1 Finding Contrast Effects in Prior Work

Although existing accounts of category generation broadly overlook the role of category contrast in determining what is novel versus familiar, it was implicitly assumed that learners in previous experiments were *successful* in creating new categories. Thus, if contrast plays a robust role in category genration, its effects should be discernable within the experimental results of these studies. To provide a test of the influence of category contrast within existing data, we conducted a novel analysis of Experiment 3 from Jern and Kemp (2013).[1]

Participants in their experiment were exposed to members of two experimenter-defined categories of 'crystals' varying in hue, saturation, and size. Each category possessed a unique hue, but varied in saturation and size: In the 'Positive' condition, there was a positive correlation between these features (i.e., larger sized crystals were more saturated), and in the 'Negative' condition this relation was reversed. In the 'Neutral' condition, there was no correlation between saturation and size. After learning about the categories from each condition, participants were asked to generate six exemplars belonging to a novel class. As noted above, Jern and Kemp (2013) found that the generated categories tended to follow the distributional properties of the experimenter-defined categories: Generated categories were tightly distributed along the hue feature, and possessed the same saturation-size correlations as in the learned categories. Jern and Kemp (2013), however, did not analyze or discuss how the generated categories *differed* from the experimenter-defined categories.

Because each experimenter-defined category in the Jern and Kemp (2013) experiment possessed a distinct hue shared by all members of the category, it is sensible that participants might generate a category with a hue distinct from the experimenter-provided categories. If category generation were influenced by category contrast in this way, the hues of generated categories should be systematically different from those of the experimenter-defined categories. Unfortunately, stimulus hue was encoded and presented in the Hue-Saturation-Value

[1] Although Jern and Kemp (2013) reported four experiments, we focus on Experiment 3 because their first two experiments tested generation of items into known categories, and their fourth experiment was identical to Experiment 3 but with a restricted generation space.

(HSV) color space, which is device-dependent and not perceptually normed such that perceived color similarity corresponds to proximity in the color space (as opposed to a color space such as CIELAB that is device-independent and equidistant sets of points correspond to pairs of colors that have the same perceptual similarity; Wyszecki & Stiles, 1967). Further, they did not calibrate their monitor, and so we cannot know the precise colors presented to participants. As Jern and Kemp (2013) were interested in relations between the saturation and length of examples in generated into novel categories, these issues do not undercut their analyses and results. However, these issues pose a significant challenge to evaluating contrast between the experimenter-defined and participant-generated categories along the hue dimension. It is plausible that two uncalibrated monitors could display the same HSV color and the colors be perceived in different color categories (especially for color boundaries that vary over lightness, such as the yellow-brown boundary).

Although we cannot know the precise colors that were displayed or perceived, we can still analyze their results from a coarse perspective to see whether there is preliminary support for contrast. To do so, we binned all possible hues into one of eight uniformly-spaced color groups: $\{$*Red*: $0 - 0.063$, $0.938 - 1$, *Yellow*: $0.063 - 0.188$, *Yellow-Green*: $0.188 - 0.313$, *Green-Teal*: $0.313 - 0.438$, *Teal*: $0.438 - 0.563$, *Teal-Blue*: $0.563 - 0.688$, *Purple*: $0.688 - 0.813$, *Pink*: $0.813 - 0.938\}$. In the Jern and Kemp (2013) experiment, the hue of each experimenter-defined category was selected from one of six possible values, each of which falls into one of the color groups above (two color groups were not used as a possible hue for the experimenter-defined categories). By categorizing the participant-generated crystals likewise, we can obtain a broad measure of category contrast by determining the proportion of participant-generated crystals that fall into the same groups as the experimenter-defined categories: If contrast influences the hues of the generated categories, we should observe minimal overlap between in the color groupings.

