A Formal Psychological Model of Classification Applied to Natural-Science Category Learning

Robert M. Nosofsky¹, Craig A. Sanders¹, and Mark A. McDaniel²,³
¹Department of Psychological and Brain Sciences, Indiana University Bloomington; ²Department of Psychological and Brain Sciences, Washington University in St. Louis; and ³Center for Integrative Research on Cognition, Learning, and Education, Washington University in St. Louis

Abstract
The field of psychological science has seen major advances in the development of formal models of perceptual classification learning; however, little work has tested such models in real-world natural-category domains. In our current project, we aim to fill that gap by testing the ability of a formal exemplar model of classification to predict learning of rock categories in the geologic sciences. As a prerequisite for testing the model in this domain, we have conducted extensive work to derive a high-dimensional feature-space representation for the rock stimuli. An eight-dimensional representation yields good accounts of naive participants’ judgments of similarity among a large battery of rock-picture samples; furthermore, the eight dimensions have natural psychological interpretations. We then use the exemplar model in combination with the derived feature-space representation to successfully predict participants’ learning and generalization of a variety of scientifically defined rock categories. We discuss further steps for making use of the model and its associated feature-space representation to search for effective techniques of teaching categories in the science classroom.

Keywords
category learning, exemplar model, feature-space representation

The field of psychological science has seen major advances in the development of formal models of perceptual category learning and representation (for a comprehensive review, see Pothos & Wills, 2011). However, essentially all tests of such models have been in simplified domains involving highly controlled, low-dimensional stimuli and artificial category structures. There is little, if any, knowledge of the extent to which the formal models may scale up successfully to account for category learning in the real world. In our current work, we have begun to fill that gap by testing the ability of formal models to account for learning of rock classifications in the geologic sciences. Rock categories provide good examples of complex, high-dimensional category structures found in the natural world, so they serve as a challenging and representative target domain for our initial explorations. Furthermore, as we outline in our concluding discussion in this article, our hope is to put the formal models to practical educational use: The idea is to use the models themselves to help guide the search for effective strategies of teaching rock classifications in the geology classroom.

Our initial investigations have focused on a well-known exemplar model of categorization known as the generalized context model (Medin & Schaffer, 1978; Nosofsky, 1986). Because the exemplar model has accounted successfully for wide varieties of category-learning phenomena in simpler domains and serves as a foundation for a number of other highly significant models in the field (e.g., Anderson, 1991; Kruschke, 1992; Love, Medin, & Gureckis, 2004), it seemed like a fruitful starting point. According to the model, people represent categories by storing individual examples of the categories in memory. Classification decisions are...
Based on similarity comparisons with the stored examples. For instance, if a test item is highly similar to examples of Category A and dissimilar to the examples of contrasting categories, then an observer would tend to classify the item into Category A (see Nosofsky, 1986, 2011, for formal statements).

**Specification of the Feature Space**

Applying the exemplar model to generate quantitative predictions requires the specification of a multidimensional feature space in which the to-be-classified objects are embedded (Ashby, 1992; Nosofsky, 1992). In numerous past tests of the model, the derivation of the feature space was straightforward because the studies involved the classification of simple stimuli varying along a small number of salient dimensions (e.g., shapes varying in size and orientation). In such cases, the derived feature space is a fairly direct reflection of the highly controlled, experimentally manipulated dimensions that compose the objects.

In a real-world natural-category domain such as rocks, however, the derivation of a feature-space representation becomes an enormously ambitious task: Natural stimuli, rocks included, vary along a large number of complex dimensions that may be difficult to discern. In addition, because we seek generality in our tests of the exemplar model, we use multiple tokens of the numerous rock categories that our participants must learn (the main stimulus set that we have used in our recent experiments is composed of pictures of 360 rock tokens: 12 tokens from each of 30 major rock categories). Thus, a large number of stimuli must be positioned in the high-dimensional feature space to which the model makes reference. Therefore, as a prerequisite for testing the model in this rocks domain, Nosofsky, Sanders, Meagher, and Douglas (2017) initiated an extensive project to derive a feature-space representation for the rock stimuli (work on the project continues).

Although we are using a variety of complementary methods to derive the feature space, here we describe only the one based on *multidimensional scaling* (MDS). In MDS, objects are represented as points in a space, and similarity between objects is a decreasing function of distance in the space (Shepard, 1980). Using naive college students as participants, Nosofsky, Sanders, Meagher, and Douglas (2017) collected direct ratings of similarity between the numerous pairs of individual tokens in their rocks stimulus set. An MDS model was then applied to the similarity-judgment data. Basically, the application involved positioning the rocks in an M-dimensional space, such that the greater the similarity rating between any two rocks, the closer together they were in the space. The beauty of the technique is that beyond summarizing large sets of similarity-judgment data, one can inspect the derived space in which the objects are embedded to determine the major psychological dimensions that compose the objects and that govern the observers’ similarity judgments.

