Bogus Concerns About the False Prototype Enhancement Effect

Donald Homa, Michael C. Hout, Laura Milliken, and Ann Marie Milliken
Arizona State University

Two experiments addressed the mechanism responsible for the false prototype effect, the phenomenon in which a prototype gradient can be obtained in the absence of learning. Previous demonstrations of this effect have occurred solely in a single-category paradigm in which transfer patterns are assigned or not to the learning category. We tested the hypothesis that any extraneous variable potentially responsible for this effect, such as compactness varying with pattern distortion (Zaki & Nosofsky, 2004), may be functional in the single-category paradigm but not when multiple categories are available at the time of transfer. In the present study, subjects received a bogus or a real category learning phase, followed by a transfer test that required assignment into 1 or 3 prototype categories. The results showed that a minimal prototype gradient was obtained in the bogus conditions, with performance approaching chance levels when classification into 3 categories was required. In contrast, a substantial prototype gradient effect was found following learning. We conclude that the prototype gradient typically obtained following multiple-category learning is primarily driven by real learning and that the false prototype effect is itself an artifact of the single-category paradigm.

Keywords: prototype, gradient, categorization, paradigm, learning

A continuing dispute is how categorical knowledge is represented. Although the literature contains a myriad of models based on rules, features, neural nets, and boundaries, the main focus remains on variations of prototype and exemplar models. Prototype theorists (e.g., Homa, Dunbar, & Nohre, 1991; Smith & Minda, 1998) generally assert that the categories are constructed around abstracted central tendencies called prototypes, forged from the integration of experiences common to each category. Exemplar theorists (e.g., Nosofsky, 1988; Zaki & Nosofsky, 2001) contend that our categories are represented solely by the storage of the particular instances that are encountered, with abstractions playing no role. An issue critical to these two theories concerns the shape of generalization gradients: Exemplar theorists predict that generalization gradients should be steep around stored instances with a minimal gradient tied to the category prototype; prototype theorists predict a sharp gradient around the prototype and a minimal gradient centered on the training instances. These predictions arise because similarity to either the stored instances or the category prototype should critically determine subsequent transfer performance of novel instances.

Support for prototype models was claimed by Smith (2002), who analyzed the shape of gradients from numerous studies that used a particular paradigm first introduced by Knowlton and Squire (1993). In general, Smith demonstrated that variations of exemplar models must predict a flattened gradient around the prototype, whereas the actual data revealed a steeper gradient, as predicted by prototype theory. Homa, Proulx, and Blair (2008) explored the shape of gradients around training patterns and the prototype under variations of category size and number of exception patterns in each category, finding support for exemplar predictions only when category size was small or when larger category sizes contained numerous exception patterns; when category size was large and the number of exception patterns was small, sharp gradients around the prototype and flattened gradients to training instances were obtained. Zaki and Nosofsky (2004) acknowledged the steepness of the prototype gradient was a problem for exemplar models of classification and “could potentially indicate that some form of prototype abstraction is taking place” (p. 391).

However, Zaki and Nosofsky (2004) questioned whether the transfer gradient might arise, at least in part, from an artifact. What they termed the false prototype enhancement effect in dot pattern categorization refers to the demonstration that a prototype gradient effect can be obtained in the absence of any prior learning, arguing that an extraneous variable—compactness of the dot pattern forms—covaried with distortion. They suggested that the algorithm used to create distortions from a dot pattern prototype tended

This criticism applies to the distorted form stimuli used in the present study, because the distortion algorithm is similar, although line length rather than compactness is probably more germane. We believe this concern raised by Zaki and Nosofsky (2004) can be avoided by routine rejection of any prototype that has peculiar characteristics such as points that cluster near the centroid. Once the coordinate values exceed ± 10 units in the 50 × 50 grid, each distortion has a near 50% chance of moving closer, rather than farther, from the prototypical value. Regardless, our contention, repeated in the present study, is that this criticism vanishes in the multiple-category paradigm, because an extraneous variable such as compactness (or line length) cannot function as a categorically discriminative cue.

1 This criticism applies to the distorted form stimuli used in the present study, because the distortion algorithm is similar, although line length rather than compactness is probably more germane. We believe this concern raised by Zaki and Nosofsky (2004) can be avoided by routine rejection of any prototype that has peculiar characteristics such as points that cluster near the centroid. Once the coordinate values exceed ± 10 units in the 50 × 50 grid, each distortion has a near 50% chance of moving closer, rather than farther, from the prototypical value. Regardless, our contention, repeated in the present study, is that this criticism vanishes in the multiple-category paradigm, because an extraneous variable such as compactness (or line length) cannot function as a categorically discriminative cue.
to produce distortions that were less compact than the prototype, with this tendency increasing with level of distortion. In effect, the obtained accuracy of classification—prototype > low-level distortions > medium-level distortions > high-level distortions—was not due to a prior prototype abstraction process that guided subsequent transfer but rather to an artifact of pattern compactness, with compactness decreasing with pattern distortion.

In their experiments, subjects were exposed to a bogus learning phase and then instructed to assign transfer patterns to the “learned” category. In actuality, no patterns were shown, and the transfer test contained four different prototypes and either 40 random high-level distortions (Experiment 1A) or 10 random low- and 30 random high-level distortions (Experiment 1B). Results showed a slight but significant decline in rate of endorsement as distortion level increased. A similar result was obtained by Palmeri and Flanery (1999), who used a bogus learning phase followed by a transfer phase involving the category prototype, distortions of the category prototype, and random patterns.

