Feature Highlighting Enhances Learning of a Complex Natural-Science Category

Toshiya Miyatsu and Reshma Gouravajhala
Washington University in St. Louis

Robert M. Nosofsky
Indiana University

Mark A. McDaniel
Washington University in St. Louis

Learning naturalistic categories, which tend to have fuzzy boundaries and vary on many dimensions, can often be harder than learning well defined categories. One method for facilitating the category learning of naturalistic stimuli may be to provide explicit feature descriptions that highlight the characteristic features of each category. Although this method is commonly used in textbooks and classrooms, theoretically it remains uncertain whether feature descriptions should advantage learning complex natural-science categories. In three experiments, participants were trained on 12 categories of rocks, either without or with a brief description highlighting key features of each category. After training, they were tested on their ability to categorize both old and new rocks from each of the categories. Providing feature descriptions as a caption under a rock image failed to improve category learning relative to providing only the rock image with its category label (Experiment 1). However, when these same feature descriptions were presented such that they were explicitly linked to the relevant parts of the rock image (feature highlighting), participants showed significantly higher performance on both immediate generalization to new rocks (Experiment 2) and generalization after a 2-day delay (Experiment 3). Theoretical and practical implications are discussed.

Keywords: category learning, educational application, geology, multimedia learning

The cognitive psychology literature on how humans acquire and represent categories is quite substantial (for reviews, see Ashby & Maddox, 2005; Murphy, 2002; Pothos & Wills, 2011). Yet, nearly all of the research on category learning in the past 50 years has been conducted with artificial stimuli, such as geometric shapes and simplified drawings. Consequently, the extent to which the principles discovered in this research apply to learning of natural categories is unclear. One point of departure between the work on artificial categorization to new rocks (Experiment 2) and generalization after a 2-day delay (Experiment 3). Theoretical and practical implications are discussed.

Keywords: category learning, educational application, geology, multimedia learning

The cognitive psychology literature on how humans acquire and represent categories is quite substantial (for reviews, see Ashby & Maddox, 2005; Murphy, 2002; Pothos & Wills, 2011). Yet, nearly all of the research on category learning in the past 50 years has been conducted with artificial stimuli, such as geometric shapes and simplified drawings. Consequently, the extent to which the principles discovered in this research apply to learning of natural categories is unclear. One point of departure between the work on artificial categorization to new rocks (Experiment 2) and generalization after a 2-day delay (Experiment 3). Theoretical and practical implications are discussed.

Keywords: category learning, educational application, geology, multimedia learning

Learning naturalistic categories, which tend to have fuzzy boundaries and vary on many dimensions, can often be harder than learning well defined categories. One method for facilitating the category learning of naturalistic stimuli may be to provide explicit feature descriptions that highlight the characteristic features of each category. Although this method is commonly used in textbooks and classrooms, theoretically it remains uncertain whether feature descriptions should advantage learning complex natural-science categories. In three experiments, participants were trained on 12 categories of rocks, either without or with a brief description highlighting key features of each category. After training, they were tested on their ability to categorize both old and new rocks from each of the categories. Providing feature descriptions as a caption under a rock image failed to improve category learning relative to providing only the rock image with its category label (Experiment 1). However, when these same feature descriptions were presented such that they were explicitly linked to the relevant parts of the rock image (feature highlighting), participants showed significantly higher performance on both immediate generalization to new rocks (Experiment 2) and generalization after a 2-day delay (Experiment 3). Theoretical and practical implications are discussed.

To our knowledge, unlike research with artificial categories, none of the experimental work with naturalistic categories has implemented paradigms in which key features are explicitly identified for the learner. More important, no research has examined the consequences of providing explicit feature information for learning of natural categories—see for instance the paradigms developed to examine learning of different types of snakes (Noh, Yan, Vendetti, Castel, & Bjork, 2014), categories of paintings (Hartley & Homa, 1981; Kang & Pashler, 2012; Kornell & Bjork, 2008), types of birds (Jacoby, Wahlheim, & Coane, 2010; Wahlheim & DeSoto, 2017; Wahlheim, Finn, & Jacoby, 2012), types of butterflies (Birnbaum, Kornell, Bjork, & Bjork, 2013), and types of rocks (Nosofsky, Sanders, Gerdom, Douglas, & McDaniel, 2017; Nosofsky, Sanders, & McDaniel, 2017). In this article, we
present three novel experiments that provide an initial investiga-
tion of the effects of providing explicit feature information to
learners who are attempting to learn to classify rock types. As we
develop below, this issue is a theoretically interesting one for
naturalistic-category learning, with uncertain outcomes. On the
one hand, there are reasons to expect that providing features will
benefit naturalistic category learning; on the other hand, there are
also plausible reasons to expect that providing features will not
benefit, and may even interfere with, generalization.

Beyond the theoretical interest, explicitly identifying key fea-
tures of to-be-learned natural categories aligns with educational
practice. Presentations of examples of to-be-learned categories in
classrooms and other educational situations are almost always
accompanied by verbal descriptions of some key features of the
category that have been prepared by experts in the field. For the
instruction of rock categories, an instructor may show her students
an example of the rock obsidian while pointing out that this type
of rock tends to be black, have a glassy texture, and come with
shell-like fractures. Textbooks often do the same in that a picture
of a particular type of object is often accompanied by some verbal
description of the key features of the given category—for an
example involving rock categories, see Figure 1, which is ex-
tracted from a popular geology lab manual (Busch & Tasa, 2009).
Clearly, the educators’ intuitions are that such feature descriptions
will facilitate students’ learning of these natural categories. Yet,
there is no experimental evidence that informs this intuition. Be-
low we develop the theoretical reasons for why providing feature
descriptions could facilitate category learning, and other theoreti-
cal reasons for why feature descriptions may not facilitate category
learning. Although we draw our reasoning from research using a
variety of category structures, we believe that the processes we
illuminate in these sections are potentially present in the learning
of novel geological rock categories. Then, we present three exper-
iments directly contrasting learning of rock categories when fea-
ture descriptions are provided relative to when they are not.

The Case for Highlighting Critical Features

There is preliminary evidence from studies involving rule-based
laboratory categories that feature provision facilitates learning.
Salatas and Bourne (1974) had participants study a series of
gеometric shapes differing in four dimensions: color (blue, yellow,
or red), size (large, medium, or small), form (hexagon, square, or
triangle), and number of figures (one, two, or three). Although they
did not directly manipulate whether they gave feature descriptions,
their participants were told which features were relevant (one
feature from each of two dimensions was always relevant; e.g., the
color “blue” and “large” size). Compared with other studies that
used similarly structured materials, Salatas and Bourne’s partici-
pants showed accelerated learning.

