Weighing common and distinctive features in a free categorization task

John P. Clapper, Benjamin A. Miller, & Gregory J. Smith

CALIFORNIA STATE UNIVERSITY SAN BERNARDINO

Department of Psychology, California State University, San Bernardino

INTRODUCTION

The categorization of objects in the environment is one of the most fundamental human cognitive abilities. Much of this categorization occurs without external guidance or feedback, a situation often referred to as unsupervised learning, category construction or free categorization.

Most previous research on free categorization has used tasks in which participants were presented with a set of stimuli and asked to sort them into two groups[1]. However, these kinds of tasks limit the number of categories created and in some cases may force participants to place stimuli into "unnatural" groups they would not otherwise have chosen.

In order to facilitate the construction of natural categories, a new free categorization task in which participants generate labels for stimuli is used in the current study[2]. Previous research utilizing this task has found that people show a clear sensitivity to overall alignability even when the objects being grouped together share no actual surface features[3].

This research also indicated that people are sensitive to instances with matching surface features, and likely to group objects that share specific features into the same subcategory[3].

OVERVIEW OF THE EXPERIMENT

The current study further explores people's sensitivity to individual matching features in constructing novel categories by systematically manipulating the number of matching features shared between pairs of stimuli in a display. The degree of match was varied from one to four (out of four) matching features in order to assess how the probability of placing stimuli in the same category would be affected by how many surface features (specific parts) they shared. The resulting sensitivity function should provide useful information for evaluating models of learning and category goodness.

EXPERIMENT 1 METHODS

Participants viewed a single 4x4 array of novel visual stimuli and were asked to create a binomial (letter-number) compound label for each to categorize them at "family" and "species" levels, e.g., A1, B1, B2, etc.

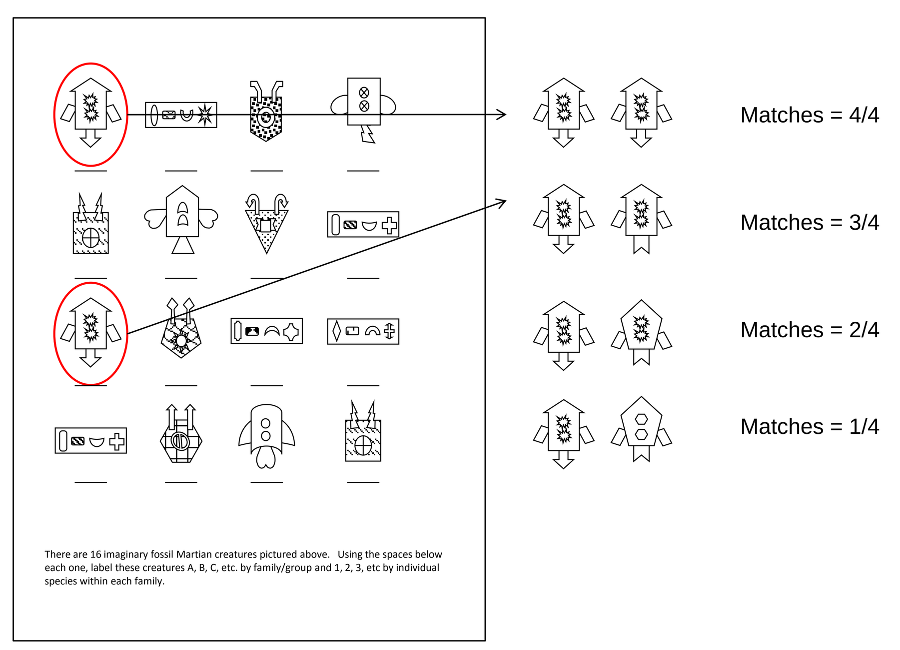
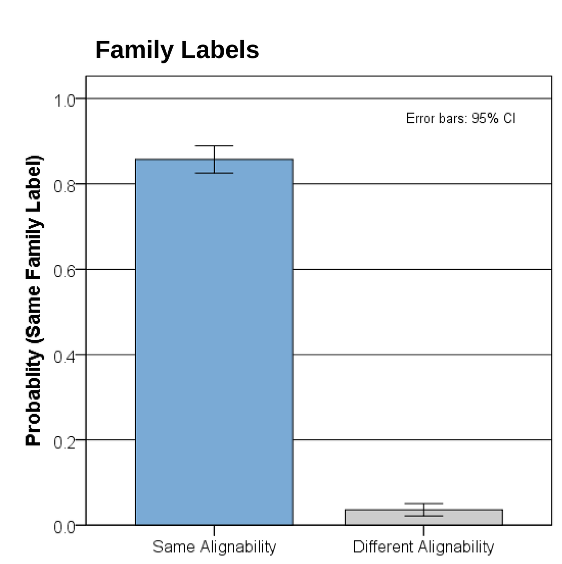