The implication that a potential artifact may be responsible, wholly or partially, for the prototype gradient is theoretically important because this interpretation minimizes evidence taken as support for prototype models in general as well as casts doubt on a number of secondary, theoretical issues—for example, whether different memory mechanisms are responsible for classification and recognition (Knowlton & Squire, 1993), a conclusion emphasized by Zaki and Nosofsky (2004): “The presence of such false prototype enhancement effects has profound implications, therefore, for the interpretation of results from this classic and highly influential paradigm” (p. 391). However, there are a number of inconsistencies and concerns that must be addressed first before this conclusion is accepted: (a) Results with the bogus learning category paradigm have not been consistently replicated, and (b) the generation of a prototype gradient effect may be an artifact itself, a result of a transfer task involving the assignment of patterns into a single learned category. These two criticisms are addressed in order.

The original paradigm that later became the primary focus of the false prototype effect was introduced by Knowlton and Squire (1993), who investigated category learning by normal and amnesic subjects. In their task, amnesic and normal subjects were first exposed to 40 dot pattern stimuli, all high-level distortions from a single prototype, followed by a transfer test composed of new category members at various distortion levels from the prototype and random patterns (foils from different prototypes). On the transfer test, subjects were asked to indicate which patterns belonged to the same category as the training instances. A recognition test involving five dot pattern stimuli, tested after a 5-min delay, was also used. Knowlton and Squire obtained transfer performance that was similar for the amnesic and normal subjects—both groups revealed a similar prototype gradient on the transfer test. However, these same subjects differed dramatically on a recognition test, with the amnesic subjects showing poor discrimination between old and new patterns. The unimpaired classification performance, combined with the reduced recognition by the amnesic subjects, led Knowlton and Squire to conclude that abstraction mechanisms engaged an implicit memory mechanism (which was intact in the amnesic patients) whereas recognition was based on a declarative system (which was damaged in the amnesic patients).

Disputes over this interpretation have come from studies that have not always produced consistent results. Knowlton and Squire (1993) also included a bogus learning control condition and found that transfer performance was significantly reduced by bogus training. In contrast, Palmeri and Flanery (1999) found a significant prototype gradient effect for the bogus learning condition (subjects were given the ruse that patterns had been subliminally presented during a word-identification task) but not following actual learning.2 Zaki and Nosofsky (2004) found a significant gradient in their bogus full condition but not their bogus subset condition, even though their model predicted a significant gradient in each condition.3 In addition, three of the four experiments run by Zaki and Nosofsky reported results in terms of a judgment on a rating scale rather than accuracy, rendering comparison to previous studies unclear. Finally, the number of subjects needed to produce a significant gradient effect following bogus learning was abnormally large, compared with most studies in categorization—179 in one condition, 123 in a second, 198 in a third, and 185 in a fourth. Their results suggest that the differences obtained across distortion levels are real but small in magnitude.

A more critical concern is that the gradient effect following bogus learning was obtained in what may be considered a single-category transfer paradigm. That is, in Knowlton and Squire (1993), as well as in the studies claiming a bogus prototype effect, subjects were required to assign transfer patterns if they believed the patterns belonged or not to the learned (or bogus) category. Although the single-category transfer paradigm was widely used in the older category literature that explored rule-learning with well-defined stimuli (e.g., Bourne, 1966; e.g., Category A might contain only objects that are red and square, and objects that failed this rule were to be called “not A”), the more recent literature on category learning rarely has used a single-category transfer paradigm. Rather, the typical paradigm presents subjects with numerous patterns that belong to two or more categories in the learning phase, followed by a transfer test in which the subject is required to assign patterns to the learned categories. This distinction—whether transfer occurs to one or multiple categories—is, we believe, critical to the issue of whether the gradient is an artifact of stimulus construction or a real outcome produced by prior learning.

2 Palmeri and Flanery (1999) reported a statistically significant effect of distortion in their learning condition but, as is clear from the degrees of freedom in their reported $F$ statistics, this analysis included classification of the random patterns, which were classified far worse than the category patterns. If analysis is restricted to the reported difference on classification accuracy of the prototype and high-level distortions (63.8% vs. 62.0%, respectively), then distortion level is not statistically significant, based on their reported $MS$ error terms.