More recently, Eglington and Kang (2017) provided evidence
that highlighting key features facilitates acquisition of a rule-based
science category. Participants learned various chemical categories
by studying diagrams of organic compounds. Highlighting diag-
nostic features of each chemical category in red (e.g., the presence
of phosphorus or chlorine in the example diagrams belonging to
organophosphate and organochloride categories respectively; see
Eglington & Kang, 2017, for examples of highlighted diagrams)
substantially improved learning (comparing across Experiments 2 and 3).

Rock categories (and natural categories in general) bring additional complexities that might lead one to expect a similar positive effect of highlighting features. In complex natural category learning—as opposed to rule-based laboratory categories like those from Salatas and Bourne (1974) or rule-based science categories like those in Eglington and Kang (2017)—learners have to first identify dimensions that characterize the corpus of instances in which they are delineated by different categories (and that presumably contribute to category differentiation). One challenge for the learner is that some of these dimensions may not be immediately apparent (Schyns & Rodet, 1997). This challenge contrasts with the vast majority of artificial categories in which the dimensions are readily available (e.g., common dimensions are size, color, shape, and number of stimuli). In the case of rocks, there are many dimensions on which rocks vary, such as color, texture, and granularity, to name a few (cf. Nosofsky, Sanders, Meagher, et al., 2017). However, learners may not encode all the dimensions on their own. For example, when first encountering several rock examples, learners may notice only more salient dimensions and overlook other dimensions. If the learners persist in attending to just salient dimensions, it is possible that they may fail to encode less salient but relevant dimensions for categorization. It is even theoretically possible that some dimensions are ambiguous and must be perceptually constructed (not directly extracted), as demonstrated in the Schyns and Rodet (1997) work with artificial categories. Accordingly, explicitly highlighting the relevant dimensions and specific feature values for particular categories would support learning by ensuring that the relevant dimensions are attended regardless of its salience.

Moreover, after identifying the dimensions, learners have to observe the variation in each dimension and identify the appropriate range of values that characterize the examples from a given category. In the case of the rock categories, after identifying color, texture, and granularity as diagnostic dimensions (for instance), learners have to identify the values of each dimension that characterize particular categories. For example, obsidian tends to be glassy black whereas marble is light-colored. Highlighting critical features could give learners an advantage in identifying the characteristic features of each category. As an example, consider the processes assumed in one particular model of category learning, the RULEX model (Nosofsky, Palmeri, & McKinley, 1994). According to RULEX, category learners generate simple rules (i.e., hypotheses about which features are critical for determining the categories) that can encompass as many instances as possible and then memorize the exceptions to these rules for each category. Highlighting feature descriptions would presumably speed the adoption of reasonably accurate simple rules, thereby allowing learners to more quickly focus on learning the exceptions. Indeed, in one way or another, a process of identifying critical dimensions for categorization is assumed in almost all of prominent models of category learning, such as prototype models (e.g., Reed, 1972), rule-learning models (e.g., Bower & Trabasso, 1963; Levine, 1975; Nosofsky et al., 1994), exemplar models (e.g., Kruschke, 1992; Nosofsky, 1984), and clustering models (e.g., Love, Medin, & Gureckis, 2004) and, thus, it might be broadly expected that highlighting features for learners would facilitate category learning.

Finally, unlike the vast majority of artificial category structures in which variation in the same dimension(s) forms the basis of all categories to be learned, characteristic features of rock categories, at least, may be associated with different dimensions across categories. For instance, whereas the characteristic features of granite include color (light color or reddish) and granularity (interlocking coarse-grained crystals), the primary characteristic feature of pumice involves texture (rough textured volcanic glass with small holes). This form of variation may create additional complexity for learners. Consequently, highlighting critical features for each category could enhance learning by focusing learners’ attention to the relevant dimensions that can be especially helpful in this type of complex category structure.

The Case Against Highlighting Critical Features

The literature on category learning and perceptual expertise also raises the possibility that highlighting features may not enhance and may even interfere with natural category learning. First, experts and novices differ in the dimensions to which they attend, as well as the precision with which they can detect the differences within a given dimension (Goldstone, 2003; Sowden, Davies, & Roling, 2000; see Palmeri, Wong, & Gauthier, 2004, for review). For example, when medical X-ray images were shown to experts and novices, experts had higher sensitivity to low-contrast dots, a dimension that is critical in detecting abnormalities (Sowden et al., 2000). Thus, feature characteristics that experts use to categorize objects may not be immediately useful for novices, because novices cannot identify these dimensions or lack the acuity necessary to detect the changes in the given dimension associated with each category.

Second, highlighting features may circumvent processes necessary for natural category learning, given that natural categories can have many dimensions. Pertinent to the present focus on rock category learning, Nosofsky, Sanders, Meagher, et al. (2017) conducted multidimensional scaling on similarity rating data of 12 different examples from each of 30 common rock categories. The resulting psychological feature space for rocks’ visual features involved at least eight complex dimensions, more than is typically found in artificial categories where it is not uncommon to generate stimuli that vary on only two dimensions (e.g., Ashby & Maddox, 1992; Nosofsky, 1986).

More important, allocating significant attention to all of these dimensions may be critical in successful learning of the rock categories. Consistent with this possibility, Nosofsky, Sanders, and McDaniel, (2017) had participants learn some of these rock categories (from which the multidimensional scaling study was based) and fitted several models (e.g., prototype model, exemplar model) to the test classification data. The exemplar model was the most successful in accounting for the observed data (the generalized context model; Nosofsky, 1986, 2011). Importantly for present purposes, the successful instantiation of the exemplar model assumed that learners gave significant attention to all dimensions (rather than weighing some dimensions dramatically more than others). Providing selected feature descriptions would clearly be expected to bias learners’ attention toward these selected dimensions at the outset of learning. Doing so, however, could conceivably distract learners from adequately encoding the training instances along the entire feature space. Thus, if the encoding of the
entire feature space is necessary for successful learning of rock categories, it is possible that provision of selective features may hinder learning.