In the case of our MDS analysis of the rocks, the results were remarkably straightforward: An eight-dimensional solution provided a good account of the similarity structure of the 360 rock tokens that composed the 30 rock categories, and the derived dimensions could be interpreted in terms of lightness/darkness of color, average grain size, roughness/smoothness, shininess, organization, chromaticity, hue, and shape-related components. Interactive displays of the derived eight-dimensional solution are provided on the website associated with Nosofsky, Sanders, Meagher, and Douglas’s (2017) study (https://osf.io/w64fv/). (Note that although our stimuli are pictures of rocks, future work can extend the solution to include nonvisual sources of information, such as hardness, heft, and so forth.) Having established the multidimensional feature space, we were now positioned to evaluate how the exemplar model would fare in accounting for human performance in learning rock categories.

**Results of Rock Category Learning: Testing the Exemplar Model**

In one of our major category-learning studies, Nosofsky, Sanders, and McDaniel (2018) had participants learn to classify the rock stimuli into the complete set of 10 igneous-rock categories in our rock-pictures collection. The goal was to use the exemplar model, in combination with the derived MDS solution for the rocks, to account for participants’ category-learning and generalization performance. The key independent variable that was manipulated was the nature of the training examples. Across two conditions, participants first engaged in an instance-based training phase involving three training examples of each of the 10 categories. In the *center* condition, the three training examples were those closest in distance to the centroid of each category distribution defined in the eight-dimensional scaling solution for the rocks (see Fig. 1, left panel, for a schematic two-dimensional illustration). In the *coverage* condition, the three training examples more completely covered the entire rock-category distribution; however, there was far less training on central examples than in the center condition (Fig. 1, right panel).

An example with actual rock pictures is shown in Fig. 2, which illustrates the training items for rhyolite (one of the more dispersed categories that our participants needed to learn) in the center and coverage conditions. Although it is difficult to intuit the centroid of
an eight-dimensional space, the center items for rhyolite are roughly average on dimensions of lightness/darkness, average grain size, smoothness/roughness, and so forth. For the coverage items, it seems intuitively clear that one training item covers the banded rhyolites, a second covers the members with small fragments, and the third covers the members with a fine-grained homogeneous texture.

Following the training phase, participants engaged in a transfer phase that included presentation of the original training examples as well as novel rock samples from the categories. Some summary results are illustrated in Figure 3, which plots the overall proportion of correct responses on different item types across the two conditions. Not surprisingly, the center items were classified most accurately in the center condition, whereas the coverage items were classified most accurately in the coverage condition—this result indicates simply that participants performed best on those items on which they were trained. The more interesting result is that the absolute magnitude of the performance differences between the center and coverage items was much greater in the center condition than in the coverage condition. The exemplar model predicts this result (as shown by the white dots in Fig. 3) because in the center-training condition, most of the coverage items are far from the center (trained) items (Fig. 1, left panel), so there is a big dip in generalization performance on the coverage items. By comparison, in the coverage condition, all the center items tend to lie proximally to at least one coverage-training item because one of the coverage-training items is centrally located in the category distribution (Fig. 1, right panel). This proximity allows participants to generalize their exemplar-based category knowledge to the untrained center items.¹

Beyond predicting this summary result, the more impressive finding is the one illustrated in Figure 4, which plots performance on the center items, coverage items, and “neither” items (i.e., items that did not serve as either center or coverage items during training) for each of the individual 10 categories across the two conditions. The figure reveals, first, that the pattern of summary results from Figure 3 tends to be observed for each of the individual rock categories, and, second, that there is dramatic variation in overall performance levels across the different rock categories. For example, participants are extremely accurate in classifying obsidian and pumice, moderately accurate in classifying basalt and pegmatite, and less accurate in classifying andesite, diorite, granite, and rhyolite. As the white dots in Figure 4 indicate, the exemplar model does an excellent job of capturing these performance variations across the different rock categories. Furthermore, as explained in detail by Nosofsky et al. (2018), these accurate quantitative predictions were achieved while estimating only two free parameters from the model. The key to achieving the predictions is that rock categories that lie close to one another in the high-dimensional feature space tend to be confused with one another, whereas rock categories that lie in isolated regions of

---

¹ Nosofsky, Sanders, and McDaniel (2018)
the space are classified with high accuracy. We find it remarkable and encouraging that a low-parameter model, one developed in past research involving simple, artificial laboratory categories, can achieve such accurate predictions of classification performance in this complex, high-dimensional natural-science category domain.

The pattern of results illustrated in Figures 3 and 4 for our igneous-rocks experiment is robust. We conducted a conceptual replication in which participants learned to classify a mix of igneous, metamorphic, and sedimentary rocks (Nosofsky et al., 2018). This second experiment yielded the same pattern of results as described above, with the exemplar model again achieving accurate quantitative predictions across the different categories of the two training conditions.