3 In their Experiments 2A and 2B, Zaki and Nosofsky (2004) reported a significant interaction between condition (full and subset) and item type (prototype, low distortion, high distortion) but failed to report whether the main effect of item type was significant in the subset condition in either experiment. The slight difference among the three item types (less than 3% between the prototype and high distortions in Experiment 2A and less than .10 on a 5-point scale in categorization judgment in Experiment 2B) and the outcomes in each experiment (high distortions were classified more accurately than low distortions) are inconsistent with the predictions of their own model fits (see Figures 3 and 4).
The concern is not whether the false prototype effect can be generalized from a one-category to a multiple-category transfer paradigm; rather, the claim is that the potential artifact raised by Zaki and Nosofsky (2004) is invalid except when a one-category transfer paradigm is used. In particular, it is unclear how an extraneous variable like “compactness” could produce a prototype enhancement effect when more than one category is learned and discrimination among these categories is required at transfer. Specifically, compactness by itself could not function as a discriminative cue when multiple categories are considered. To make this point clear, suppose you knew nothing about the different types of animals on a particular planet except for one fact—you are told the best examples of each of three species are the reddest, with nonexamples being less red in color. If asked to separate the best members of these three species from the nonexamples, you could do this with some accuracy. However, if you were asked to sort the members of these three species into their respective groups, sorting by redness alone would produce chance performance. The former case—separating the members from the nonmembers—corresponds to the one-category paradigm used by Zaki and Nosofsky. The latter case—sorting the members of these three species into their appropriate groups—was not explored by Zaki and Nosofsky but is investigated in the present study. Because the vast majority of studies in categorization, from the seminal studies by Fisher (1916) and Hull (1920) to the modern study of ill-defined categories (Posner & Keele, 1968, 1970), have used a multiple-category learning phase followed by transfer into these same categories, the finding of a false prototype effect that was manifested only in a single-category paradigm would render meaningless the criticism of Zaki and Nosofsky.

In their discussion, Zaki and Nosofsky (2004) acknowledged that the false prototype effect might be confined to the single-category paradigm:

We should acknowledge that such effects of “compactness” and “goodness” are likely to be most pronounced in the present type of single-category paradigm. . . . It is an open question whether similar stimulus specific effects may also play some role in multiple-categorization paradigms. (footnote 3, p. 398)

The present study addresses precisely this issue.

In Experiment 1, subjects were exposed to either a real or a bogus learning phase, followed by transfer to a single category or to multiple categories. In the bogus category conditions, patterns generated from three different prototypes were presented so briefly (and followed by a noise mask) that nothing could be seen except for a brief flash. At the time of transfer, the bogus one-category subjects were told the patterns actually came from a single category and that they should assign the transfer patterns into either the category seen subliminally or a junk category. The bogus three-category subjects received similar instructions, except they were told that the patterns actually came from three different categories. On the transfer task, they were instructed to assign patterns into these three categories (A, B, C) or into a junk category if the pattern belonged to none of the subliminal categories. The real learning condition required that subjects first learn to classify the patterns into three categories prior to transfer. The composition of the transfer task was identical for all conditions, composed of three prototypes, low-, medium-, and high-level distortions of these prototypes, and random (unrelated) patterns. As a consequence, the only difference among the conditions was in the prior learning phase, either bogus learning with one or three categories or actual learning with the three categories.

Therefore, if the concerns of Zaki and Nosofsky (2004) are extendable to multcategory learning using pattern distortions, then the generalization gradient at the time of transfer should be undiminished by the number of categories available at the time of test, regardless of learning. However, if the gradient is minimal or absent following bogus learning for a three-category case and robust following the learning of three categories, then the demonstration of a false prototype enhancement effect is likely confined to (or an artifact of) the one-category paradigm. In our discussion, we examine more fully the factors that are responsible for producing transfer gradients, including why a minimal, nonzero gradient might be expected following bogus learning for the three-category case that has little to do with an artifact of stimulus compactness or goodness.

**Experiment 1**

In Experiment 1, subjects were exposed to either a bogus learning phase or a real learning phase. For the bogus learning condition, 30 patterns were shown briefly (10 ms) and followed by a noise mask. This procedure minimized the encoding of pattern information and, therefore, category information. A postexperiment questionnaire confirmed that subjects rarely reported anything other than a “blink” or, at most, a pattern segment. This procedure did have the effect, therefore, of convincing the subjects that something had been shown.4 This bogus learning was followed by a transfer test to one or three categories. The transfer test contained 93 patterns, 15 patterns each from three prototype categories (five low, medium, and high distortions each), the three category prototypes, and 45 random patterns (high-distortion patterns from 45 different prototypes). Subjects assigned the transfer patterns into Category A, B, C, or junk for the three-category transfer; for the one-category transfer, subjects assigned the transfer patterns into Category A or junk. For the real learning condition, subjects were exposed to a real learning phase involving the same three categories that preceded the transfer test. During the learning phase, subjects were shown 30 high-distortion patterns (the same patterns presented in the bogus conditions) for five study/test learning trials, where 10 patterns belonged to each of the three prototype categories. This learning phase was followed by a transfer test identical to that for the bogus three-category condition above.

In the bogus one-category condition, subjects were told to assign the patterns to Category A or junk. In the bogus condition involving three categories at the time of transfer, subjects were told to assign the patterns to the Categories A, B, C, or junk (identical to the learning subjects). Subjects were free to use the labels as they

---

4 Although this issue was not addressed by Zaki and Nosofsky (2004) or Palmeri and Flanery (1999), our pilot data indicated that some subjects were suspicious whether anything was actually presented in a bogus condition that lacked any patterns being presented. To minimize this concern, we presented the patterns very briefly (10 ms), followed by an immediate noise mask. This rendered an impression that something “popped” but was otherwise not discernible. We did include a postquestionnaire that confirmed this, including a test to ask subjects to draw any pattern, which no subject could.
wished, with the sole stipulation that they adopt a strategy where they sorted together under a common label those patterns they felt belonged together. At the conclusion of the study, scoring was optimized for each subject by counting, as the subject labels, whatever labels produced the highest classification score (see also Homa & Cultice, 1984, where a similar strategy was used in a task that contrasted learning with and without feedback during learning).\(^5\)

**Method**

**Subjects.** Subjects were 83 Arizona State University undergraduates who received course credit in their introduction to psychology classes for participation in the experiment. There were 27 subjects in each bogus category condition (bogus one-category and bogus three-category) and 29 in the real learning condition (learn three-category). One subject was dropped from the learning condition because he placed 85 of the 93 patterns into the junk category.