Lastly, the findings from studies using artificial categories that favor providing defining features to learners (Salatas & Bourne, 1974) may not be directly applicable to more complex categories like rocks because of the numerous differences between artificial and natural categories. Unlike many artificial laboratory categories that vary in a few dimensions (e.g., color, size, and form) with only a few discrete features (e.g., blue, yellow, or red in the color dimension), as noted above rock categories have a greater number of dimensions (Nosofsky, Sanders, Meagher, & Douglas, 2017) and the feature variation within a dimension is continuous. Moreover, natural categories tend to have high variability within dimensions (Murphy, 2002), and the rock categories in particular appear to have rather complex structures (Nosofsky, Sanders, Gerdon, et al., 2017). For example, in some cases, an instance from one category looks more similar to instances from other categories than to the instances from the category to which it belongs. Thus, the facilitative effect of providing features for clearly defined rule-based categories (Salatas & Bourne, 1974) may not extend to learning rock categories.

Experiment 1

One straightforward way to provide feature descriptions is to place them right below the rock images used for category training, as is often done in textbooks (see Figure 1). Thus, in this first experiment, during training, some participants were presented with rock images and an associated feature description below each image (along with the category name), whereas other participants studied the same rock images with only the category name shown.

We also manipulated training-exemplar variability. Increasing the number of unique training exemplars as opposed to repeating the same training exemplars has been shown to enhance classification performance of new examples (Wahlheim & DeSoto, 2017; Wahlheim et al., 2012). However, it is possible that participants may not need as many unique training exemplars if feature descriptions are provided. One potential mechanism through which training-exemplar variability helps learning is that it allows learners to observe the variation associated with each dimension more precisely. This enhancement in precision can lead to a better identification of unique features of each category in a given dimension. To explore this possibility in Experiment 1, half the participants studied the categories of rocks in a condition with six unique training exemplars per category, with each individual exemplar repeated twice; whereas the other half studied with two unique training exemplars per category, with each individual exemplar repeated six times.

Following training, we evaluated participants’ ability to classify the old training items and new transfer items in a test phase. In addition, we assessed participants’ learning strategy with a post-experimental questionnaire that asked participants to rate whether they took a more abstraction-based or memorization-based approach. Much debate in the category learning literature focused on how learned categories are represented (Murphy, 2002). According to an abstraction-based approach, learners develop some sort of a summary representation of the learned categories. Specifically, learners may develop a set of rules that determines categorization (rule-based) or prototypes of the learned categories (prototype-based), and the rules or the prototypes are applied to classify exemplars that are subsequently encountered (e.g., Allen & Brooks, 1991; Erickson & Kruschke, 1998; Posner & Keele, 1968, 1970; Regehr & Brooks, 1993; Trabasso & Bower, 1968). On the other hand, a memorization-based (or exemplar-based) approach assumes that learners store exemplars from each category and the subsequent categorization judgment is made according to the similarity of the newly encountered exemplars to the stored exemplars (e.g., Brooks, 1978; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; Nosofsky & Kruschke, 1992). Based on the artificial category learning literature, we thought it possible that learners would vary in terms of adopting a more abstraction-based or memorization-based approach within a single categorization task (Bourne, Healy, Kole, & Graham, 2006; Little & McDaniel, 2015a; Wahlheim, McDaniel, & Little, 2016; see also Little & McDaniel, 2015b).

Method

Participants and design. There were 204 Amazon Mechanical Turk workers who participated in exchange for monetary compensation. To qualify for participation in the experiment, Mechanical Turk workers had to have met the following three qualifications: A HIT (Human Intelligence Task) approval rate for all requesters’ HITs greater than or equal to 95%; the number of HITs approved was greater than or equal to 500; and must reside in the United States. The mean age of the sample was 36.5 years of age, and 62% had obtained an Associate’s degree or higher. Participants who failed to complete the experiment or who reported a high prior knowledge of the material on a postexperiment questionnaire (n = 28) were excluded from the analyses (final N = 176). A 2 × 2 between-subjects factorial design was used, with feature highlighting (feature-description present, absent) and number of unique training exemplars (2, 6) as independent variables.

Materials. Pictures of rocks from 12 categories (four categories from each of the three higher-order categories of igneous, metamorphic, and sedimentary) were assembled. The images were scaled to be similar sizes and presented against a white background. The materials were taken from a larger set created by Nosofsky, Sanders, and Meagher, et al. (2017). The present materials consisted of 12 rocks from each of the following 12 categories: granite, obsidian, pegmatite, pumice, amphibolite, gneiss, marble, slate, breccia, conglomerate, rock gysum, and sandstone. For each of these categories, a geological science faculty member at Indiana University provided short feature descriptions that were presented in conjunction with the rocks in the highlighted feature conditions. Note that because these are characteristic features (i.e., features that are typically but not always present in category members), some of the rock tokens were missing one or more of the described category-relevant features. The pictures of rocks used in all the experiments reported in the current article can be found at Open Science Framework (https://osf.io9vg8hn/#). Appendix A provides the feature descriptions for each category. All experiments reported in the current article were programmed in Collector (http://github.com/gikeymarcia/Collector), a PHP-based open-source experiment program designed to conduct psychological experiments through Web-browsers.
Procedure. Participants first completed an observation phase, where they were presented with a randomly chosen rock from each of the 12 categories and their corresponding category label for 6 s (feature descriptions were not provided). No feedback was provided in this phase.

Then, participants completed 12 blocks of feedback learning. Each feedback learning block contained 12 trials. For each block, across trials a training exemplar from each of the 12 categories was presented. For a trial, the training exemplar was presented for 6 s, during which time participants clicked one of the 12 category names (using the mouse) to indicate a response. Feedback was then presented for 4 s. If participants failed to make a response within the 6 s, the program nevertheless progressed to the feedback screen. For the feature description-absent conditions (FD-absent), the feedback was the picture of the just-tested rock with the correct category label. For the feature description-present (FD-present) conditions, the feedback in addition presented a feature description below the picture.

For each participant, the appropriate number (2 or 6) of training exemplars were randomly selected from the 12 exemplars in each category, and the presentation sequence of these training exemplars was randomized. In the 2-unique-training-exemplar conditions, the feedback learning blocks contained six repetitions of the same two randomly chosen training exemplars from each of the 12 categories. In the 6-unique-training-exemplar conditions, the feedback learning blocks contained two repetitions of these six training exemplars. In all conditions, the first feedback learning block contained the same rocks as those presented in the observation phase. The repetition of training exemplars was spaced by block; in other words, in the 6-unique-training-exemplar conditions, for example, an item was presented once in the first six blocks and then repeated once in the last six blocks.

