) or the remapped version (remapped labels condition).

Results

Accuracy did not differ between label configurations; participants learned both the original ($M = 0.37, SD = 0.09$) and remapped ($M = 0.36, SD = 0.09$) label-stimulus assignments equally well ($t(156) = 0.92, p = .36$). For the bias analysis, a linear mixed effects model revealed significant main effects of inter-stimulus distance ($\beta = -0.14, p < .001$; the aggregate of both slopes in Figure 3C) and of label configuration ($\beta = -0.17, p < .001$; indicating that there was also an effect of label-to-category assignment), as well as a significant interaction between the two ($\beta = 0.20, p < .001$).

Discussion

This experiment yielded a few findings. First, we replicated the original effect from Experiment 1 with a different population. Second, we found that this effect was modulated by the particular configuration of labels mapping onto stimulus space; remapping the labels in stimulus space somewhat reversed the proximity bias, as evidenced by the significant interaction between inter-stimulus distance and label configuration. Despite this, we still observed the

predicted main effect of stimulus distance in the aggregate: Here again, the negative slope indicates that when stimuli were closer together, people showed a preference for responding with the numerically proximal label, but this preference reversed as the stimuli got further apart. Thus, Experiment 2’s results suggest that the earlier findings were not *solely* driven by the specific mapping of labels to categories—it seems that both the specific label-to-category mapping *and* a proximity bias produced the observed effect. To corroborate this interpretation and further test the generalizability of this effect beyond the features of spectral and temporal modulation, we conducted Experiment 3.

Experiment 3

The results from Experiment 2 suggested that certain label-stimulus mappings can produce a proximity bias when categorizing artificial nonspeech auditory ripples. Experiment 3 tested whether the effect was restricted to those stimuli (perhaps due to their specific featural properties) or if it generalizes to other stimulus dimensions and modalities. To this end, we again turned to open data, this time from the second experiment of Roark et al. (2022). The experiment was almost identical to the experiments described above. One key difference, however, was that participants completed both auditory and visual categorization tasks. The auditory stimuli comprised tones of varying pitch and duration, while the visual stimuli consisted of single lines of varying length and orientation. Roark et al. (2022) recruited both musicians and non-musicians as part of their experiment design. In keeping with that distinction, we report it here as an additional measure of “familiarity” with a given modality. The underlying assumption is that musicians may be more perceptually attuned to differences along the auditory dimension than non-musicians.

Methods

Eighty-seven participants (36 musicians) were recruited through Prolific and completed the experiment on Gorilla. Each participant performed the categorization task for four stimulus distributions in total: two each for both the auditory and visual tasks, spaced 1 week apart. Each stimulus distribution consisted of 300 trials, which proceeded in the same manner as Experiments 1 and 2 described above. Analysis was similar to Experiment 1, with the addition of stimulus modality and musician status as predictors in the linear mixed-effects model.

Results

Results again showed a significant main effect of inter-stimulus distance on preference for the proximal label, and in the predicted direction ($\beta = -0.13, p < .001$; Figure 4). There was also a main effect of stimulus modality ($\beta = -0.08, p < .01$) and an interaction between distance and modality ($\beta = 0.07, p < .05$). No other main effects or interactions were statistically significant. Follow-up analyses on each modality separately showed that although the main effect of distance was present for both visual stimuli ($\beta = -0.06, p < .01$) and auditory stimuli ($\beta = -0.13, p < .001$), it was weaker for visual stimuli (hence the main effect of modality).

Discussion

We again found a bias to select numerically proximal labels when stimuli were closer together in stimulus space. This effect was found with auditory tones differing in pitch and duration as well as visual lines varying in length and orientation, indicating the effect in Experiment 1 was not specific to the auditory ripple stimuli. We also found no group differences between musicians and non-musicians—the

