

USING LANGUAGE TO DISCOVER CATEGORIES: MORE NAMEABLE FEATURES IMPROVE CATEGORY LEARNING

MARTIN ZETTERSTEN^{*1} and GARY LUPYAN¹

^{*}Corresponding Author: zettersten@wisc.edu

¹Psychology Department, University of Wisconsin-Madison, Madison, USA

Does the lexicon of a language have consequences for cognition? Here, we provide evidence that the ease with which category features can be named can influence category learning. Across two experiments, participants learned to distinguish images composed of colors (Experiment 1) and shapes (Experiment 2) that were either easy or more difficult to name in English. Holding the category structure constant, when the underlying features of the category were easy to name, participants were faster and more accurate in learning the novel category. We argue that these findings suggest that labels allow learners to form more compact hypotheses, which in turn can be confirmed or disconfirmed in the course of learning. These results have consequences for considering how cross-linguistic differences in lexical inventory affect how readily novel categories are learned.

1. Introduction

Categorization is central to cognition (Harnad, 1990; Murphy, 2002). Learning categories requires recognizing which features or dimensions of stimuli are relevant and which are not. For example, color is a relevant dimension for categorizing foods because many foods have characteristic colors. In contrast, color is largely irrelevant for categorizing vehicles: knowing that an object is brown does not help in categorizing it as a car. But how do we discover which features and dimensions are relevant to category membership? One common proposal is that learners track how strongly different features are associated with a category (Rosch & Mervis, 1975). But how do learners identify candidate features in the first place (see e.g., Schyns, Goldstone, & Thibaut, 1998 for discussion)?

Here, we ask whether part of the answer lies in language (Lupyan, 2012). Might the vocabulary of a language provide the learner with a powerful set of candidate features (priors) to use when learning new categories? In particular, we test the hypothesis that the ease with which features can be named influences the likelihood that they are considered as relevant for category membership. For example, on encountering a novel stimulus, learners can more easily form a verbal description of some candidate features (e.g., “red” as a description of a color or “tree” as a description of a complex shape) compared to others (e.g., “greenish-yellow-brown mix” or “nonsymmetrical spacecraft looking thing”). Features that are more nameable may be represented in a more stable way, making it more likely to provide a hypothesis about category membership (e.g., “items with the color red belong in the same category”) that can be tested during category learning. If language plays a role in our ability to form hypotheses about novel categories, this has important implications for understanding the consequences of cross-linguistic differences: when a language has a readily

accessible name for a feature or dimension, it makes that feature/dimension easier to represent during category learning.

In the current experiment, we compare participants' ability to learn novel categories when category exemplars were composed of more nameable or less nameable color features (Experiment 1) or shape features (Experiment 2). Both category types had similar logical structure: the categories were structured such that one color feature or one shape feature always predicted category membership (e.g., images containing red belonged to category A while those containing brown belonged to category B). We hypothesized that categories would be easier to learn when the underlying color and shape features were easier to name, and therefore could be more readily formulated as a hypothesis about category membership.

2. Experiment 1: Color-based category features

2.1. Participants

We recruited 201 participants through Amazon Mechanical Turk. Participants were randomly assigned to the High Nameability Condition ($n = 101$) or to the Low Nameability Condition ($n = 100$) and were paid \$0.60 - \$0.75 for completing the task, which lasted approximately 4 minutes.

2.2. Stimuli

The exemplars were circles ("color wheels") composed of 3 different colors (see Figure 1A). Following the design of (Couchman, Coutinho, & Smith, 2010), one of the colors was perfectly predictive of category membership. The other two were correlated at 66.6% with category membership. The critical manipulation involved the nameability (Guest & Laar, 2002) of the colors comprising each color wheel exemplar. To assess nameability we used the results of a large-scale online color naming study ($N=134,727$, Munroe, 2010), restricting the analysis to RGB values named by ≥ 100 individuals. To ensure that the results could not be affected by differences in color discriminability between the low- and high-nameability conditions, we selected the colors such that all the pairwise CIE-LAB distances were equated using ΔE_{2000} (Sharma et al., 2005). We constructed image sets with different color pairs in the critical position to ensure that any effect was not due to idiosyncratic properties of colors in the category-diagnostic position. The critical colors in the more-nameable categories were "brown" RGB=(120, 80, 40) and "red" RGB=(220, 20, 0), "blue" RGB=(30, 90, 210) and "orange" RGB=(250, 120, 30), or "blue" and "brown". These colors were named according to their modal label by 80% - 85% of the population. The critical colors in the low-nameability were RGB=(170,160,40) (modal name "mustard") and RGB=(200,170,170)

(modal name “lavender”) or RGB=(200, 100, 70) (modal name “brown”) and RGB=(70, 100, 90) (modal name “grey”). These names were only used by 6% - 10% of participants in the original naming task. One drawback of the current approach is that ΔE requires a device-independent color space while our participants viewed the colors on their (often uncalibrated) monitors. However, while this introduces noise, there is no indication that this variability in viewing conditions favorably affects the high-nameability colors compared to the low-nameability colors.