After participants completed the 12 feedback learning blocks, they completed a distractor task (Tetris) for three minutes. Following this distractor task, participants then completed the final test phase. In the final test phase, participants were presented with 72 test items, containing two rocks from each of the 12 categories they had been trained on (old training items) and four randomly chosen rocks that they had not been trained on (new transfer items) from each of the 12 categories. The order of presentation was randomized for each participant. The participants were instructed to select, at their own pace, which category each presented item belonged to by clicking one of the 12 category names using the mouse. No feedback was provided during the final test phase.

Following the final test phase, participants completed a strategy questionnaire (adopted from Wahlheim et al., 2016) where they reported the strategy used during the training phase. Participants were instructed to think back to the training phase and remember whether they were more focused on developing a rule for each category, or on memorizing rocks and their corresponding category labels. They made their ratings on a Likert scale ranging from 1 (memorization strategy) to 7 (rule-establishing strategy). Participants were instructed to give an extreme rating only if they used that strategy exclusively throughout training, but to otherwise use the full range of the scale to indicate the degree to which they preferred one strategy over the other. Participants were also told to give a rating of 4 if they used both strategies equally or if they were unsure about their strategy preferences.1

Results

Training phase performance. Figure 2 shows the learning curves for each condition across the 12 feedback learning blocks. A 2 (FD present or FD absent) × 2 (number of unique training exemplars: 2 or 6) × 12 (feedback learning blocks) mixed analysis of variance (ANOVA), with the presence or absence of feature descriptions and the number of unique training exemplars as the between-subjects variables and the feedback learning blocks as the within-subjects variable, was conducted on these data.

There was no main effect of the presence or absence of feature descriptions, \(F(1, 172) < 1\). There was a significant main effect of the number of unique training exemplars, \(F(1, 172) = 32.11, p < .001, MSE = 11.70, \eta^2_p = 0.16\), indicating that the performance during the training phase was higher when participants were provided 2 unique training exemplars (\(M = .69, SD = .18\)) than when they were provided 6 unique training exemplars (\(M = .54, SD = .18\)). There was a significant main effect of learning block, \(F(11, 172) = 121.44, p < .001, MSE = 1.72, \eta^2_p = 0.41\), such that accuracy increased over the course of the 12 feedback learning blocks.

The interaction between number of unique training exemplars and the presence or absence of feature descriptions was not significant, \(F(1, 172) < 1\). There was a significant interaction between feedback learning block and the number of unique training exemplars, \(F(11, 172) = 19.39, p < .001, MSE = 0.28, \eta^2_p = 0.10\). Inspection of Figure 2 suggests that accuracy for those in the 2-unique-training-exemplar condition increased more from the first half (\(M = 0.59, SD = .18\)) of training to the second half (\(M = 0.78, SD = .19\)) than it did for those in the 6-unique-training-exemplar condition (first half: \(M = 0.48, SD = 0.17\); second half: \(M = 0.60, SD = 0.19\)).

There was also a significant interaction between feedback learning block and the presence or absence of feature descriptions, \(F(11, 172) = 3.38, p < .001, MSE = 0.05, \eta^2_p = 0.02\). As seen in the lower panel of Figure 2, the pattern of training performance across the 12 learning blocks does not seem to reveal any systematic differences between the FD-present and FD-absent conditions.

We will further examine this interaction if it turns out to be reliable (i.e., also emerges in Experiments 2 and 3). The three-way interaction between feedback learning block, the number of unique training exemplars, and the presence or absence of feature descriptions was not significant, \(F(11, 172) = 1.04, p > .05, MSE = 0.02, \eta^2_p = 0.01\).

Final test performance. Figure 3 shows participants’ final test performance as a function of conditions and test type. A 2 (test type: old training or new transfer) × 2 (number of unique training exemplars: 2 or 6) × 2 (FD present or FD absent) mixed ANOVA, with the test type as the withinsubjects variable and the number of unique training exemplars and the presence or absence of feature description as the between-subjects variables, was conducted on these data.

1 The strategy questionnaire results did not reveal anything directly pertinent to the main questions of the current study; however, for completeness these results are reported in Appendix B.
significant interaction between test type and the number of unique training exemplars, $0.62$, $0.41$. There was also a significant main effect of the number of participants were provided 2 unique training exemplars ($F$, $M$). Performance on the new transfer items was identical across the two $6$-unique exemplar conditions: $M = .59$, $SD = .16$.

The two-way interactions between the presence or absence of feature descriptions and number of unique training exemplars and the presence or absence of feature descriptions and test type were not significant (largest $F(1, 172) = 1.60$, $MSE = 0.02$, $\eta^2_p = 0.01$, for the latter interaction), nor was the three-way interaction ($F < 1$).

**Discussion**

For present purposes the most important finding was that providing feature descriptions did not enhance learning. One interpretation is that this pattern aligns with the theoretical perspectives outlined in the introduction that anticipated no benefits of highlighting features for learning these natural rock categories. An alternative, however, is that the presentation of the feature descriptions did not function in its intended manner for several potential reasons. First, the expert-like terminology that was included in the feature descriptions (e.g., extrusive, vesicles) may not be familiar to laypeople (MTurk workers) and, thus, what the descriptions were referring to may have been unclear. Moreover, even if the participants could understand what the terms meant, it is possible that they could not tie the terms to the corresponding perceptual dimensions in the rock pictures. For example, when a participant saw a picture of *amphibolite* with a description that read “dark colored and weakly grained structure,” she might not understand what perceptual features reflected “grain.” Lastly, even if the participants could understand the meaning of the terms and identify the described dimension, they might not be able to tie the description to specific perceptual features that characterized the values within the dimension (e.g., what perceptual features counted toward characterizing the grains as “weak”).

Closely related to the above point, it appears in retrospect that the features were not presented in a way that supported optimal processing of the information. Multimedia learning principles (Mayer, 2002, 2005) assume that to maximize the benefits of presenting words and pictures together, extraneous processing (i.e., any processing that does not contribute to learning and drains limited cognitive capacity) needs to be minimized, so that learners can allocate more cognitive resources to essential processing (i.e., selecting relevant information and organizing them in working memory). In Experiment 1, gaze shifts from the caption below the picture to corresponding parts of the picture, as well as the act of looking for the parts of the picture described in the feature descriptions, could be considered extraneous processing. Reducing these aspects of extraneous processing may be crucial in drawing conclusions.
out the benefits of feature highlighting, especially given the limited
time participants had to process the feature feedback (4 s).

