Nameability predicts subjective and objective measures of visual similarity

Martin Zettersten (zettersten@wisc.edu)
Department of Psychology, University of Wisconsin-Madison,
1202 W. Johnson Street, Madison, WI 53706 USA

Ellise Suffill (suffill@wisc.edu)
Department of Psychology, University of Wisconsin-Madison,
1202 W. Johnson Street, Madison, WI 53706 USA

Gary Lupyan (lupyan@wisc.edu)
Department of Psychology, University of Wisconsin-Madison,
1202 W. Johnson Street, Madison, WI 53706 USA

Abstract

Do people perceive shapes to be similar based purely on their physical features? Or is visual similarity influenced by top-down knowledge? In the present studies, we demonstrate that top-down information – in the form of verbal labels that people associate with visual stimuli – predicts visual similarity as measured using subjective (Experiment 1) and objective (Experiment 2) tasks. In Experiment 1, shapes that were previously calibrated to be (putatively) perceptually equidistant were more likely to be grouped together if they shared a name. In Experiment 2, more nameable shapes were easier for participants to discriminate from other images, again controlling for their perceptual distance. We discuss what these results mean for constructing visual stimuli spaces that are perceptually uniform and discuss theoretical implications of the fact that perceptual similarity is sensitive to top-down information such as the ease with which an object can be named.

Keywords: visual similarity; nameability; perceptually uniform; top-down processing; language

Introduction

What determines the perceived similarity of two objects? Researchers have sought to measure perceptual similarity for at least two central reasons. First, perceptual similarity has a pervasive influence on cognitive processing, including attention, memory and categorization processes (e.g., Jiang et al., 2016; Kravitz & Behrmann, 2011; Sloutsky, 2003). Second, precisely because perceptual similarity consistently influences cognitive processes, measuring perceptual similarity has crucial methodological importance, since researchers will often seek to control for perceptual similarity (Li et al., 2019). Past efforts have led to the development of stimulus spaces in which perceptual similarity is carefully mapped. One example of a perceptually uniform space is the CIELAB color space (Roberston, 1990). Pairs of equidistant stimuli in this space are (roughly) equally discriminable (Cheung, 2016).

One challenge for these efforts is that recent theories suggest that lower-level representations of similarity and higher-level representations (e.g., category knowledge) mutually influence one another in a dynamic fashion (e.g., Lupyan 2015; Hohwy, 2014). A consequence of these views is that representations of similarity are not stable – instead, they can vary dramatically across tasks and contexts (Çukur et al., 2013). For example, in the context of clouds, the color "grey" is more similar to the color "black" than to the color "white" (since darker clouds index rain); in the context of hair, on the other hand, "grey" and "white" are more similar because grey and white hair color is associated with older age (Roth & Shoben, 1983).

The dynamic nature of representations has consequences for measuring perceptual similarity. In particular, different contexts and tasks are differentially sensitive to top-down representations (such as category knowledge) influencing judgements about lower-level properties (such as visual similarity). Contextual effects on perceptual judgments are ubiquitous in domains such as color perception (Goldstone, 1995). For example, people typically demonstrate an advantage for between-category color discrimination (e.g., for English speakers, discriminating a shade of blue and a shade of green) compared to within-category discrimination (e.g., for English speakers, discriminating two shades of blue), even when the perceptual distance is equated on a perceptually uniform color space. However, when people make many within-category judgements, the between-category advantage largely disappears (Witzel & Gegenfurtner, 2015). Perceptual similarity spaces can be warped by the perceptual context and goal.

One source of top-down influence on perception is language (Lupyan & Clark, 2015). While our perceptual systems deal in *particulars (*every red is a particular shade of red), language deals in *categories* (the word "red" denotes a category). Once learned, a name becomes a powerful cue, warping color representations into a more categorical form. For example, Forder and Lupyan (2019) showed that immediately after hearing a basic color term (e.g., "red", "green"), people were much more accurate in discriminating the named color from nearby colors in an odd-one-out task.