2.3. Task

The participants were tasked with learning to place the color wheels into one of two categories by dragging it into the appropriate box, labeled as box “A” and box “B” (Figure 2). Participants completed a total of 24 training trials, split into 3 blocks. On each block, participants sorted the prototype exemplar (the top image in Figure 1A) for each category twice, and the remaining two exemplars of each category once. Participants received immediate feedback on whether their choice was correct or incorrect. Trials were repeated after an incorrect response. Box locations were counterbalanced across participants.

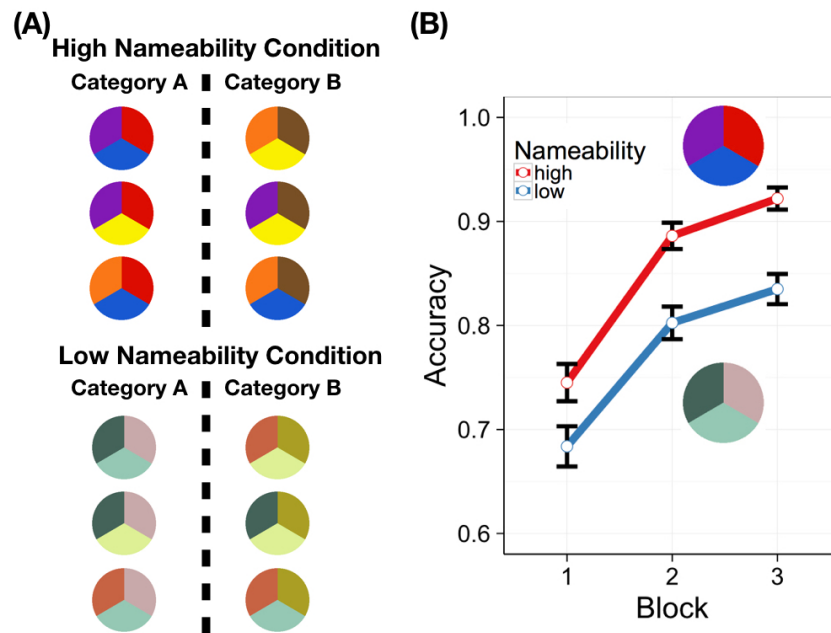


Figure 1. (A) Example stimuli for the high and low nameability condition and (B) Accuracy on categorization across blocks depending on nameability condition in Experiment 1 (color-based features).

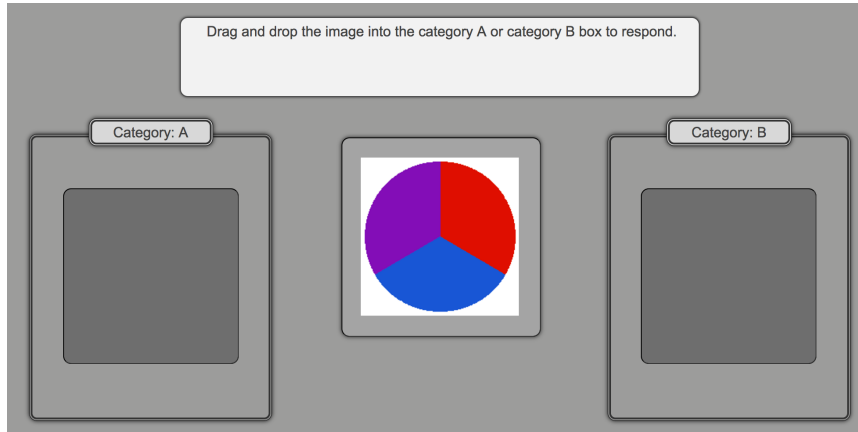


Figure 2. Task setup for category learning experiment

2.4. Results

Training results are shown in Figure 1B. We tested the effects of condition (low nameability vs. high nameability), block, and their interaction on participants' accuracy using a logistic mixed-effects model. Variables were centered and the model included a by-subject random intercept and random slope for block.