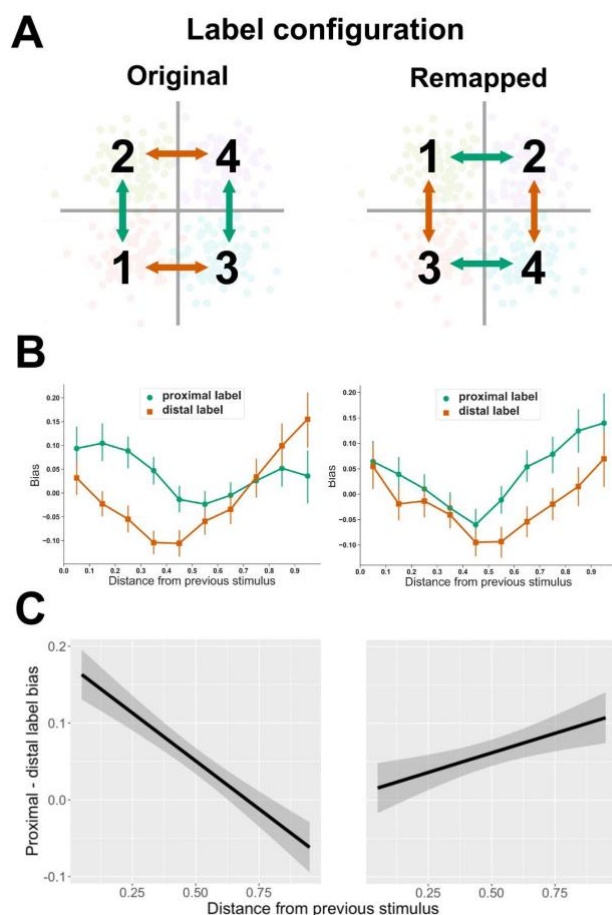


Figure 3: **A:** Visualization of label configurations. The underlying stimulus distribution was exactly the same—labels for the categories were rotated 90° clockwise in space in the “remapped” condition. **B:** Bias as a function of inter-stimulus distance, plotted separately for proximal labels (green) and distal labels (red) and conditions (original: left; rotated: right). **C:** Linear slopes of proximal–distal bias difference as a function of inter-stimulus distance

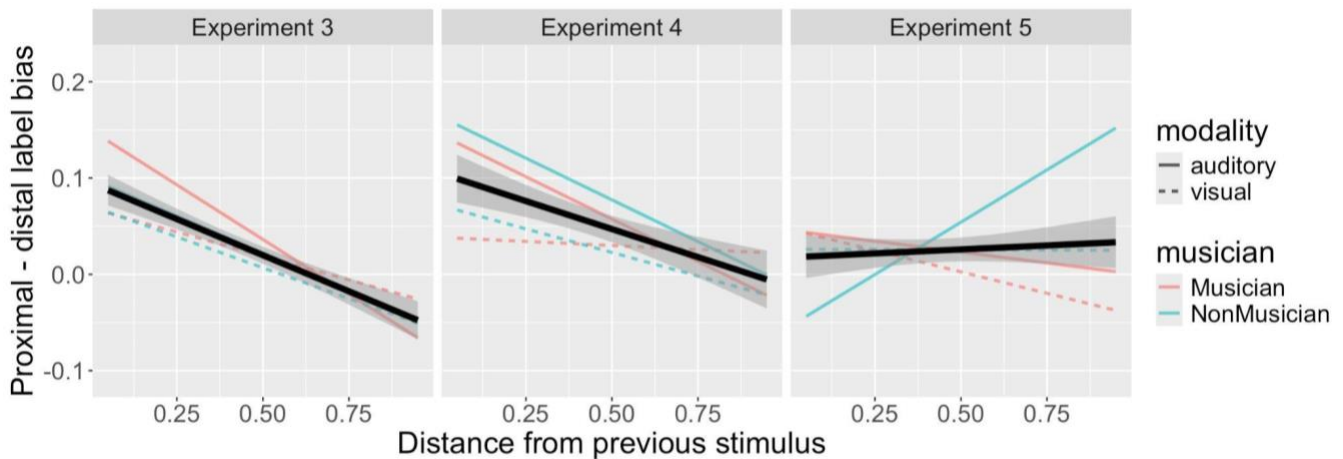


Figure 4: Linear slopes of proximal–distal bias difference as a function of inter-stimulus distance for Experiments 3-5. Slopes are plotted for auditory (solid lines) and visual (dotted lines) stimuli separately for each group (musicians: red, non-musicians: green). The overall mean is plotted in solid black.

bias affected them equally regardless of (presumed) experience with auditory features.