The present studies

In a recent study, Li et al. (2019) developed a perceptually uniform space of shape stimuli – in analogy to CIELAB space

– validated through subjective ratings of similarity between shape items that morph into one another in a continuous perceptual space (Figure 1). Given past research on the sensitivity of perceptual similarity to top-down factors such as verbal encoding, we predicted that viewing more nameable shapes would activate verbal labels, which would in turn warp the perceptual space around the shape. The consequence of this warping is that easier-to-name shapes should become easier to distinguish from their neighbors in shape space, particularly when the neighbors are likely to not share a label. In Experiment 1, we demonstrate that the degree to which visual items share similar names predicts the likelihood of clustering them together, over and above the normed perceptual distance between items. In Experiment 2, we demonstrate that nameability predicts people's visual discrimination: the more nameable a shape is, the easier it is to visually discriminate it from its neighbors.

Experiment 1: Clustering images based on visual similarity

We tested whether participants consistently used nameability to guide their decisions about which visual items were more similar to one another. First, we collected nameability ratings for a validated set of shape stimuli, normed to be equidistant in subjective visual similarity (Li et al., 2019). Then, we instructed a new group of participants to arrange these images based on their similarity. If people use verbal labels to guide their explicit judgments of visual similarity, then the degree to which visual items share similar names should predict their likelihood of being grouped together, over and above what is predicted by the items' visual similarity within the normed perceptual space.

Method

Participants

Naming task. We recruited 120 participants (64 female; 118 native speakers of English) through Amazon Mechanical Turk (mean age: 37.5 years; *SD* = 11.6; range: 20 – 74 years). Participants were paid $0.30 for completing the task.

Sorting task. We recruited 28 (17 female; 27 native speakers of English) undergraduate students at a large Midwestern university, who took part in the study for course credit (mean age: 18.46 years; *SD* = 0.69; 18 – 20 years).

Stimuli

We selected 36 items sampled at 10 degree increments (see Fig 1) along the Validated Circular Shape (VCS) space (Li et al., 2019).

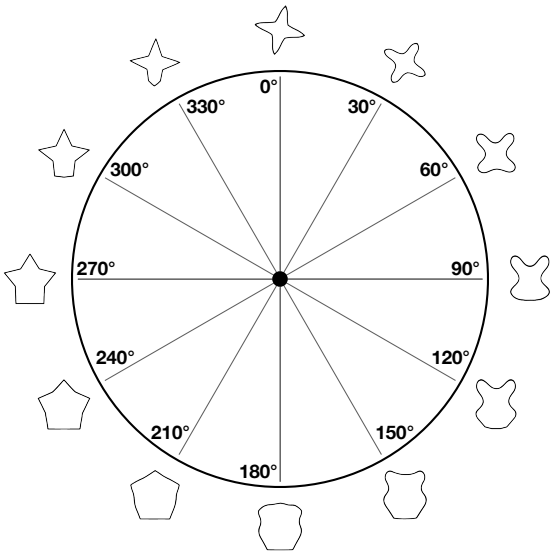

Figure 1: Example stimuli from the Validated Circular Shape (VCS) space (sampled at an angular distance of 30°).

Design & Procedure

Naming Task. The 36 images were grouped into three sets of 12 images equidistant in the validated circular shape space. This was to ensure that participants did not judge items that were extremely close in the validated shape space and to reduce the likelihood that participants would frequently re-use the same label in naming shapes. Participants were randomly assigned to one of the three image sets (e.g., a participant would see the images corresponding to the angles of 10, 40, 70, 100, 130, 160, 190, 220, 250, 280, 310, and 340 in the circular shape space). On each trial, participants viewed one of the 12 images presented in random order. Participants were instructed to name the image in 1-2 words, as quickly as possible. Participants who did not provide an appropriate name for the three familiar images (n = 4) or who did not comply with task instructions (n = 1) were excluded from the analysis, leaving a final sample of n = 115.

Sorting Task. For the sorting task, items were presented around the outer edges of the screen in a randomized order. Participants were instructed to sort the items on the basis of similarity by dragging the items into groups. Once they had finished sorting, they entered the labeling phase, in which they were asked to provide a 1-2 word label for each of the shape clusters. Items could not be moved during the labeling phase. One participant's data was excluded due to a technical error.

Results

The data and analysis code for all results are openly available (https://github.com/mzettersten/vcs).

Relationship between nameability and shape similarity

First, we investigated the degree to which participants' naming judgments were related to the angular distance between visual objects in the Validated Circular Shape space.