One surprising finding was that training with a larger sample of
different exemplars did not enhance generalization to new transfer
items. From the perspective of prototype theory, more complete
category learning requires appreciating the allowable variation
from a prototype that category members can exhibit (e.g., Posner
& Keele, 1968). In line with this perspective, training with a larger
set of different exemplars has been shown to enhance learning of
other natural categories, that of birds (Wahlheim et al., 2012). It is
beyond the scope of the present research, however, to pinpoint
why these results with rock categories versus bird categories are
discrepant.

Experiment 2

In Experiment 2 the same feature descriptions used in Experi-
ment 1 were embedded within the rock pictures, and the corre-
sponding parts of the image that were reflected in the feature
description were circled (see Figure 4 for examples; for conve-
nience we refer to this particular feature-description implementa-
tion as feature highlighting). We reasoned that these modifications
should clarify the perceptual aspects of the rock to which the
features referred (i.e., signaling; Meyer, 1975) as well as facilitate
the scanning needed to integrate reading the described features and
processing the images. Specifically, these modifications follow the
spatial contiguity principle of the multimedia principle (Mayer,
2002, 2005) that allows learners to minimize extraneous process-
ing and maximize essential and generative processing. Supporting
this reasoning, Moreno and Mayer (1999) had participants study
how lightning forms with an animation and words. Participants’
performance on a transfer test was higher when the words were
embedded in the animation next to the specific parts being de-
scribed compared with when the words were placed below the
animation as a caption. Similarly, Eglington and Kang’s (2017)
successful implementation of feature highlighting was embedded
within the training pictures of organic-compound diagrams.

Figure 4. Examples of rock pictures with embedded feature descriptions used in Experiment 2 and 3. See the
online article for the color version of this figure.
If the participants in Experiment 1 failed to benefit from the feature provision because of their inability to understand or locate the perceptual aspects described by the feature descriptions, their inability to overcome the burden posed by extraneous processing, or both, then the participants in the present feature highlighting condition should outperform those in the control condition. Alternatively, if the participants in Experiment 1 failed to benefit from feature provision because of other reasons (as outlined in the introduction), Experiment 2 should not show the advantage of feature highlighting.

Method

Participants and design. There were 80 Washington University in St. Louis undergraduate students who participated in this experiment in exchange for course credit or monetary compensation. There were two between-subjects conditions: FD present or absent. All participants were trained on 6 unique training exemplars, each repeated twice across the 12 feedback learning blocks.

The materials and procedure were identical to Experiment 1, except the feature descriptions in the FD-present condition were embedded in the rock pictures and the parts being described were circled as shown in Figure 4. Specifically, each feature description was split into shorter fragments that referred solely to one particular dimension (e.g., color, pattern, texture, and grain size). Features were highlighted using a colored circle and the corresponding feature description fragment was placed directly adjacent to the circle along with the category name.

Results and Discussion

Training phase performance. Figure 5 shows the learning curves for each condition across the 12 feedback learning blocks. A 2 (FD present or FD absent) × 12 (feedback learning blocks) mixed ANOVA, with feature-description condition as the between-subjects variable and the feedback learning blocks as the within-subjects variable, was conducted on these data. There was no significant main effect of feature description, F(1, 78) < 1. There was a significant main effect of learning block, F(11, 78) = 52.05, p < .001, MSE = 0.80, ηp² = 0.40, such that performance increased as training progressed.

There was a significant interaction between feature-description condition and learning block, F(11, 78) = 6.79, p < .001, MSE = 0.10, ηp² = 0.08. To try to simplify the source of this interaction based on our observations from the Experiment 1 training results we collapsed performance across the first and second halves of the feedback learning blocks. We found no significant interaction between the first and second halves of feedback learning and condition, however, F(1, 78) = 2.14, p > .05, MSE = 0.01, ηp² = 0.03. Therefore, we conducted simple comparisons, and found that the interaction appears to have been mostly driven by a significant advantage in learning in the FD-absent condition (M = 0.81, SD = 0.18) relative to the FD-present condition (M = 0.63, SD = 0.20) in Block 7, t(78) = 4.23, p < .001, d = 0.95. Therefore, although there was a significant interaction in both Experiments 1 and 2, there appears to be no systematicity in the pattern. In summary, as was the case in Experiment 1, providing feature descriptions did not generally affect categorization performance during the training phase.

Final test performance. Figure 6 shows participants’ final test performance as a function of conditions and test type. A 2 (FD present or FD absent) × 2 (test type: old training or new transfer) mixed ANOVA, with the feature description condition as the between-subjects variable and the test type as the within-subjects variable, was conducted on these data. In general, performance on the old training items was significantly better (M = .85, SD = .10) than on the new transfer items (M = .80, SD = .08), F(1, 78) = 46.39, p < .001, MSE = 0.13, ηp² = 0.37.

More important, there was a significant main effect of feature description, F(1, 78) = 25.49, p < .001, MSE = 0.74, ηp² = 0.25, such that those in the FD-present condition performed significantly better on the final test (M = .89, SD = .12) than did those in the FD-absent condition (M = .76, SD = .12). This effect was qualified by a significant interaction with test type, F(1, 78) = 132.39, p < .001, MSE = 0.37, ηp² = 0.63. Post hoc independent-sample t tests showed that new transfer performance was substantially advantaged when features were provided during training relative to when they were not provided (FD present M = .91, SD = .10; FD absent M = .68, SD = .12), t(78) = 9.38, p < .001, d = 2.08. By contrast, performance on old training items did not significantly differ between the groups (FD present: M = .87, SD = .13 vs. FD absent: M = .83, SD = .15), t(78) = 1.28, p > .05, d = 0.28, consistent with the patterns during the training phase.

In summary, the results supported the possibility that the Experiment 1 participants failed to benefit from feature provision because the feature description did not function in its intended manner. The present results suggest that learners can benefit from feature descriptions if those descriptions are presented such that learners can locate and identify the described perceptual aspects. Also, the placement of the feature descriptions may have supported optimal processing of the verbal descriptions in conjunction with the perceptual input (according to the spatial contiguity principle). An interesting find was that feature highlighting selectively enhanced performance on the new transfer items; it did not enhance performance on the old training items or initial training accuracy. This pattern indicates that providing features did not facilitate learning particular training instances, but it did facilitate learning the categories—that is, aspects that supported generalization. It should be noted, however, that the interaction may have emerged because of a ceiling effect given the high performance on both old
training (M = .87) and new transfer items (M = .91) in the FD-present condition. Before discussing the theoretical implications of this result, we first report another experiment to establish the reliability of the benefit of highlighting features for learning rock categories.