**Materials and apparatus.** Members of three form categories served as stimuli and have been described previously (Homa, 1978). In brief, a form category is created by first generating a random nine-dot configuration within a 50 × 50 grid and then connecting the dots with lines. This pattern is designated as the category prototype; different members of this category are then generated by statistically moving each of the dots of the prototype.

A statistical distortion algorithm is applied to each point in the category prototype, with the amount of dot displacement determining the distortion level of a pattern. For high-level distortions, each dot is displaced, on the average, by about 4.6 Euclidean units from each corresponding dot of the prototype. The topography of a category can be thought of as a sphere with the prototype located in the center, and with the high-level distortions on the surface of the sphere having a radius of 4.6 units. The average distance between any two learning patterns is about 7.0 Euclidean units, with no two patterns being closer than 4.5 units. Only high-level distortions were used in the learning phase. On the transfer test, new low-level distortion, medium-level distortion, high-level distortion, and random patterns were used. For the low- and medium-level distortions, each dot is displaced about 1.2 or 2.8 Euclidean units, respectively, from the corresponding dot of the prototype. Random patterns are statistically unrelated to the three prototypes and range from 10 to 15 Euclidean units from any other pattern.

Subjects were run individually in a small cubicle with all displays on a computer screen; programming, timing, and data recording were controlled by E-Prime 1.2 (Schneider, Eschman, & Zuccolotto, 2002). In the bogus one-category condition, the subject pressed either the A or J keys on the transfer test to signify a pattern in those categories. In the bogus three-category and real learning conditions, the subject pressed either the A, B, C, or J keys to signify a pattern in those categories. A brief questionnaire was administered following completion of the experiment to assess what, if anything, was observed in the learning phase.

**Procedure.** In each bogus learning condition, the subject engaged in a “subliminal training” phase in which a series of patterns were briefly shown for 10 ms each. A total of 30 patterns were presented in a random order, followed by an immediate visual noise mask. The 30 patterns were composed of 10 high-level distortions from three different prototypes.

Following this phase, each subject in the bogus one-category transfer task was told that the learning set actually came from a single category and that his or her job was to assign each transfer pattern into either Category A if it came from the subliminal category or Category J (for junk) otherwise.\(^6\) Each subject was then presented with 93 patterns, 16 from each of three prototype categories and 45 random patterns. The 16 patterns from each category were composed of the prototype and five low-, five medium-, and five high-level distortions; the 45 unrelated patterns were 45 random patterns unrelated to the three prototypes.

For subjects in the bogus three-category condition, the subject instructions and learning phase were identical to those used in the bogus one-category condition. At transfer, the subject was told that the learning set actually came from three categories and that his or her job was to assign each transfer pattern into either Category A, B, or C, or junk. Immediately following learning, the subject observed the same 93 patterns and composition as the one-category transfer condition. No feedback was provided following each response, and the procedure was self-paced. The order of patterns in the learning and transfer phases was randomized separately for each subject.

In the learn three-category condition, the procedure was similar to that for the bogus conditions except that the learning phase was real and only a three-category condition was used. There were five learning blocks, each having a different category order. For Learning Block 1, the subject saw the patterns from three prototype classes, blocked by prototype class (labeled A, B, and C for each subject), with the pattern label appearing in the upper right corner for 2 s. This was followed by a blank screen shown for 1 s, followed by the next pattern from this group. The patterns were blocked by prototype but otherwise shown in a randomized order. For example, the 10 patterns from one prototype class were labeled A1, A2, ..., A10; the 10 patterns from the second prototype class were labeled B1, B2, ..., B10; and the 10 patterns from the third prototype class were labeled C1, C2, ..., C10. For Learning Block 1, the subject might see, in order, A1, A7, A3, A4, A6, A2, A10, A8, A8, A7, A1, A2, ..., A10, B1, B2, ..., B10, C1, C2, ..., C10.

---

\(^5\) For example, suppose a subject assigned the 16 patterns from Prototype 1 as follows: 3 in A, 7 in B, 2 in C, and 4 in junk. For Prototype 2, 8 were put in A, 3 in B, 4 in C, and 1 in junk. For Prototype 3, the assignments were 2 in A, 4 in B, 5 in C, and 5 in junk. Maximizing this subject’s performance would assume that Prototype 1 was called B, Prototype 2 was A, and Prototype 3 was C, resulting in an overall performance of 7, 8, and 5 patterns correctly assigned. For subjects in the bogus one-category condition, no assumptions were necessary, because the subject need only separate the category from the random patterns.