Nameability measures. We quantified nameability in two ways. First, we computed the *nameability* of each item in terms of the diversity of participants' responses, calculated using Simpson's diversity index *D* (Simpson, 1949). Simpson's diversity index is sensitive to the frequencies of each word used (Majid et al., 2018). Formally, for a given stimulus, if speakers produce *N* description tokens, including *R* unique description types from *1* to *R*, each with frequencies of n_1 to n_R, then Simpson's diversity index *D* is computed as

$$D = \frac{\sum_{i=1}^{R} n_i(n_i - 1)}{N(N-1)}$$

This measure ranges from 1 – indicating high nameability (all respondents gave the same response type, i.e. $i = 1$ and $n_i = N$) – to 0 – indicating low nameability (all respondents gave unique response types, i.e. $n_i = 1$ for all i).

Second, we considered the extent to which the naming responses provided for each pair of images were related. To compute the degree of *name-based similarity* between two images, we first computed a vector of counts for each unique response provided for each image. We then computed the cosine similarity between the count vectors of each image. Values ranged from 0 (no overlap in the names given to two images) to 1 (perfect overlap in the verbal responses).

Nameability and normed visual similarity. Nameability varied substantially across the shape similarity space (M = .09, 95% CI = [.06, .11], range = [.01, .27]; Figure 2), transitioning between roughly five modal labels (*star*, *blob*, *rabbit*, *vase*, *pentagon*) across the circular similarity space. The shape with the highest name agreement (angle 350) was described as a "star" by 73.0% of respondents, while the shape with the lowest name agreement (angle 100) was described as a "bunny" or "rabbit" by 10.5% of respondents.

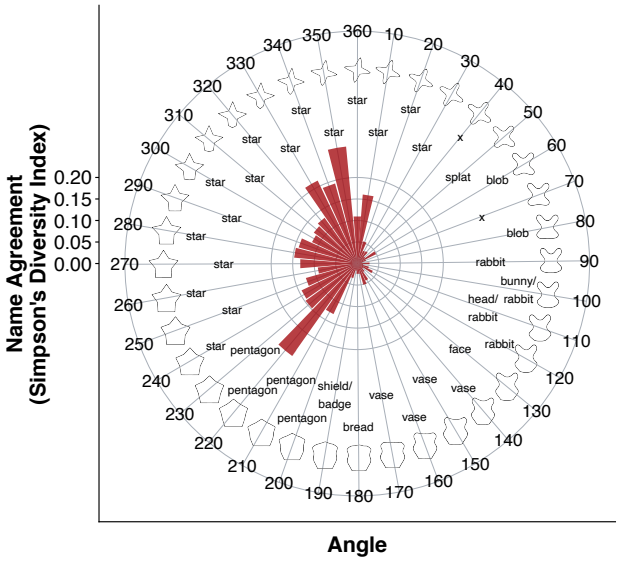

Figure 2: Simpson's diversity index for the VCS shapes. Modal names are presented alongside each shape.

Relationships between nameability measures and clustering

To determine whether name-based similarity predicts how people cluster the shapes according to their perceived similarity, we computed the final coordinates of every item from a participant's sorting solution and used the 'pamk' function ('fpc' package; Hennig, 2019) to identify medoid-based clusters of items (Kaufman & Rousseeuw, 1990). We then computed the average probability that each possible pair of items would be placed within the same cluster. We also recorded each participant's labels for their clusters (n = 25; two participants completed the sorting task, but failed to provide labeling data).

Name-based similarity predicts clustering probability. On average, participants grouped the shapes into 5-6 clusters (M = 5.52, 95% CI = [4.50, 6.53], range = [2, 10]). There was substantial variability in how likely pairs of images were to be clustered together (mean clustering probability = 0.23 ,95% CI = [.21, .25], range = [0,1]). To test whether name-based similarity predicts the likelihood that participants placed images in the same cluster, we fit a general linear model predicting the average probability that two images would be grouped together from the name-based similarity of each image pair (i.e., the cosine similarity of responses for each image in the naming task), while controlling for angular distance between images. Name-based similarity was a strong predictor of the probability that two images would be clustered together, after controlling for angular distance, $b = 0.14$, $t(627) = 5.50$, $p < .001$.