Experiment 3

Experiment 3 was conducted to extend the experiment 2 findings to a delayed final test (a 2-day delay).

Method

Participants, design, and procedure. There were 59 undergraduate students at Washington University in St. Louis who participated in this experiment in exchange for course credit or monetary compensation. There were 2 between-subjects conditions: FD present (n = 29) or FD absent (n = 30). The procedure was identical to Experiment 2, except for the introduction of a 48-hr delay between the training and the final test phases.

Results

Training phase performance. Figure 7 shows the learning curves for each condition across the 12 feedback learning blocks. The performance increase observed in the feedback learning Block 7 was because of the fact that that was the first time the participants saw the same training exemplars for the second time (six training exemplars per category were repeated twice). A 2 (FD present or FD absent) × 12 (feedback learning blocks) mixed ANOVA, with the feature description condition as the between-subjects variable and the feedback-learning blocks as the within-subjects variable, was conducted on these data. Converging with the previous experiments, there was no significant main effect of feature description, F(1, 57) = 3.07, p = .09, MSE = .36, η² = .05. There was a significant main effect of feedback learning block, F(11, 57) = 42.59, p < .001, MSE = 0.66, η² = 0.43, such that performance increased over the course of the 12 feedback learning blocks. The interaction between feature description and feedback learning block was not significant, F(11, 57) = 1.16, p > .05, MSE = 0.02, η² = 0.02.

Final test performance. Figure 8 shows participants’ final test performance as a function of conditions and test type. A 2 (FD present or FD absent) × 2 (test type: old training or new transfer) mixed ANOVA, with the feature-description condition as the between-subjects variable and test type as the within-subjects variable, was conducted on these data. There was a significant main effect of test type, F(1, 57) = 104.13, p < .001, MSE = 0.46, η² = 0.65, such that performance on the old training items (M = .84, SD = .12) was significantly better than on the new transfer items (M = .72, SD = .11).

More important, as in Experiment 2, providing highlighted features during training produced more accurate test performance (M = .81, SD = .10) compared with not providing features during training (M = .75, SD = .10), F(1, 57) = 5.58, p < .05, MSE = 0.11, η² = 0.09. A marginally significant interaction with test type suggested that the advantage of providing critical features was limited to new transfer items, F(1, 57) = 3.60, p = .06, MSE = 0.02, η² = 0.06. Post hoc independent-sample t tests confirmed this suggestion. On new transfer items, the participants provided with feature descriptions during training (M = .76, SD = .10) significantly outperformed those who were not provided with feature descriptions (M = .68, SD = .11), t(57) = 3.12, p < .05, d = 0.76. On old training items, the groups did not statistically differ (FD present: M = .86, SD = .10 vs. FD absent: M = .83, SD = .13), t(57) = 1.27, p > .05, d = 0.26.

In short, Experiment 3 extended the main finding in Experiment 2 to a final test that was administered 2 days after training. Feature
provision during training selectively enhanced the performance on new transfer items but not on old training items.

**General Discussion**

Providing feature descriptions of rock categories as a caption below images of training instances failed to improve learning of the rock categories relative to training without the feature descriptions (Experiment 1). However, when these verbal feature descriptions were displayed so that the described feature was explicitly linked to particular perceptual aspects of the rock image (that we termed feature highlighting), learning of rock categories was significantly enhanced as evidenced on an immediate generalization test (Experiment 2) and a 2-day-delayed generalization test (Experiment 3). As reviewed in our introduction, classic previous work has shown that providing explicit information concerning defining features facilitates learning of artificial rule-based categories (Salatas & Bourne, 1974). The present work extends such findings in a significant manner by showing that parallel results hold for the learning of natural categories. Given that natural categories have more complex category structures than laboratory rule-based categories (e.g., unlikely to have defining features), a richer array of perceptual dimensions, continuous dimensional values (unlike many laboratory categories), and high variation within dimensions, it was uncertain a priori that feature provision would enhance natural-category learning. We first discuss these results in terms of their theoretical implications and then highlight the important practical implications.

One theoretical idea suggested at the outset was that providing novice learners with feature descriptions generated by an expert might not be helpful for novices because novices may not be able to interpret the verbal descriptions or may not have the ability to perceive and extract the visual aspects to which the descriptions refer (unlike for artificial categories in which feature descriptions can be readily interpretable, such as “blue” and “square”). Consider that a feature provided for *pumice* was “rough textured volcanic glass.” A novice may have a poor idea of what “volcanic” glass is. As another example, for *conglomerate* a provided feature was “rounded fragments cemented together.” Though a novice may be able to understand the terms, it is not certain that the novice would be able to identify the perceptual qualities of a rock image that counts as “fragments” and “cemented together.” The results from the study as a whole support a more refined specification of this initial theoretical view, at least for rock category learning. With a standard presentation of features often found in educational settings, in which the features were presented separately but concurrently with a training instance (e.g., the feature description was placed below an image), feature descriptions failed to improve learning (Experiment 1). That is, an expert’s summary of characteristic features of a training instance (e.g., the feature description was placed below the image), feature descriptions failed to improve learning (Experiments 2 and 3). Within one successful rule learning model of ill-defined categories (like rocks), learners search for relatively simple rules that classify many of the instances and then learn exceptions (e.g., RULEX; Nosofsky et al., 1994). The feature provision condition essentially identified simple rules for the learners (e.g., “*obsidian* is glassy black and has shell-like fracture”) and, thus, learning would be

A second theoretical idea was that providing features might not benefit rock category learning because encouraging learners to focus on one or two dimensions might restrict the encoding of a rich array of perceptual information from rocks, which in turn could limit category learning. Previous work has shown that when novice learners are given no feature information and asked to rate similarity of rock images, their ratings reflect a complex feature space of at least eight perceptual dimensions (as reported in Nosofsky, Sanders, Meagher, & Douglas’, 2017 multidimensional scaling study). Moreover, additional studies found that rock classification performances after training (without feature descriptions) were well accommodated by an exemplar-based model that assumed equivalent attention to all eight dimensions in representing the exemplars (Nosofsky, Sanders, & McDaniel, 2017). It appears, however, that highlighting features did not result in functionally more impoverished encodings of the training items than when features were not provided. This conclusion is based on several observations. First, when features were highlighted, classification performance for trained items was high (over 80% correct) and virtually equivalent (even nominally better) to that found when features were not highlighted (see the old training performances in Figures 6 and 8). Second, when the test was delayed for 2 days after training, performance for trained items did not decline relative to when the test was immediate, and this pattern held when features were provided. Thus, classification performance for trained items was highly accurate and resistant to forgetting even in the feature-provision condition, suggesting that providing features did not produce impoverished encodings of the training items, at least under the present training conditions.