\(^6\) In Palmeri and Flanery (1999), subjects were told that the “subliminal” patterns belonged to a single category; in Zaki and Nosofsky (2004), subjects were told that the patterns came from five categories (Experiments 1A and 1B) or a single category (Experiments 2A and 2B). In each experiment, subjects were then asked to assign transfer patterns to the subliminal category (or categories) or not. In our bogus one-category condition, subjects were told that the subliminal patterns belonged to a single category. Because it is unclear whether the number of categories supposedly presented in the bogus phase matters, at least in the bogus one-category condition, we also replicated this procedure with the sole exception that subjects in the bogus one-category condition were told that the patterns actually came from three categories. The resulting transfer performance was virtually identical to that in the bogus one-category condition reported here.
A5, and A9, each for 2 s with the label A appearing in the upper right corner. This was followed by the B and C sets, which were also blocked and randomized within the block, but with the labels B and C shown in the upper right corner. A test block followed each learning block. On each test, all 30 patterns were presented in a random order without the category label. After the subject’s “A,” “B,” or “C” response, the correct label was shown again in the upper right corner for 1 s. Following feedback, the screen went blank for 1 s, and then the next pattern was shown. Performance on the test trials was self-paced, with most responses occurring within 2–3 s.

Learning Blocks 2–5 were identical to Learning Block 1 with the sole exception that the order of prototype class corresponding to A, B, and C was randomly changed and the patterns within each prototype category were randomized again. That is, if the patterns were shown in the order A, B, C on Learning Block 1, Learning Block 2 might present these patterns in the order B, C, A, and so forth. After the fifth learning/test block, the transfer test was given. The transfer test for the real learning condition was identical to that of the bogus three-category condition—the 93 patterns were shown in a random order, and the subject selected Categories A, B, C, or junk to classify the patterns. No feedback was provided to the subjects at any time.

**Design.** A mixed design was used, with condition (bogus one-category, bogus three-category, learn three-category) as the between-subjects variable and type of transfer item (prototype, new-low, new-medium, new-high, random) as a within-subject variable.

**Results**

**Learning.** Figure 1 shows the mean error rates of learning across the five trial blocks for the three-category learning condition. The main effect of learning block (1, 2, 3, 4, 5) was significant, $F(4, 112) = 33.10$, $MSE = 7.83$, $\eta^2 = .542$, $p < .001$. Because errors decreased across training blocks, significant learning did occur.

The transfer test for the real learning condition was identical to that of the bogus three-category condition—the 93 patterns were shown in a random order, and the subject selected Categories A, B, C, or junk to classify the patterns. No feedback was provided to the subjects at any time.

**Transfer: Gradient test.** The initial analysis focused on the shape of the gradient for transfer patterns at varying levels of distortion (low, medium, high) from the prototype. Figure 2 shows the mean correct classification rates for the bogus one-category, bogus three-category, and learn three-category conditions; also shown are the classification rates for the prototypes. The effect of condition was significant, $F(2, 80) = 13.35$, $MSE = 20.61$, $\eta^2 = .250$, $p < .001$, as was the effect of distortion, $F(2, 160) = 41.50$, $MSE = 3.24$, $\eta^2 = .342$, $p < .001$. The Condition × Distortion interaction was also significant, $F(4, 160) = 10.61$, $MSE = 3.24$, $\eta^2 = .210$, $p < .001$. A subsequent analysis revealed that the two bogus conditions (bogus one-category and bogus three-category learning) did not differ from each other, $F(1, 52) = 0.07$, $p > .20$, but both bogus conditions differed from the learning condition ($p < .05$ in each case). The bogus conditions did not interact with distortion. $F(2, 104) = 1.27$, $p > .20$. In general, it appeared that performance decreased across distortion level and that the learn three-category condition was superior to the bogus conditions. Because this latter interaction was significant, separate analyses were done for each condition.

For each of the bogus one-category, bogus three-category, and learn three-category conditions, the main effect of distortion was significant, $F(2, 52) = 5.18$, $MSE = 3.04$, $\eta^2 = .166$, $p < .01$; $F(2, 52) = 4.96$, $MSE = 2.91$, $\eta^2 = .160$, $p < .05$; $F(2, 56) = 48.07$, $MSE = 3.72$, $\eta^2 = .632$, $p < .001$, respectively. There was an 8%–10% change across low-, medium-, and high-distortion levels for the bogus conditions; in the three-category learning condition, the gradient was steeper, resulting in a 33% drop.

**Transfer: Prototype classification.** Classification accuracy of the category prototype was .84 for the learn three-category condition, versus .62 and .62 for the bogus one- and bogus three-category conditions, $F(2, 80) = 5.96$, $MSE = .701$, $\eta^2 = .13$, $p < .01$. A Bonferroni test subsequently confirmed that the learn three-category condition differed from each bogus condition ($p < .05$), with the two bogus conditions not differing from each other ($p > .20$).

**Transfer: Junk classification and a compositional analysis.** Classification of most transfer patterns, including random patterns, into the learned category would give the appearance of highly accurate transfer performance. It is, therefore, instructive to measure how often random (unrelated) patterns were correctly classified as “junk” and how often they were incorporated into the category (or categories) represented in the learning phase. For the three-category conditions, we can also measure how often intrusions from other categories occurred. Table 1 shows the mean number of random patterns correctly classified into the junk category; also shown are category intrusions and a compositional analysis (Homa, Burrel, & Field, 1987), explained shortly.