Nameability of individual shapes predicts nameability of clusters. We also investigated whether the nameability of

individual shapes predicted the nameability of the clusters in which the shapes were placed. For the nameability of images based on the clustering task, we assigned each image the label given to its respective cluster by participants in the sorting task and then computed Simpson's diversity index for each image, based on the names of the clusters they belonged to. The name agreement for clusters was generally lower (M = .04, 95% CI = [.03, .05], range = [.01, .09]) than for individual shapes. This is likely due to there being more opportunity for variation in the cluster names (because each cluster contained a variety of shapes with potentially different names), as compared to naming shapes in isolation. Crucially, name agreement (Simpson's diversity index) for individual shapes was highly correlated with the name agreement of clusters that images were sorted into (r = .85, 95% CI = [.73, .92], $t(34)$ = 9.46, p < .001; Figure 3). In other words, independently collected naming norms predicted the consistency of the labels associated with each shape's clusters.

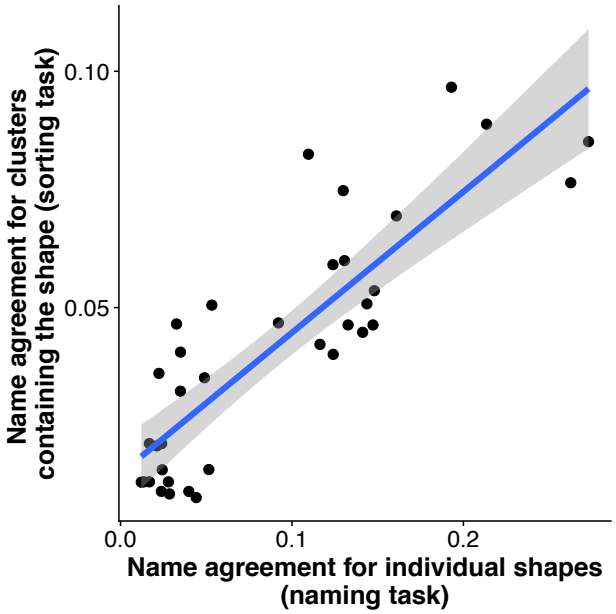

Figure 3: Correlation between Simpson's diversity index for naming an individual image and Simpson's diversity index for the same image's cluster names.

Discussion

In Experiment 1, we investigated whether nameability predicted how participants would cluster shapes based on perceptual similarity within a space of visual images normed to be perceptually uniform. We found three main results. First, the distribution of name agreement across the normed stimulus space was not uniform: some shapes were substantially more nameable than others. Second, the nameability of individual shapes (assessed in a separate naming task) predicted how nameable their clusters were in a sorting task: more nameable shapes were also placed into more nameable groups of images. Most importantly, we found that name-based similarity skewed which images participants grouped together: two images that shared more similar verbal descriptions were more likely to be placed into the same cluster, while controlling for their perceptual distance in the validated circular shape space.

The current results suggest that the nameability of individual images, and the degree to which nameability overlaps between items, can affect people's subjective judgements of which images are more similar to one another. One possibility is that these effects are relatively transitory and only appear when participants are asked to make slow, deliberative decisions in which they incorporate verbal information. However, an alternative possibility is that nameability may skew perceptual similarity even early on in the processing stream. To test this possibility, in Experiment 2, we investigated the effects of nameability in a visual discrimination task designed to provide a more objective measure of perceptual similarity.

Experiment 2: Perceptual discriminability

In Experiment 2, we investigated whether the nameability of VCS shapes would predict the perceptual discriminability of shapes in a speeded match-to-sample task. Response times in speeded visual discrimination tasks are a sensitive measure of perceptual processing of visual images (Lupyan, 2008) and allow us to gain more fine-grained insight into the nature of participants' perceptual representations.

Method

Participants

We recruited 58 (29 female; 51 native speakers of English) new participants from the undergraduate psychology student research pool at a large Midwestern university (Mean age: 18.5 years; *SD* = 1.07; range: 18-24 years). Students participating in the study received course credit. Five additional participants participated but were not included due to having less than 80% useable trials (n = 1 due to a technical error and n = 4 due to a high rate (>20%) of trial exclusions).

Stimuli

The stimuli were the same 36 visual items selected from the Validated Circular Shape space in Experiment 1.

Design

Since testing all pairwise comparisons between items would have led to far too many experimental trials to collect within one participant, we split item-wise comparisons into two counterbalanced, between-subjects conditions. One counterbalancing condition controlled the angular distance between target and foil items on each trial. Participants were randomly assigned to judge items either at a 10 degree, 20 degree, or 30 degree angular distance (manipulated between-subjects). The second counterbalancing condition split the 36 items into two equidistant groups to ensure that we had a sufficient number of trials for each item comparison within each participant (n = 32 per item pair). For example, if participants were in the group judging item pairs that were 10

degrees apart along the shape circle, they would judge one of two mutually exclusive sets of 18 item pairs spaced equidistantly around the shape circle (set 1: 10°-20°, 30°-40°, 50°-60°,…; set 2: 20°-30°, 40°-50°, 60°-70°,…). Participants were randomly assigned to the two counterbalancing conditions.