Accuracy increased over the course of training, $b = .98$, Wald 95% CI = [.81, 1.16], $z = 11.03$, $p < .0001$. Overall accuracy was higher in the high nameability condition ($M = 85.1\%$, 95% CI = [82.7%, 87.5%]) compared to the low nameability condition ($M = 77.4\%$, 95% CI = [74.3%, 80.4%]), $b = 0.74$, Wald 95% CI = [.39, 1.09], $z = 4.13$, $p < .0001$. There was significant interaction between block and condition, $b = .43$, Wald 95% CI = [.14, .71], $z = 2.89$, $p = .003$, indicating that participants learned the categories more quickly in the high nameability condition than in the low nameability condition.

3. Experiment 2: Shape-based category features

3.1. Participants

We recruited 120 participants through Amazon Mechanical Turk. Participants were randomly assigned to the High Nameability Condition ($n = 58$) or to the Low Nameability Condition ($n = 62$) and were paid \$0.90 for completing the task, which lasted 8 minutes on average.

3.2. Stimuli

The exemplars were circles similar to Experiment 1 composed of 2 different shapes (see Figure 3A). The shapes were chosen from a previous study on

the nameability of complex polygon shapes (Vanderplas & Garvin, 1959). To ensure that shapes used in the experiment were equally discriminable, we conducted a separate norming task in which participants ($n = 24$) performed a speeded same/different task—an extremely sensitive method for measuring represented visual similarity (Lupyan, 2008). We collected these visual discriminability data for the eight most nameable and the eight least nameable shapes from Vanderplas & Garvin (1959)’s ratings. The high nameability shapes were slightly more discriminable than the low nameability shapes, making it impossible to find enough shapes to recreate the same feature structure as in Experiment 1 while still matching the two conditions on shape discriminability. We therefore selected four highly nameable shapes and four shapes with low nameability such that the pairwise discriminability (as measured by reaction times in the norming task) was matched between the two shape sets. The modal names for the highly nameable shapes were “bear”, “tree”, “frog”, and “swan” (left to right in Fig. 3A). The modal names for the low nameability shapes were “mountain”, “rabbit”, “bird” and “hook”. As in Experiment 1, the two categories were defined by the presence of a single critical shape. To increase the difficulty of the task and increase the variability of the stimuli, we randomized the location of the shapes, such that the critical image appeared in all three “slice” locations rather than in a fixed location, as in Experiment 1.

3.3. Task

The category learning task was identical to Experiment 1. Participants completed 24 training trials split into 3 blocks.

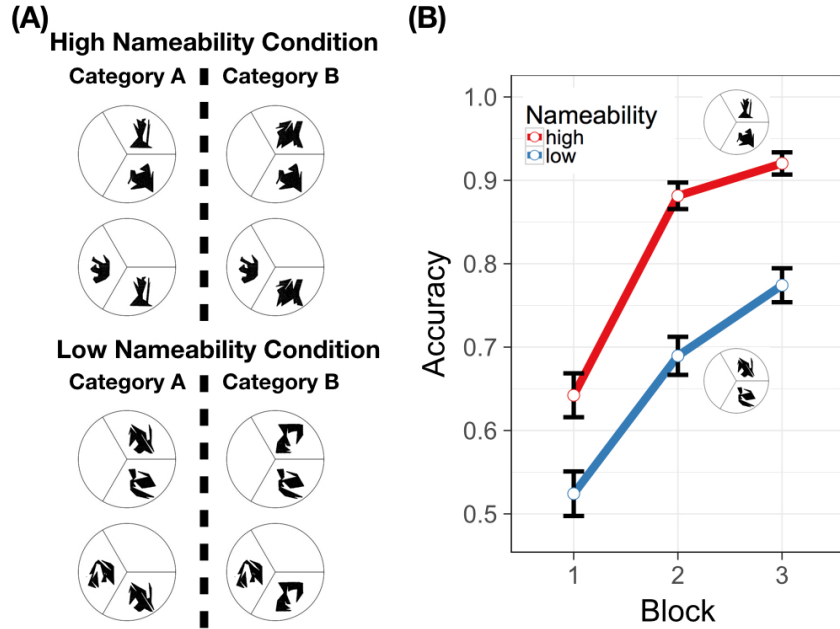


Figure 3. (A) Example stimuli for the high and low nameability condition and (B) Accuracy on categorization across blocks depending on nameability condition in Experiment 2 (shape-based features).