Figure 1. Sample visual display with four matching features, manipulation of matching features is shown to the right.

Three alignability-based categories (groups of stimuli defined by abstract alignability, but which vary in surface features) were included in each display (horned, winged, and box stimuli, above). Within each alignable category there was one pair of stimuli that shared at least one feature in common. The number of matching features shared by these "repeated instances" was manipulated in a between-groups design from one to four, resulting in a total of four conditions.

RESULTS

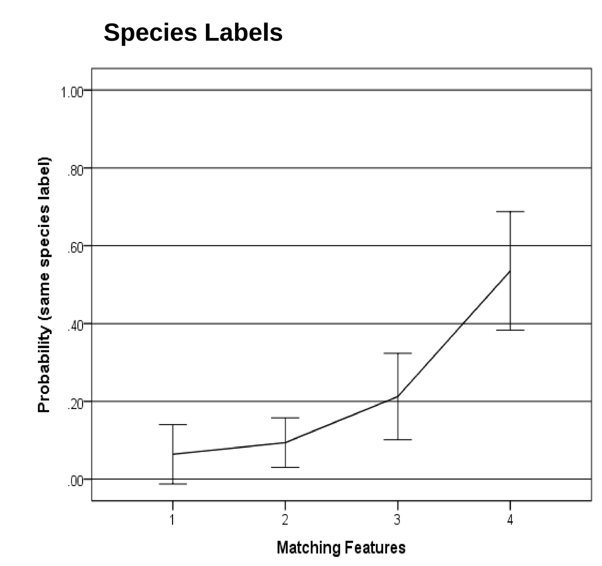
Consistent with previous findings, items from the same alignable categories were much more likely to receive the same labels than items from different, non-alignable categories. This effect was most obvious in the family labeling data, t(137) = 40.04, p < .001.



Same Alignability	0.857
Diff Alignability	0.036

These results are consistent with the hypothesis that alignability is an important grouping principle in free categorization, and further validates the reliability and applicability of the binomial labeling task.

The number of shared features strongly affected the probability of giving the two matching instances the same label. This effect was most noticeable in the species labeling data, F(3,134) = 17.20, p < .001. The linear and quadratic trends were both significant, F(1,134) = 42.86, p < .001 and F(1,134) = 7.82, p < .01, respectively.



Features Shared	Effect
1	0.064
2	0.094
3	0.212
4	0.535

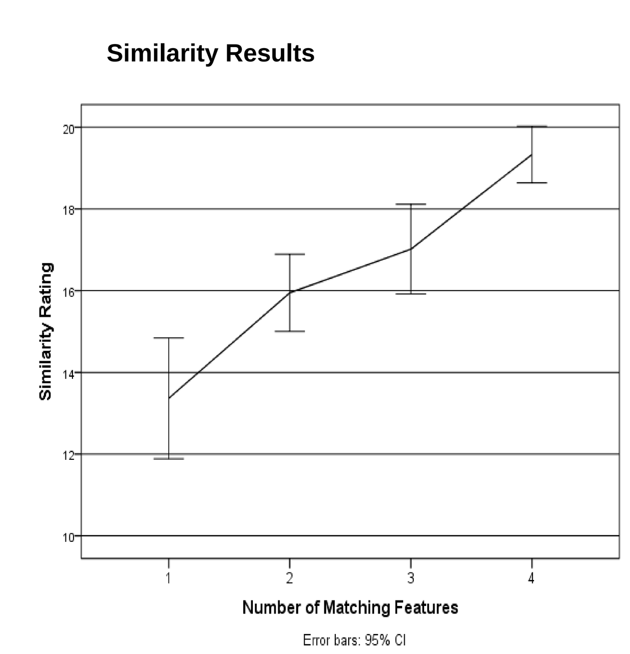
These results show that people were sensitive to the number of matching features shared by two objects. Much of this effect occurred between 3 and 4 matching features, but the linear trend computed over 1 to 3 matches was also significant, F(1,100) = 6.19, p < .02.