Participants completed 32 match-to-sample trials for each item pair under speeded conditions. Each of the two items served as the standard – and therefore also the target – on half of these trials. The location of the target was counterbalanced across trials for each participant. Each participant judged 18 equidistant item pairs, leading to a total of 576 test trials per participant.

Procedure

On each speeded match-to-sample trial (Figure 4), participants judged whether a visual stimulus (the standard) matched one of two images (the target and the foil) appearing below the standard. At the onset of each trial, three placeholder boxes appeared on the screen for 500ms to orient the participant to the locations of the three shape images on each triad trial. Then, the standard was presented in the top box for 1s. Next, the target and the foil appeared below the standard in the left and right boxes. Participants were instructed to judge as quickly as possible, without making errors, whether the top shape matched the left or the right shape. Participants used the 'z' (left response) and the '/' (right response) keys on the keyboard to indicate their response. If participants responded incorrectly, a short buzzing sound was played over headphones to provide participants with feedback.

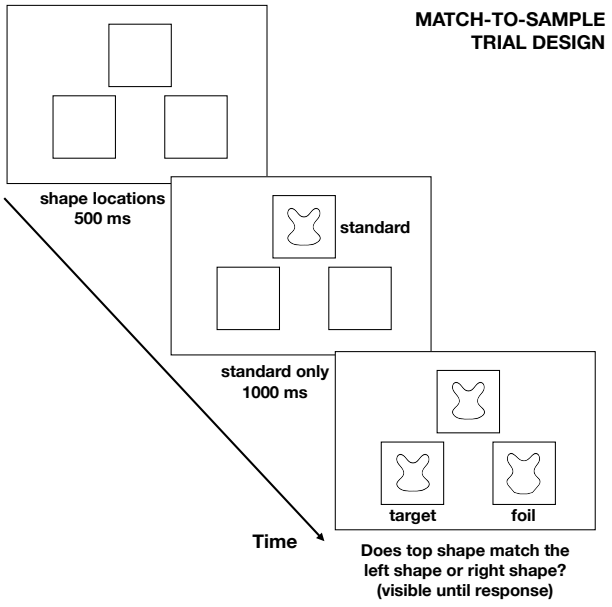

Figure 4: Speeded match-to-sample trial design in Experiment 2.

Results

Overall speeded verification performance

Trials with either very short (<200 ms) or very long (>5000 ms) reaction times were excluded from the analysis (~1.3% of all trials). Overall, participants performed highly accurately at the task while making speeded judgments, with higher accuracy and speed when the angular distance between target and foil shapes was greater (Table 1).

Table 1: Accuracy and Reaction Times in Experiment 2

Distance Condition	Mean Accuracy	Mean RT (correct trials)
10	M = 82.4% 95% CI = [77.9%, 86.8%]	M = 1443ms 95% CI = [1249ms,1638ms]
20	M = 92.9% 95% CI = [91.0%,94.8%]	M = 1094ms 95% CI = [987,1201ms]
30	M = 95.2% 95% CI = [94.0%,96.4%]	M = 808ms 95% CI = [723ms,893ms]

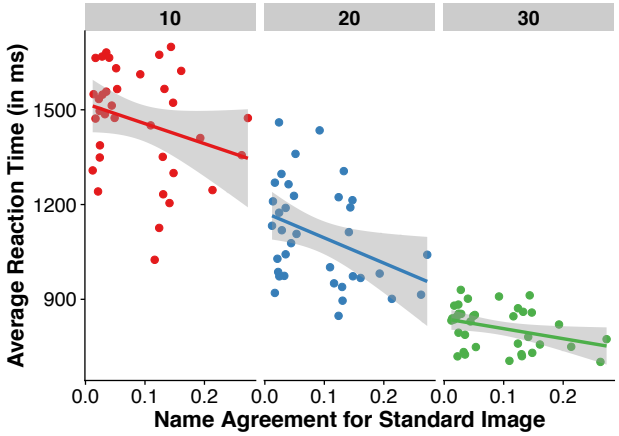

Figure 5: Relationship between name agreement (Simpson's diversity index) for the standard and correct RTs (averaged across subjects). Colors/panels represent angular distances between the standard and foil (smaller values correspond to more visually similar shapes).