These data, shown in Figure 1, reveal a clear pattern: The majority of participants in each condition ($n = 22$) generated categories possessing entirely distinct hues; with 0/6

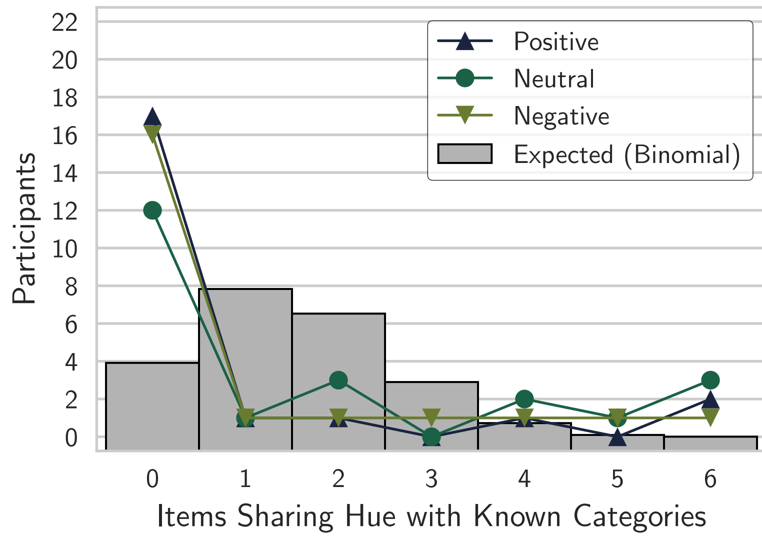

Figure 1: Analysis of data from Jern and Kemp (2013), Experiment 3. Plotted is the number of generated items that share a color group with one of the experimenter-defined classes. The "Expected" data follows a Binomial distribution with $p = 2/8 = 1/4$, given there were two experimenter-defined classes, and eight color groups.

exemplars sharing a hue with the experimenter-defined categories. These results can be compared to the predictions of a Binomial model, which proposes that participants generate hues at random. That is, if hue selection is not systematic, the probability that any given example will lie in the same color group as an experimenter-defined category is given by a Binomial distribution with $p = 2/8 = 1/4$, as there were two experimenter-defined categories and eight possible color groups. Chi-square goodness-of-fit tests reveal that the observed distribution in each condition is highly inconsistent with the hues being chosen at random, all $\chi^2(6, N = 22) > 200$, $p < .001$. Participants tended to generate items that were perceptually distinct from the categories they had learned, and were less likely to generate hues possessed by members of the experimenter-defined categories.

It is possible, however, that this data can be explained by a process that does not involve contrast. Specifically, since participants were trained on exemplars that share the same hue within a category, it is plausible that they generate exemplars that all share a common hue. If this were the case, given a $6/8 = .75$ probability that a hue distinct from the experimenter-defined categories is selected, there is a corresponding .75 probability (or about 17 out of

22 participants) that the generated category will have no exemplars that share a hue with any of the experimenter-defined categories. In addition, that there is a $2/8 = .25$ probability (about 6 out of 22 participants) that the generated category will have all six exemplars that share a hue with any of the experimenter-defined categories. This prediction of regularity across categories is much closer to the data than the prediction from a Binomial model and does not require any mechanism of category contrast.

However, upon analyzing the consistency of hues in the generated categories, it is clear that participants do not generally produce categories with entirely similar hues across all exemplars. Specifically, only 55% of all generated categories comprised exemplars with a common hue. Consequently, the assumption of regularity across categories may be inappropriate for this data set, leaving the contrast-based explanation as the most plausible for our purposes.

Re-analyzing the results from Jern and Kemp (2013) provides some corroborating support for contrast playing a role in category generation. Taken alongside the analyses reported by Jern and Kemp (2013), our analysis suggests that generated categories tend to be distinct from *and* distributionally similar to what is already known. However, it is worth noting that our analysis is still limited: The color groups defined above are imprecise, and it is not clear that our color grouping is consistent with the psychological color boundaries perceived by participants. While we did obtain similar results using a variety of alternative groupings, the hue dimension used in the Jern and Kemp (2013) study does not lend itself straightforwardly to the computation of similarities, and thus we cannot be certain of whether our coding accurately approximates the psychological space of the stimuli. This precludes traditional applications of categorization models to their data as it is usually necessary to encode objects in psychological space in order to accurately determine the similarity between objects. By consequence, although these results likely indicate that contrast exerts *some* influence, they do not precisely describe the nature of that influence.

References

Jern, A., & Kemp, C. (2013). A probabilistic account of exemplar and category generation. *Cognitive Psychology*, *66*(1), 85–125.

Wyszecki, G., & Stiles, W. S. (1967). *Color science*. New York: John Wiley.