Correct assignments of unrelated patterns into the junk category for the bogus one-category, bogus three-category, and learn three-category conditions were .685, .285, and .724, respectively, $F(2, 80) = 51.73$, $\eta^2 = .564$, $p < .001$. Subsequent tests revealed that the bogus one-category and learn three-category conditions did not differ from each other ($p > .20$), with both conditions exceeding the performance of the bogus three-category condition ($p < .05$). Thus, the bogus one-category condition produced significant discrimination between the category members and unrelated patterns, but this discrimination was considerably reduced for the bogus three-category condition.

![Figure 1](image-url)
The compositional analysis provides a sensitive index of category knowledge by identifying the kinds of information assigned to a category, including correct assignments as well as erroneous category intrusions and unrelated stimuli. The purity measure (which is separate from the hit rate) reflects the proportion of information assigned to a category that is correct. For example, in the bogus three-category condition, the mean number of category patterns correctly assigned to each category was 0.62 prototypes, 2.64 low-level distortions, 2.20 medium-level distortions, and 2.25 high-level distortions, or a total of 7.71 patterns correctly assigned to each category out of 16 possible. However, each category also contained 5.61 category intrusions and 10.80 unrelated patterns, resulting in an average of 24.12 patterns classified into each of the three categories. Of these, 7.71 were correct assignments, resulting in a purity value of .320. In other words, of the patterns assigned to each category, 32% were correct, and 68% were intrusions from other categories or unrelated patterns. For comparison, random assignment of the category patterns into the three categories (with an equal number of random patterns into each category) would result in a purity value of .25.

The purity value for the bogus one-category condition was .625; random assignment of the entire transfer set, with an equal number of patterns assigned to Category A and the junk category, would produce a purity value of .516. For the learn three-category condition, purity was .653 (again, chance with three categories would produce a purity value around .25). A comparable analysis can be made for the junk category. A high purity value for junk would indicate that most patterns assigned to junk were, in fact, patterns unrelated to the learning categories. For the bogus one-, bogus three-, and learn three-category conditions, junk purity was .559, .609, and .779, respectively. This indicates that in both bogus conditions, random patterns were discriminated from category patterns with limited success (a .50–.50 split would indicate no discrimination), with the two bogus conditions functioning similarly. In contrast, the learn three-category condition was far more successful in this discrimination.

Discussion

Experiment 1 demonstrated that prior category learning produced a sizable gradient across pattern distortion, with performance decreasing about 35% from low- to high-level distortions. The magnitude of this gradient across pattern distortion mirrors what is typically obtained in categorization research (e.g., Homa & Little, 1985; Homa et al., 2008). In contrast, bogus learning produced a minimal gradient—about 8%–10%. When transfer required the subject to discriminate among the categories as well (bogus three-category condition), the generalization gradient became erratic across distortion level, and overall, classification accuracy was poor. The compositional analysis, which provides an index of the kinds of information assigned to a category, indicated that, following bogus learning, purity dropped from over 60% to less than 35% when subjects were required to discriminate among the categories. This is because most patterns classified into the same category were intrusions either from other categories or random patterns. In effect, any potential extraneous variable that might be available to discriminate category members from random patterns could not be effectively used in the multiple-category condition. Furthermore, the ability to discriminate category from random patterns was poor, with about 80% of all random patterns incorporated into the categories.

It is important to realize that performance in the bogus three-category condition is probably a more accurate measure of the level of category knowledge acquired during transfer than that in the bogus one-category condition. That is, if subjects are asked to discriminate category from random patterns following bogus learning, the appearance of slight, but significant, performance is obtained. However, this apparent level of knowledge is somewhat illusory, because this same training results in subjects who then demonstrate little ability to accurately sort patterns into the multiple categories. In fact, when we ran about 100 pseudo-subjects who randomly sorted the transfer stimuli and then maximized their performance, the obtained accuracy (40%) was only slightly worse than that obtained by subjects in the bogus three-category condition (48%).

Because transfer following pattern learning usually entails transfer to multiple categories, not just one, the concerns expressed by Zaki and Nosofsky (2004) mostly vanish. This is not to deny that some minimal, nonrandom learning can occur during transfer, only that its contribution is slight and far less than what is typically obtained following actual learning.

Experiment 2

Experiment 2 was a replication of the learning condition in Experiment 1 with the exception that the junk patterns were removed from the transfer test. All researchers who use a one-category paradigm typically include random patterns (e.g., Knowlton & Squire, 1993; Palmeri & Flanery, 1999; Zaki & Nosofsky, 2007), and their inclusion is necessary to determine whether subjects can separate category from random patterns in a one-category task. However, the inclusion of so many random patterns at the time of transfer is not typically done when real learning is followed.
by transfer to multiple categories. It is, therefore, necessary to show that the large gradient effect obtained in the three-category learning condition of Experiment 1 is not diminished when unrelated patterns are excluded from the transfer phase. We hypothesized that the removal of numerous random patterns in transfer might affect the transfer gradient to a slight degree, because the inclusion of random patterns can influence decisional processes at the time of transfer separate from acquired categorical knowledge (Homa et al., 1987). However, the resulting gradient, if primarily driven by learning, would be similar to that obtained in Experiment 1 and, therefore, still far greater than that obtained in the bogus learning conditions.