Experiment 4

The findings above suggest participants use trial-to-trial label similarity and perceptual similarity to make categorical judgments in a general sense. If participants are anchoring their responses on a given trial to the previous one (and making their decisions based on similarity rather than absolute feature values), they might be doing so because they do not have a clear idea which dimensions are relevant for categorization. While they may perceive changes in stimulus features (e.g. pitch), they may be neglecting one or the other in their categorization judgments. By chance, proximal labels (e.g. 1-2 and 3-4) could align along the dimension the participants *do* use, leading to the notion that these categories are similar to each other by virtue of them co-occurring along this stimulus dimension. Such an account would frame the proximity bias effect as a byproduct of not knowing the full space of relevant stimulus features. To address this possibility, we analyzed data from the third experiment in Roark et al. (2022), which explicitly informed participants of the relevant stimulus dimensions for categorization.

Methods

Sixty-four participants (28 musicians) took part in this experiment. The set up was nearly identical to that of Experiment 3; the only difference was that participants were explicitly told about the dimensions that would be relevant for categorization (tone pitch and duration for auditory stimuli; line length and orientation for visual stimuli). Note that they were not told where the category boundaries lay along these dimensions, only that the dimensions themselves were relevant for categorization.

Results

Once again, there was a significant main effect of inter-stimulus distance ($\beta = -0.10, p < .01$; Figure 4) and stimulus modality ($\beta = -0.11, p < .01$) and an interaction between the two ($\beta = 0.09, p < .05$). Analyzing the data for both stimulus modalities separately revealed that while there was still a main effect of inter-stimulus distance for auditory stimuli ($\beta = -0.10, p < .01$), the effect was absent for visual stimuli ($\beta = -0.01, p = .74$).

Discussion

Explicit knowledge of the dimensions partially affected the proximity bias. When participants were told in advance of the dimensions along which to base their categorization judgments, they still showed a proximity bias for auditory stimuli. However, they did not show such a bias for visual stimuli. It is possible that they were better able to extract and thus identify the category boundaries in visual feature space when informed about the relevant dimensions, leaving them less susceptible to the proximity bias. We again found no group interaction, suggesting that the auditory effect was not driven by expertise in that domain (or lack thereof).

Experiment 5

Experiment 5 explored the contribution of the response keys in determining the bias. Experiments 1-4 had participants respond using the keys 1/2/3/4 on their keyboard. These keys, at least on a traditional QWERTY keyboard, confound spatial and numeric distance—1 and 2 are close, while 1 and 4 are distant both in space and in magnitude. To test the effect of spatial distance in the absence of numeric magnitude, we examined data from the first experiment of Roark et al. (2022).

Methods

Forty-nine participants (24 musicians) were recruited from the University of Texas at Austin community. Stimuli were the same as Experiments 3 and 4 above, and participants were

informed of the stimulus dimensions along which categories were defined. There were, however, a few changes to the design. Participants completed 600 trials instead of 300. Thus, for direct comparison with Experiments 3 and 4 above, only the first 300 trials are analyzed here. This experiment was also conducted in person instead of online, although we consider this an incidental element of the study that should not affect results.

The principal change for our purposes was that instead of responding with the keys 1/2/3/4, participants used the keys Z/W/M/I (corresponding to the categories 1/2/3/4). This manipulation preserves the spatial proximity structure of response keys in our comparisons (2-1/W-Z and 2-4/W-I are still “near” and “far” respectively on the keyboard) while divorcing it from numeric magnitude, allowing us to assess the contribution of spatial distance independently³.

Results

Contrary to the previous experiments, there was no significant main effect of inter-stimulus distance on preference for the proximal label ($\beta = -0.05, p = 0.33$; Figure 4). There was, however, a significant main effect of group ($\beta = -0.10, p < .01$) and a significant interaction between inter-stimulus distance and group ($\beta = 0.26, p < .001$). Follow-up analyses for each group independently showed no main effects or interactions for the musicians ($\beta_{\text{distance}} = -0.05, p = .28$; $\beta_{\text{modality}} = -0.00, p = .97$; $\beta_{\text{interaction}} = -0.04, p = .46$), but for the non-musicians there were significant positive main and interaction effects of distance and modality ($\beta_{\text{distance}} = 0.22, p < .001$; $\beta_{\text{modality}} = 0.08, p < .05$; $\beta_{\text{interaction}} = -0.22, p < .01$). Non-musicians were more likely to prefer the spatially *distal* response when stimuli were more perceptually similar, but only for auditory stimuli.