**Fig. 2.** The 12 samples of the category rhyolite. In the top panel, the red ovals indicate the training items in the center condition. In the bottom panel, the blue rectangles indicate the training items in the coverage condition. Figure reprinted with permission from Nosofsky, Sanders, and McDaniel (2018; copyright American Psychological Association).

**Fig. 3.** Mean proportion of correct responses for the main item types in the center and coverage conditions, averaged across all categories, in Nosofsky, Sanders, and McDaniel’s (2018) Experiment 1. White dots indicate predictions of the exemplar model. Error bars show ±1 SEM. Figure reprinted with permission from Nosofsky et al. (2018; copyright American Psychological Association).
Fig. 4. Mean proportion of correct responses for the main item types in the center and coverage conditions, broken down by each individual rock category, in Nosofsky, Sanders, and McDaniel’s (2018) Experiment 1. White dots indicate predictions of the exemplar model. Error bars show ±1 SEM. Figure reprinted with permission from Nosofsky et al. (2018; copyright American Psychological Association).
Departures From the Family-Resemblance Principle

An important finding from our similarity-scaling and classification-learning studies is that the various rock categories do not appear to obey the famous "family-resemblance" principle that is presumed to organize most natural categories (Murphy, 2002; Rosch & Mervis, 1975). According to this principle, members of natural categories are composed of bundles of correlated features, producing compact and coherent clusters in a multidimensional feature space. The high-level categories of igneous, metamorphic, and sedimentary rocks appear to violate this principle and instead show highly dispersed and disorganized structures. For example, there are many igneous rocks that are highly similar to metamorphic ones and many igneous rocks that are highly dissimilar to one another (for details, see Nosofsky, Sanders, Gerdon, Douglas, & McDaniel, 2017). Even at the subtype level (i.e., different types of igneous rocks, such as obsidian, granite, and rhyolite), our preliminary evidence suggests that the category structures show analogous forms of complexity (Nosofsky, Sanders, Meagher, and Douglas, 2017). Perhaps for this reason, although the exemplar model fared well at accounting for the classification-learning data in our above-reviewed study, an alternative prototype model, which assumes that categories are learned in terms of simple summary representations (e.g., Reed, 1972; Smith & Minda, 1998), failed to account for the data.

Future Directions

Although our initial tests have been promising, we are still at the early stages of the project, and a number of challenges lie ahead. For example, geologic experts have undoubtedly learned to attend to subtle features that are required to discriminate among some of the categories, and our current feature-space representation will need to be expanded to include such information (e.g., Palmeri, Wong, & Gauthier, 2004). Certain selective-attention mechanisms that are part of the exemplar model will need to be brought into play to reflect the learning of these subtle features (Kruschke, 1992; Nosofsky, 1986, 1987). We are also exploring various teaching techniques that may bring such subtle features into focus and speed the attention-learning process (Meagher, Carvalho, Goldstone, & Nosofsky, 2017; Miyatsu, Gouravajhala, Nosofsky, & McDaniel, in press). In addition, the context and environment in which rocks are located provide important clues regarding their identity, and expert classifiers would make use of such information. Such processes can be formalized by developing theories of the settings of certain response-bias parameters that are part of the model (e.g., Cohen, Nosofsky, & Zaki, 2001). Some participants also bring prior knowledge about rock categories to our learning experiments; such knowledge can be modeled in terms of examples stored in memory prior to the participants' arrival to the laboratory (Heit, 1994).

As the project proceeds, we are hopeful that the model can be used to guide research that may enhance the teaching of rock categories in the geology classroom. For example, one question that arises in teaching scientific categories is which training examples should be used. In our study presented above, we limited consideration to the "center" versus "coverage" strategies. These cases are but single examples from an essentially infinite collection of different possibilities. Applications of the model itself could provide theoretical guidance to suggest which possibilities are the most promising. The idea would be to use the model itself to predict the learning and generalization outcomes that would be yielded by different training sets. Empirical studies could then focus on the training sets that the model predicts are optimal (for illustrations of such an approach in simpler domains, see, e.g., Patil, Zhu, Kopeć, & Love, 2014). To the extent that learning and generalization outcomes are indeed improved, this process would provide converging evidence in further support of the exemplar model while also suggesting principles and techniques for choosing training examples that might be translated to the science classroom.

Recommended Reading

Murphy, G. L. (2002). (See References). A comprehensive review of theory and research on human categorization, including sections that challenge the present exemplar-model approach.


Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by National Science Foundation Grant 1534014 (EHR Core Research) to R. M. Nosofsky, Bruce Douglas, and M. A. McDaniel.

Note

1. In some cases, one item served as both a center- and coverage-training item. Such “both” stimuli were separated out
in the detailed analyses reported by Nosofsky et al. (2018), so the center and coverage stimuli were nonoverlapping.

References