Method

Subjects. A total of 30 subjects were used, drawn from the same pool as in Experiment 1, except all of these subjects were in the three-category condition. Data from one subject were deleted for failure to show learning across the five learning blocks. Therefore, all analyses were based on the results from 29 subjects.

Materials and apparatus. Subjects were run according to the same guidelines and stimuli as in Experiment 1.

Procedure. The procedure was identical to that of Experiment 1, except that there were no random patterns in the transfer test, and patterns were assigned to Category A, B, or C.

Results

Learning. Figure 3 shows the error rates of learning across the five trial blocks. The main effect of learning block was significant, \( F(4, 112) = 40.05, MSE = 7.51, \eta^2 = .589, p < .001 \). The initial and terminal error rates mirrored those of Experiment 1.

Transfer: Prototype and distortions. Figure 4 shows the three-category real learning condition without junk patterns in the transfer. For the three-category real learning condition without random patterns in the transfer, the effect of item type was significant, \( F(3, 84) = 53.95, MSE = 3.82, \eta^2 = .658, p < .001 \). Because the procedure and patterns for Experiment 2 were virtually identical to those used in Experiment 1, an analysis that directly compared the three-category learning conditions (with random patterns in Experiment 1 and without random patterns in Experiment 2) was performed. A mixed analysis of variance was done, with learning condition (three-category learning with and without junk) as the between-subjects variable and item type (prototype, low, medium, high) as the within-subject variable. This analysis revealed that the effect of condition was not significant, \( F(1, 56) = 0.204, MSE = 24.40, \eta^2 = .004, p > .20 \), nor was the Condition \( \times \) Distortion level interaction, \( F(3, 168) = 0.53, MSE = 4.60, \eta^2 = .009, p > .10 \), but the effect of distortion level was significant, \( F(3, 168) = 77.04, MSE = 4.60, \eta^2 = .58, p < .001 \).

Transfer: Gradient test. An analysis confined to the pattern distortions revealed that the main effect of distortion from the prototype (low, medium, high) was significant, \( F(2, 56) = 69.71, MSE = 3.00, \eta^2 = .713, p < .001 \). The two learning conditions (three-category learning with and without random patterns) did not differ from each other, \( F(1, 56) = 0.46, MSE = 19.09, \eta^2 = .01, p > .20 \), nor was the Condition \( \times \) Distortion interaction significant, \( F(2, 112) = 0.49, MSE = 3.21, \eta^2 = .01, p > .20 \). However, the effect of distortion was significant, \( F(2, 112) = 119.89, MSE = 3.21, \eta^2 = .68, p < .001 \).

A comparison of the transfer performance for all conditions of this study—the three conditions from Experiment 1 and the three-category learning without random patterns condition from Experiment 2—is shown in Figure 5. Inspection of Figure 5 reveals the close similarity in performance for the two learning conditions and the dramatically worse performance for the two bogus conditions (which also produced similar results to each other).

General Discussion

In their study, Zaki and Nosofsky (2004) claimed the prototype enhancement effect—the tendency for classification accuracy at transfer to be highest for the low-level distortions, intermediate
for the medium-level distortions, and lowest for the high-level distortions—typically obtained in categorization research is, partially or totally, a pseudo-outcome, an artifact of the compactness of the prototype in relation to the other distortion levels. This issue is important because generalization gradients centered on the category prototype, rather than particular learning instances, have typically been used as support for a prototype abstraction process (e.g., Smith & Minda, 2002). However, limitations in the results of Zaki and Nosofsky should be noted. First, the effect they obtained was of small magnitude, not consistently obtained in their own experiments or by other researchers, and required upwards of 200 subjects to obtain statistical significance. Second, Zaki and Nosofsky reported confidence data based on a rating scale in three of their four experiments, rather than on accuracy rates, and therefore, direct comparison to other studies cannot be easily made. Third, and most important, Zaki and Nosofsky used only a one-category paradigm, and therefore, their results cannot extend to the learning of and transfer to more than a single category. Compactness, or any other extraneous variable that is correlated with construction of the prototype or its distortions, might work to explain transfer performance with a single category but cannot be logically applied when more than one category is used. That is, if the prototypes for all categories are more compact that their instances, then compactness alone cannot function as a distinctive cue to discriminate among these categories.

In the present experiment, the results revealed a slight but significant gradient around the prototype in both the bogus one-and bogus three-category conditions, although the gradient for the bogus three-category condition was erratic, with higher classification for the high-level distortions than the medium-level distortions. In contrast, the magnitude of this gradient was far greater following learning—about 35%—in both Experiments 1 and 2. Therefore, it is clear that the magnitude and shape of this gradient is minimally driven by potential artifacts like stimulus compactness and primarily driven by category learning. In effect, the pseudo-prototype effect obtained by Zaki and Nosofsky (2004) is itself a pseudo-effect of the single-category paradigm. Because the vast majority of studies in human categorization have used paradigms involving multiple categories in learning and transfer, the concern that a substantial part of the prototype gradient is due to an artifact is unwarranted.