3.4. Results

To analyze the training results (see Figure 3B), we fit the same model as in Experiment 1. Accuracy increased over training blocks, $b = 1.45$, Wald 95% CI = [1.17, 1.73], $z = 10.09$, $p < .0001$. As in Experiment 1, overall accuracy was higher in the high nameability condition ($M = 81.5\%$, 95% CI = [77.7%, 85.6%]) compared to the low nameability condition ($M = 66.3\%$, 95% CI = [61.7%, 70.8%]), $b = 1.45$, Wald 95% CI = [.89, 2.01], $z = 5.04$, $p < .0001$. We also found a significant interaction between block and condition, $b = .91$, Wald 95% CI = [.42, 1.39], $z = 3.67$, $p < .001$, indicating faster increases in accuracy in the high nameability condition compared to the low nameability condition.

4. Discussion

Categories defined by a more nameable color or shape were learned substantially faster than categories with an identical structure but whose defining feature was less nameable. This result supports the idea that, at least for learning simple categories with diagnostic features, more easily formed verbal descriptions may make it easier for participants to formulate and test hypotheses about which features define a category. This result supports and extends

previous work showing that controlling for categorization experience, named categories are easier to learn than unnamed categories (Lupyan, Rakison, & McClelland, 2007).

An important question is whether labels truly are the causal force driving the difference in categorization accuracy. One potential worry is that more nameable colors and shapes are more nameable *because* they are somehow cognitively simpler, rather than because labels aid learners in representing them in the context of category learning. One way to address this worry is to teach labels for the low-nameability categories and test the specific efficacy of such label-based training. Another way is to conduct cross-linguistic experiments: to the extent that language does the driving, categorization performance of e.g., Mandarin speakers, ought to be better predicted by shape/color nameability of Mandarin rather than English. We are currently using both approaches to address the question of causality.

Another worry is that nameability is confounded with familiarity. Perhaps more nameable items are more familiar and it is familiarity rather than language that is causing the difference between conditions. This is possible in the case of shapes (though note that the *specific* shapes were equally novel). We are unsure how a familiarity-based explanation would work in the case of color, however. Unsaturated colors are considerably harder to name than saturated colors. However, saturated colors span only a small region of color space, and most of our experiences involve less than fully saturated colors, making less saturated colors, if anything, *more* perceptually familiar.

The current findings have important implications for thinking about why and in what way cross-linguistic differences in vocabulary may impact cognition. When one language provides a compact label for a feature that another language does not, it provides a verbal encoding that allows learners to more readily represent this feature. A more easily represented feature can then be entertained as a potential high-level hypothesis about the structure of a novel category. This may help to explain the role played by language in performance across various domains, including representing exact numerosity (Frank, Everett, Fedorenko, & Gibson, 2008; Gordon, 2004) and relational reasoning (Christie & Gentner, 2014; Gentner, Ozyürek, Gürcanli, & Goldin-Meadow, 2013). As languages develop, they also change the verbal repertoire available to their users. These changes in verbal repertoire not only have consequences for communication, they also have consequences for forming novel categories.

Acknowledgements

This material is based upon work supported by NSF-PAC 1331293 to GL and NSF-GRFP DGE-1256259 to MZ. We thank Lynn K. Perry for developing central ideas underlying this study.

References

- Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, 38(2), 383–397.
- Couchman, J. J., Coutinho, M. V. C., & Smith, J. D. (2010). Rules and resemblance: Their changing balance in the category learning of humans (*Homo sapiens*) and monkeys (*macaca mulatta*). *Journal of Experimental Psychology. Animal Behavior Processes*, 36(2), 172–183.
- Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, 108(3), 819–24.
- Gentner, D., Ozyürek, A., Gürcanli, O., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition*, 127(3), 318–30.
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, 306(5695), 496–9.
- Guest, S., & Laar, D. V. (2002). The effect of name category and discriminability on the search characteristics of colour sets. *Perception*, 31(4), 445–461.
- Harnad, S. (Ed.). (1990). *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.
- Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, 108(2), 566–577.
- Lupyan, G. (2012). What do words do? Toward a theory of language-augmented thought. In Brian H. Ross (Ed.), *Psychology of Learning and Motivation* (Volume 57, pp. 255–297). Academic Press.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1082.
- Munroe, R. P. (2010, May 4). Color survey results. Retrieved January 30, 2016, from <http://blog.xkcd.com/2010/05/03/color-survey-results/>
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605.
- Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *The Behavioral and Brain Sciences*, 21(1), 1–17–54.

- Sharma, G., Wu, W. C., & Daa, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*, 30(1), 21–30.
- Vanderplas, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology*, 57(3), 147–154.