Nameability and name-based similarity predict discriminability

We predicted that the verbalizability of visual shapes would affect participants' ability to rapidly identify the shape matching the standard in two ways. First, we predicted that more nameable standard images would help participants more quickly identify the target image. However, the name-based similarity between the target image and the foil image may alter the degree to which the activation of a verbal label is helpful: if both the target and the foil image activate similar verbal labels (e.g., if both images activate the label "star"), this should dampen the benefit of a more nameable standard by increasing the confusability of the target and foil. Thus, our second prediction was that the nameability of the standard image and the name-based similarity between target and foil

image would interact, such that target-foil pairs with higher name-based similarity would lead to slower response times.

To test these predictions, we used the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015; version 1.1-21) to fit a linear mixed-effects model predicting participants' trial-by-trial reaction times (correct trials only; in ms) from the nameability of the standard (Simpson's diversity index; centered), the name-based similarity between the target and foil (centered), and their interaction, while controlling for the angular distance between the target and the foil. We first fit the model with the maximal random effects structure for participants and items, and then iteratively pruned random effects until model convergence was achieved. The final model included a by-participant random intercept and random slope for the nameability of the standard, as well as by-item random intercepts for the target and the foil. We used Satterthwaite's method to estimate degrees of freedom.

Nameability of the standard predicted faster response times on correct trials, b = -596.3, Wald 95% CI = [-1065.7, -126.8], $t(47.6) = -2.49$, $p = .016$ (Figure 5). Moreover, consistent with our second prediction, there was a significant interaction between the nameability of the standard and the name-based similarity of the target and foil images, $b = 688.4$, Wald 95% CI = [98.3, 1274.5], $t(3879) = 2.29$, $p = .022$. As name-based similarity between the target and foil images increased, the boost to reaction times conferred by nameable standard images decreased. There was also a significant overall effect of name-based similarity, $b = 65.8$, Wald 95% CI = [15.7, 115.8], $t(6130) = 2.58$, $p = .01$, such that reaction times were slowed when target and foil images were more similar (controlling for angular distance between the shapes and the nameability of the standard).

Discussion

Across varied tasks, nameability predicted visual similarity. In a shape space specifically constructed to create perceptually equidistant shapes, nameability predicted how people clustered items based on similarity. Items that were more likely to be given similar verbal descriptions were more likely to be grouped together in similarity space, over and above what one would predict based on shapes' normed similarity distance. The effect of nameability extended beyond explicit clustering judgments into more objective measures of perceptual similarity. When participants were asked to discriminate images in a speeded match-to-sample visual discrimination task, a more nameable standard led participants to more quickly identify the target, again controlling for the normed perceptual similarity distance between shapes.

Why were these nameability effects not already "baked into" the VCS norms? Li et al. (2019) relied on iterative norming using odd-one-out tasks to generate shapes that are perceptually equidistant. Had we used their exact tasks, we have every reason to think that we would replicate their results. But altering the tasks even subtly led to very different estimates of which shapes are similar. Specifically, the instruction in Experiment 1 to group shapes led people to produce similar clusters. Although Experiment 2 used a triad task similar to that used by Li et al. (2019), the inclusion of a delay between the presentation of the standard and target/foil (Fig 4) may have led to an automatic, verbally aided categorization of the standard (as in Lupyan et al., 2010). If true, this would predict that decreasing the delay would reduce the effect of nameability. Relatedly, it is possible that the benefits of verbally supported categorization were amplified due to the speeded nature of the match-to-sample task. For example, past work suggests that verbal labels may facilitate deployment of attention to named objects early in visual processing (e.g., Lupyan & Spivey, 2010).

Nameability vs. alternative explanations. The current studies provide evidence that a shape space constructed to be perceptually uniform nevertheless exhibits consistent similarity structure, and that this similarity structure is related to the verbalizability of individual shapes. Our preferred explanation of why nameability and visual similarity go hand-in-hand is that labels play a causal role in visual processing, by guiding categorization judgements (Sloutsky, Lo, & Fisher, 2001; Perry & Lupyan, 2014) and shaping perceptual expectations (Lupyan & Clark, 2015; Samaha et al., 2018). However, the current findings do not rule out alternative non-linguistic explanations. For example, the top-down influence of non-linguistic category-based representations may simultaneously explain differences in how images are perceived and in how easy images are to name. One way to disentangle these alternative explanations would be to test differences in perceptual similarity across languages that differ in how they verbally encode a variety of shape sets: if language plays a causal role in visual similarity, then perceptual processing should vary with how verbalizable a given object is in each language. Our results can also be further generalized by applying our methods to new stimuli.