Discussion

It is clear that the switch from using Z/W/M/I as response keys to 1/2/3/4 affected the pattern of results and interacted with domain expertise. Why did we find a reversed bias for the auditory stimuli with non-musicians in Experiment 5? One possibility is that spatial distance was not the relevant factor of proximity across the response keys, and participants instead used some other similarity metric. With Z/W/M/I as response keys, participants would (presumably) be using both hands to make their responses (left: Z/W, right: M/I). If so, *motoric* similarity may have driven our observed results—the same finger on both hands might be considered more similar than different fingers on the same hand. Since this analysis was constrained by the limitations of using existing data (and many factors are unaccounted for), we refrain from speculating further. Pinning down the parameters giving rise to a proximity bias will be the focus of future research.

General Discussion

Over the course of five category learning experiments, we show evidence of a bias to make numerically proximal label responses (e.g. 2-1) when stimuli are perceptually more similar to each other and characterize some of the bias’s limits. The effect, which we term a *proximity bias*, can be obtained across different modalities (auditory and visual) and multiple dimensions within modality (pitch and duration vs. spectral and temporal modulation). However, the bias is not always present. It appears to depend partially on the alignment between stimulus and label magnitudes, whether participants are aware of the relevant stimulus features, and possibly also on a conceptual distance between the labels themselves.

The effect, when it is present, applies fairly uniformly across participants and does not appear to vary by the participant’s performance or familiarity with the dimensions involved. This suggests a relatively robust bias. At the very least, we emphasize this as a point of consideration when designing categorization experiments. Crucially, the label similarity bias cannot always be controlled for by counterbalancing, since that is a method which assumes symmetric effects. As we have seen (Experiment 2), a bias may surface in one response configuration and not be nullified by its counterpart, resulting in a still-detectable effect in the aggregate.

What might be the cognitive origins of this effect? When faced with uncertainty about categorical boundaries, participants use heuristics to help them structure their conceptual space and classify stimuli (Murphy & Ross, 1994). We hypothesize that labels may contribute to categorization as an additional feature along which to structure category knowledge. Depending on the quality of information they have about the stimuli (e.g. knowledge about what dimensions to attend to, or the perceptual sensitivity to detect boundaries along those dimensions), similarity among the *labels* may provide an additional dimension to ground their knowledge to, i.e. a metaphorical mapping between domains (Lakoff & Johnson, 1980). The structural product of this is a “second-order isomorphism” between labels and their referents (Shepard & Chipman, 1970), wherein the similarity patterns within the set of labels maps onto the similarity patterns within the stimuli such that across domains, the relational structure is preserved. Similar arguments have been articulated elsewhere in the context of space (Kirsh, 1995), gesture (Goldin-Meadow & Beilock, 2010), and events and action (Hommel et al., 2001). However, further research is needed to determine if indeed the effect we have identified generalizes beyond numeric labels, as well as to fully characterize the limits of when and where it will do so.

³ Switching the response keys to Z/W/M/I does introduce other dimensions of distance (e.g. alphabetical or phonetic), which might be investigated in future studies.