EXPERIMENT 2 METHODS

Participants viewed a single 4x4 array of novel visual stimuli, similar to those used in Experiment 1, and were asked to rate the similarity of selected pairs on a 20-point scale (20 being very similar, and 1 being very dissimilar). Pairs were counterbalanced within and between alignability-based categories and the inclusion of repeated instances.

RESULTS

Items from the same alignable categories were rated much more similar (M = 10.88) than items from different categories (M = 2.10, t(93) = 24.13, P < .001). The number of matching features strongly affected similarity ratings given to repeated instances, with instances sharing more features being rated significantly more similar than those with fewer features in common, F(3,90) = 21.95, p < .001. A significant linear trend was observed for similarity ratings, F(1,90) = 64.60, p < .001, but the quadratic trend was nonsignificant, F(1,90) = 0.07, p = .799.



Features Shared	Mean
1	13.36
2	15.95
3	17.02
4	19.33

GENERAL DISCUSSION

Both similarity and categorization (labeling) data show an increase with the proportion of matching features in a given target pair. This suggests that the probability of grouping two objects together into a new category may be a straightforward function of their similarity in terms of shared features. This results contrasts somewhat with that usually obtained in traditional sorting tasks[4], in which people tend to focus on a single dimension and ignore overall similarity. This may be related to the use of high-variability stimuli in the present experiment (up to seven values per dimension); under these conditions, individual feature matches are unlikely to occur by chance, and should thus be considered highly informative when observed.

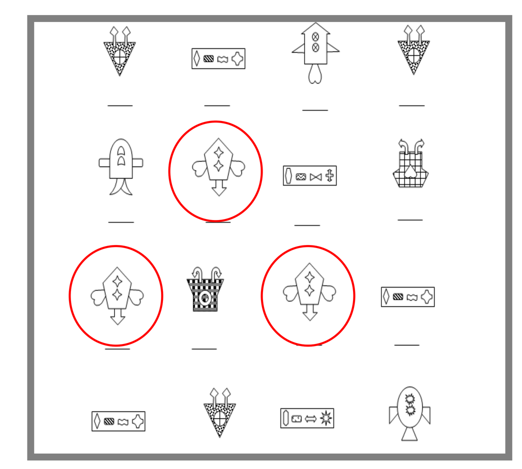
On the other hand, the two measures did show a somewhat different pattern of increase. In particular, most of the increase occurred between three and four matching features in the labeling task, while the overall pattern was more linear in the similarity task. One possibility is that categorization is affected not only by similarity by also by the probability that people will notice (or not notice) the matching features in the first place; people might be more likely to notice matching instances as such when they are identical than when they differ on one or more discrete features.

NEXT STEPS

A related question is whether the shape of the feature match sensitivity function depends on category size, i.e., will the same sensitivity function obtain or will people default to unidimensional sorting in the presence of multiple matching instances?

Repeated Instances

The number of repeated instances will be manipulated; will individual matches have larger or smaller impact as category size is increased?



Three repeated instances for each alignable category

References

[1,4] e.g., Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242-279.
[2] e.g., A binomial labeling task for category construction. Clapper, J. P. (2012). *Annual Convention of the Psychonomic Society*.
[3] e.g., Discovering categories in multi-object visual displays. Clapper, J. P. (2011). *Annual Convention of the Western Psychological Association*.