Measuring perceptual similarity. Our work emphasizes that visual similarity requires taking into account both the task used to make the measurement and nonvisual factors such as nameability and name-based similarity. A space that appears perceptually uniform as assessed in one task, e.g., using untimed subjective ratings, can become non-uniform as assessed through a different task such as speeded discrimination. For a given task, the similarity of two stimuli is influenced not only by perceptual factors, but also by how the stimuli relate to previously learned named categories. Researchers should always be cautious to use measurements of similarity that will appropriately generalize to the conditions under which participants are tested in their design and are meaningful for the inferences they require.

Why is it so hard to measure visual similarity that is independent of non-visual information? One may be justifiably puzzled as to why non-visual information – such as how easy it is to name a shape – should have such a consistent impact on visual similarity. Why is it so difficult to obtain an "objective" index of perceptual similarity that remains consistent across testing conditions? While a fully satisfying answer to this question lies outside the scope of the

current paper, we believe the persistent influence of top-down information – even early on in perceptual processing (see e.g., Samaha et al., 2018) – is likely connected to the fundamental goals of perception. Rather than encoding perceptual features in an objective fashion, the goal of perceptual processing is to generate representations that are useful to the perceiver (Hoffman, 2016). As David Marr (1982) put it, "vision is a process that produces from images of the external world a description that is useful to the viewer" (p. 31). Incorporating top-down information when computing similarity is broadly useful, and verbal information may gain particular weight given the pervasive need to talk and communicate about what we see. When judging whether two shapes are similar, what we might call them cannot help but come to mind.

Acknowledgements

This research was supported by NSF-GRFP DGE-1747503 awarded to MZ and NSF BCS 1734260 to GL.

References

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Cheung, V. (2016). Uniform color spaces. In J. Chen, W. Cranton, & M. Fihn (Eds.), *Handbook of visual display technology* (pp. 187–196). Heidelberg, Germany: Springer.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770.

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General, 148*(7), 1105–1123.

Goldstone, R. L. (1995). Effects of Categorization on Color-Perception. *Psychological Science*, *6*(5), 298–304.

Hennig, C. (2019). fpc: Flexible Procedures for Clustering. https://CRAN.R-project.org/package=fpc

Hohwy, J. (2014). *The predictive mind*. Oxford: Oxford University Press.

Hoffman, D. D. (2016). The interface theory of perception. *Current Directions in Psychological Science*, *25*(3), 157–161.

Jiang, Y. V., Lee, H. J., Asaad, A., & Remington, R. (2016). Similarity effects in visual working memory. *Psychonomic Bulletin and Review, 23*(2), 476–482.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. Hoboken, NJ: John Wiley & Sons.

Kravitz, D. J., & Behrmann, M. (2011). Space-, object-, and feature-based attention interact to organize visual scenes. *Attention, Perception, and Psychophysics, 73*(8), 2434–2447.

Li, A. Y., Liang, J. C., Lee, A. C., & Barense, M. D. (2019). The validated circular shape space: Quantifying the visual similarity of shape. *Journal of Experimental Psychology: General*.

Lupyan, G. (2008). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369.

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

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

Lupyan, G., & Spivey, M. J. (2010). Redundant labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics, 72*(8), 2236–2253.

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science, 21*(5), 682–691.

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., … Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376.

Marr, D. (1982). *Vision: A computational approach*. San Francisco: Freeman & Co.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain & Language, 135*, 66–72.

R Development Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Robertson, A. R. (1990). Historical development of CIE recommended color difference equations. *Color Research & Application, 15*(3), 167–170.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, *15*(3), 346–378.

Samaha, J., Boutonnet, B., Postle, B. R., & Lupyan, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. *Scientific Reports*, *8*(1), 1–14.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(1943), 688.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*(6), 246–251.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72*(6), 1695–1709.

Witzel, C., & Gegenfurtner, K. R. (2015). Categorical facilitation with equally discriminable colors. *Journal of Vision, 15*, 22.