References

- Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham, and Jo K. Evershed. "Gorilla in Our Midst: An Online Behavioral Experiment Builder." *Behavior Research Methods* 52, no. 1 (February 2020): 388–407. <https://doi.org/10.3758/s13428-019-01237-x>.
- Ashby, F Gregory, and W Todd Maddox. "Relations between Prototype, Exemplar, and Decision Bound Models of Categorization." *Journal of Mathematical Psychology* 37, no. 3 (1993): 372–400.
- Ashby, F Gregory, Sarah Queller, and Patricia M Berretty. "On the Dominance of Unidimensional Rules in Unsupervised Categorization." *Perception & Psychophysics* 61, no. 6 (1999): 1178–99.
- Dehaene, Stanislas, Serge Bossini, and Pascal Giraux. "The Mental Representation of Parity and Number Magnitude." *Journal of Experimental Psychology: General* 122, no. 3 (1993): 371.
- Estes, WK. "Classification and Cognition." Oxford University Press, 1994.
- Goldin-Meadow, Susan, and Sian L Beilock. "Action's Influence on Thought: The Case of Gesture." *Perspectives on Psychological Science* 5, no. 6 (2010): 664–74.
- Goldstone, Robert L. "Influences of Categorization on Perceptual Discrimination." *Journal of Experimental Psychology: General* 123, no. 2 (1994): 178–200. <https://doi.org/10.1037/0096-3445.123.2.178>.
- Hommel, Bernhard, Jochen Müsseler, Gisa Aschersleben, and Wolfgang Prinz. "The Theory of Event Coding (TEC): A Framework for Perception and Action Planning." *Behavioral and Brain Sciences* 24, no. 5 (October 2001): 849–78. <https://doi.org/10.1017/S0140525X01000103>.
- Kirsh, David. "The Intelligent Use of Space." *Artificial Intelligence* 73, no. 1–2 (1995): 31–68.
- Lakens, Daniël, Iris K. Schneider, Nils B. Jostmann, and Thomas W. Schubert. "Telling Things Apart: The Distance Between Response Keys Influences Categorization Times." *Psychological Science* 22, no. 7 (July 1, 2011): 887–90. <https://doi.org/10.1177/0956797611412391>.
- Lakoff, George, and Mark Johnson. "The Metaphorical Structure of the Human Conceptual System." *Cognitive Science* 4, no. 2 (1980): 195–208.
- Livingston, K. R., J. K. Andrews, and S. Harnad. "Categorical Perception Effects Induced by Category Learning." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, no. 3 (May 1998): 732–53. <https://doi.org/10.1037//0278-7393.24.3.732>.
- Lupyan, Gary. "Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis." *Frontiers in Psychology* 3 (2012): 54.
- Lupyan, Gary, David H Rakison, and James L McClelland. "Language Is Not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories." *Psychological Science* 18, no. 12 (2007): 1077–83.
- Maddox, W. Todd, and F. Gregory Ashby. "Comparing Decision Bound and Exemplar Models of Categorization." *Perception & Psychophysics* 53, no. 1 (January 1, 1993): 49–70. <https://doi.org/10.3758/BF03211715>.
- Murphy, Gregory L, and Brian H Ross. "Predictions from Uncertain Categorizations." *Cognitive Psychology* 27, no. 2 (1994): 148–93.
- Nosofsky, Robert M. "Similarity Scaling and Cognitive Process Models." *Annual Review of Psychology* 43, no. 1 (1992): 25–53.
- Roark, Casey L., and Bharath Chandrasekaran. "Stable, Flexible, Common, and Distinct Behaviors Support Rule-Based and Information-Integration Category Learning." *Npj Science of Learning* 8, no. 1 (May 13, 2023): 14. <https://doi.org/10.1038/s41539-023-00163-0>.
- Roark, Casey L., Kirsten E. Smayda, and Bharath Chandrasekaran. "Auditory and Visual Category Learning in Musicians and Nonmusicians." *Journal of Experimental Psychology: General* 151, no. 3 (March 2022): 739–48. <https://doi.org/10.1037/xge0001088>.
- Shepard, Roger N, and Susan Chipman. "Second-Order Isomorphism of Internal Representations: Shapes of States." *Cognitive Psychology* 1, no. 1 (1970): 1–17.
- Sloutsky, Vladimir M., and Anna V. Fisher. "Linguistic Labels: Conceptual Markers or Object Features?" *Journal of Experimental Child Psychology* 111, no. 1 (January 2012): 65–86. <https://doi.org/10.1016/j.jecp.2011.07.007>.
- Smith, Linda B, Susan S Jones, and Barbara Landau. "Count Nouns, Adjectives, and Perceptual Properties in Children's Novel Word Interpretations." *Developmental Psychology* 28, no. 2 (1992): 273.
- Waxman, Sandra R., and Dana B. Markow. "Words as Invitations to Form Categories: Evidence from 12-to 13-Month-Old Infants." *Cognitive Psychology* 29, no. 3 (1995): 257–302.
- Yamauchi, T., and A. B. Markman. "Inference Using Categories." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, no. 3 (May 2000): 776–95. <https://doi.org/10.1037//0278-7393.26.3.776>.