**Theoretical Explanation of the Feature Highlighting Benefit**

One general hypothesis is that providing features speeds several aspects of category learning. One idea developed was that feature provision could help learners more quickly identify and orient to dimensions along which rocks differ. If so, then the expectation would be that learners in the feature-provided condition would have a Head Start at encoding and differentiating rocks, which might benefit training initially (and possibly throughout). In opposition to this idea, the learning curves were not significantly steeper or systematically better in the early stages of training for the conditions in which features were provided than for conditions in which features were not provided (see Figures 5 and 7).

Another idea was that providing features would facilitate the processes by which learners identify characteristic features of each rock category. One such process might be hypothesis-testing. Within one successful rule learning model of ill-defined categories (like rocks), learners search for relatively simple rules that classify many of the instances and then learn exceptions (e.g., RULEX; Nosofsky et al., 1994). The feature provision condition essentially identified simple rules for the learners (e.g., “*obsidian* is glassy black and has shell-like fracture”) and, thus, learning would be
expected to benefit, at least early on in training and possibly throughout training, because the need for trial and error hypothesis testing would be minimized. For instance, other studies have shown that training conditions that facilitate learning a useful (but not perfect) rule produced steeper learning curves (Mathy & Feldman, 2009). Clearly, such a pattern was not associated with the category learning benefit of feature highlighting: Highlighting features did not change the rate of learning or the learning asymptote during training. Yet, providing features did produce better performance on a generalization test for untrained instances (Experiments 2 and 3), suggesting that providing features produced more accurate acquisition of category-level information.

Another potential process by which characteristic features gain ascendance is through learning appropriate attention weights to the category-relevant dimensions (e.g., ALCOVE; Kruschke, 1992). Providing features would be expected to facilitate setting effective attention-weights (e.g., in a connectionist implementation of an exemplar model; see Kruschke, 1992) or to allow the system to preset the weights (cf. Choi et al., 1993, for knowledge-based presetting of ALCOVE parameters). In particular, if observers give greater attention to category-level-relevant features, then such information would tend to be highly beneficial in generalizing to the novel transfer instances. The reason is that the novel instances are likely to vary in numerous idiosyncratic respects from the old training instances; instead, by definition, what they would share with the training items are the invariant category-level-relevant features.

Without feature highlighting, by contrast, instead of biasing their attention toward category-relevant features, learners presumably focused on more idiosyncratic features associated with individual training instances. In favor of this claim, in rock category learning experiments in which features were not highlighted, an exemplar model with unbiased attention weights across eight dimensions successfully accounted for rock category learning and generalization (Nosofsky, Sanders, & McDaniel, 2017). Idiosyncratic features would be less useful for supporting generalization, consistent with the worse test performance on untrained instances for conditions with no feature descriptions relative to conditions with feature highlighting.

It is a more open question whether greater attention to characteristic features of the category (stimulated by feature highlighting) would lead to enhancements in performance on the old training instances themselves relative to no feature highlighting. On the one hand, attending to the category-level-relevant features might also be expected to benefit training performance, because those features are also useful for discriminating between the training examples of the contrasting categories. On the other hand, for these complex objects, there are numerous idiosyncratic features associated with individual instances that an observer could also use to support initial learning. For example, an observer may remember that a particular rock with a red color patch in the upper-right location was an example of breccia, and this information would support performance on the training item. The detailed predictions from alternative models (selective attention to characteristic features vs. attention to a range of idiosyncratic features as well) would depend on factors such as the precise configuration of objects in the multidimensional similarity space and the extent to which attention can be shared across the multiple features that compose the objects.

The obtained patterns might also be consistent with the idea that highlighting characteristic features during training promoted a more substantial qualitative difference in category representations relative to when these features were not provided. Specifically, the current patterns share similarities with findings from the function learning domain in which learners show equivalent performances during training, but diverge in terms of accuracy in generalization (extrapolation; McDaniel, Cahill, Robbins, & Wiener, 2014; see also Hoffmann, von Helversen, & Rieskamp, 2014). The theoretical account for those findings is that some learners acquired exemplar representations during training, whereas others acquired an abstraction of the function rule; both representations supported good training performance but the abstract representation was more useful for transfer. A speculative possibility is that when features are provided, the rock category representations (perhaps based on exemplar storage) are enriched with some abstract summary information (e.g., possibly rule-like as in RULEX; Nosofsky et al., 1994; or prototype-like as when training involves substantial number of instances; Homa, Sterling, & Trepel, 1981; see also Regehr & Brooks, 1993); whereas, when features are not provided, a solely exemplar-based representation is favored (but see Appendix B for the learning-strategy self-reports which did not substantially diverge as a function of whether features were provided).

Specification of the nature of the category representations formed when features are provided relative to when features are not provided clearly awaits further work.

A final theoretical observation is that, if our interpretation regarding the role of feature highlighting on selective attention to category-level information is correct, then our results have implications for numerous formal models of category-learning. In particular, for a wide variety of such models, the construct of selective attention plays a central role (e.g., differential weighting of component dimensions in computing distances to prototypes: Reed, 1972; selective attention stretching and shrinking of the psychological similarity space in which the exemplars are embedded: Kruschke, 1992; Nosofsky, 1986). In a similar vein, clustering models (e.g., Love et al., 2004) and rule-based models (e.g., Bower & Trabasso, 1963; Levine, 1975; Nosofsky et al., 1994; Shepard et al., 1961) also place emphasis on selective attention.

As currently formalized, the key factors that are presumed to influence the patterns of dimensional selective attention in such models include the inherent salience of the component dimensions as well as the structure of the categories to be learned (e.g., attention-optimization in an exemplar-based model, Nosofsky, 1986, and trial-by-trial learning of optimal weights in a connectionist architecture, Kruschke, 1992; in rule-learning models observers learn to selectively attend to the dimensions that allow for the development of the most effective rules). The current article points to another crucial mechanism that can influence patterns of selective attention that has been somewhat neglected in these formalized models: explicit instructions and clues provided by the teacher. To provide a complete account of category learning, the manner in which such explicit and illustrated clues are integrated with other attention-learning mechanisms would need to be spelled out. Thus, the present work motivates an important direction for further development of a wide class of formal models of category learning.
Practical Implications

The current findings are in line with the multimedia principles of signaling and spatial contiguity (Mayer, 2002, 2005), and extend their reach to learning of perceptually based natural categories. These principles have previously been demonstrated in science concept learning (how braks work; how planes fly). In the present Experiments 2 and 3, but not Experiment 1, the verbal feature descriptions were positioned to be spatially contiguous to the exact aspect of the rock image to which the descriptions referred; furthermore, those aspects were explicitly signaled (by encircling part of the rock image). Relative to Experiment 1, the spatial contiguity of the feature description presumably reduced extraneous processing (e.g., scanning the rock images to try to find the feature) so that learners could devote more time to processing the pertinent parts of the rock (e.g., Mayer, 1989). Signaling the exact parts that were described in the feature description would help learners appreciate the particular perceptual qualities of the relevant dimension (e.g., Mautone & Mayer, 2001), unlike in Experiment 1 where learners could not take advantage of the feature descriptions placed below the images. In general, our results in conjunction with the literature on multimedia principles provide guidance for effective communication and placement of feature descriptions to enhance learning of natural perceptually based categories.

More specifically, our results have possible implications for instruction of natural categories in educational settings. First, as illustrated in Figure 1 feature descriptions (at least for rocks) in textbooks are provided as captions placed adjacent to the image (e.g., above or below). Our results (Experiment 1) suggest that this standard practice might be revisited in further work. Such work could take steps to achieve closer alignment (than did the present experiments) with actual textbook practice, such as using fewer example images and perhaps familiarizing learners with terms in the feature descriptions before learners viewing example rock images. Second, to the extent that instructors fail to signal relevant aspects of the images when providing descriptions of characteristic features of the categories, learning may not be benefitted. Instructors’ intuitions that providing feature descriptions will enhance learning of natural science categories—seem well founded, provided that those features are explicitly signaled in the example images or tokens used during instruction.

References

FEATURE PROVISION IN NATURAL CATEGORY LEARNING


(Appendices follow)
Appendix A

Material Overview

Feature descriptions for each category as well as the rock pictures used in the current experiments. Visit Open Science Frame Work (https://osf.io/9vg8m/#) for high resolution version of these images and the ones with feature highlighting that were used in Experiments 2 and 3.

**Amphibolite** – dark colored and weakly grained structure

**Breccia** – broken angular fragments cemented together

**Conglomerate** – rounded fragments cemented together

**Gneiss** – alternating dark and light colored bands with oriented crystals

**Granite** – light colored or reddish, interlocking coarse-grained crystals

**Marble** – gray to white crystalline material with darker swirls and veins

**Obsidian** – glassy black, scallop shell-like fracture

**Pegmatite** – uneven texture with large-size interlocking crystal components, can be greenish

(Appendices continue)
**Pumice** – rough textured volcanic glass with small holes

**Rock Gypsum** – silvery and chalky

**Sandstone** – sand-sized grains, often prominently layered

**Slate** – fine-grained, layered, homogeneous

---

**Appendix B**

**Learning-Strategy Questionnaire Results**

Results from the learning-strategy questionnaire responses. The reported learning strategy did not modulate the learning performance and the final test performance.

**Experiment 1**

We computed average strategy ratings for the two feature description (FD)-present conditions grouped together and the two FD-absent conditions grouped together. There was no significant difference in the strategy use ratings between the FD-present conditions ($M = 4.13, SD = 1.67$) and the FD-absent conditions ($M = 3.94, SD = 1.82$), $t(172) = 0.69, p > .05, d = 0.11$. Figure B1 shows the number of participants who reported each rating as a function of presence or absence of feature description.

**Experiment 2**

There was no significant difference in the strategy use ratings between the FD-present condition ($M = 5.18, SD = 1.50$) and the FD-absent condition ($M = 5.10, SD = 0.98$), $t(78) = 0.27, p > .05, d = 0.06$. Despite the similarity in the tasks between Experiments 1 and 2, participants’ self-reported strategy seemed to be more rule-based in Experiment 2. This is evident when comparing the extreme ratings (21% reported relying on memorization and 26% reported relying on rules in Experiment 1 versus 5 and 43%, respectively, in Experiment 2, see Figures B1 vs. B2) and when comparing the average ratings ($M = 4.03, SD = 1.75$ in Experiment 1 vs. $M = 5.14, SD = 1.26$ in Experiment 2, $t(252) = 5.07, p < .001, d = 0.73$). Figure B2 shows the number of participants who reported each rating as a function of presence or absence of feature description. We tentatively suggest that the strategy differences could reflect the differential populations from which participants were sampled in Experiments 1 (Amazon Mechanical Turk) and 2 (students from a highly selective private university). The key finding remains that providing features did not appear to change learners’ self-reported strategies relative to not providing features.

**Experiment 3**

Participants’ reported more rule-based strategies than in the prior experiments, and there were no extreme memorization-based strategies reported (see Figure B3). In interpreting this pattern, we caution that participants’ memory of how they tried to learn 2 days before providing the self-report is likely less accurate than in the previous experiments. Accordingly, perhaps participants relied primarily on how they tried to classify the items on the just given test task.

(Appendices continue)
Figure B1. The number of participants who reported each rating (where 1 indicates a strong preference for a memorization strategy and 7 indicates a strong preference for a rule-abstraction strategy) as a function of presence or absence of feature description in Experiment 1 postexperimental questionnaire.

Figure B2. The number of participants who reported each rating (where 1 indicates a strong preference for a memorization strategy and 7 indicates a strong preference for a rule-abstraction strategy) as a function of presence or absence of feature description in Experiment 2 postexperimental questionnaire.

Figure B3. The number of participants who reported each rating (where 1 indicates a strong preference for a memorization strategy and 7 indicates a strong preference for a rule-abstraction strategy) as a function of presence or absence of feature description in Experiment 3 postexperimental questionnaire.