The results of Experiment 2, in which category learning was followed by a transfer test lacking the random patterns, largely mirrored the results of Experiment 1. This result is important because random patterns are used when transfer is considered for the one-category condition, where they are necessary, but they are infrequently used when multiple categories are explored. However, it is clear that the magnitude of the gradient around the category prototype is unaffected by whether random patterns are used or not, at least when three categories are learned. As such, our conclusion that the gradient around the category prototype is due primarily to actual learning is undiminished by whether the transfer set contains random patterns or not.

Three related methodological points deserve comment. First, the claim that some learning may occur during transfer (Zaki & Nosofsky, 2004, 2007) is likely true, and therefore, nonchance performance following bogus training is to be expected. Learning during transfer is functionally indistinguishable from schematic concept formation (Evans, 1967; Smallwood & Arnoult, 1974). In schematic concept formation, researchers explore whether categories can be learned in the absence of external, verbal feedback. In both cases—a transfer test provided without feedback, and learning in the schematic concept formation task—a series of patterns is presented, the subject is required to make a classification judgment, and no feedback is provided. The sole difference is that subjects in the schematic concept formation task explicitly attempt to learn; in most transfer tasks, any learning that occurs might be done implicitly. What has been found in the schematic concept formation task is that significant, but usually minimal, learning occurred. For example, in the free-sorting task used by Evans and Arnoult (1967), only 20% of the subjects performed better than chance. In other studies (e.g., Aiken & Brown, 1971; Brown & Evans, 1969; Smallwood & Arnoult, 1974), learning increments of less than 10% were obtained. Homa and Cultice (1984) explored

![Figure 4](image1.png)

Figure 4. The mean proportion correct classification rates on the transfer test for the three-category real learning condition, as a function of distortion level from the prototype (Pro), without junk patterns in the transfer, Experiment 2.

![Figure 5](image2.png)

Figure 5. The mean proportion classification rates on the transfer test for all four conditions from Experiments 1 and 2 (bogus one-category learning [B1-C], bogus three-category learning [B3-C], three-category real learning without junk [L3-NJ], and three-category real learning with junk [L3-J]) as a function of item type and distortion from the prototype (Pro).
learning with and without feedback in eight conditions, finding that learning occurred when patterns were low-level, medium-level, or a mixture of distortions, although the rate and terminal level of learning was far greater when feedback was provided. However, when patterns were exclusively high-level distortions, little evidence of learning was found; that is, the learning curve across eight learning blocks was flat, although above chance. Therefore, in the typical category learning paradigm, the incorporation of mixed distortion levels or the mixing of multiple copies of the category prototype in the transfer test can produce some apparent learning following a bogus learning phase. However, as the present results clearly demonstrate, the magnitude of the generalization gradient around the category prototype, and the overall level of performance, is far greater when real, rather than bogus, learning precedes transfer, at least when multiple categories must be discriminated.

Second, the slight generalization gradient obtained in the bogus three-category condition is undoubtedly elevated by the maximization of scoring used in the present study. When we ran a Monte Carlo simulation on 100 pseudo-subjects, with the entire transfer set randomized, maximizing performance resulted in an overall classification rate of .40, only slightly below the .48 obtained in Experiment 1. The addition of a single constraint—assume that the subject classifies two to three low-level distortions from the same prototype together and then randomly assigns the remaining patterns—produces an overall classification rate that matches the level obtained by the bogus three-category condition. This, combined with the compositional analysis, reveals that the transfer performance was approaching chance levels in the bogus three-category condition. Specifically, the “best” categories produced by our subjects resulted in a purity value of 32%. That is, only 32% of all patterns classified together belonged to the same prototype category; the remaining 68% of all patterns classified within each category came from category intrusions and random patterns. This can be contrasted with the bogus one-category condition, which had a purity value of .62. This latter value suggests that subjects could separate category from noncategory patterns but had little success in separating among categories.

Third, the study of human concepts cannot be adequately investigated in the single-category paradigm. This is not to deny that preliminary and even provocative findings can be generated by this paradigm (e.g., Ashby & Maddox, 2005; Knowlton & Squire, 1993), which can be further explored by converging operations. But what is lacking in the single-category paradigm are stimulus cues that would permit discrimination between the studied category and any other category of concern, or what Gibson and Gibson (1955) called distinctive features. Showing the subject a handful of apparently unrelated objects is like Socrates demonstrating a point to Menon and claiming, “These are all Manks”; it should leave Menon asking, “But what do non-Manks look like?” There exists a virtual infinity of features that could, potentially, discriminate Manks from non-Manks in this situation, and our hypothetical subject would be left to his or her own devices upon which to base his or her category decisions on a subsequent transfer test. This was nicely illustrated by Zaki and Nosofsky (2007), who demonstrated that the composition of the transfer set—in this case, multiple copies of a high-level distortion and numerous low-level distortions of this same pattern—can, following one-category learning, dramatically alter performance. We disagree, however, that this demonstration requires that researchers “reevaluate whether past results involving the steepness of the typicality gradient are, in fact, due to the abstraction of a prototype from the training instances” (Zaki & Nosofsky, 2007, p. 2090). Rather, the results of the present study make it clear that the single-category paradigm lends itself to data outcomes that need not be mirrored in a paradigm requiring discrimination among multiple categories.

We believe these concerns touch upon a broader issue that needs to be considered. Any comprehensive theory of human categorization should consider the potential data space shaped by learning variables (Homa, 1984), especially when generalization gradients are of major concern. It has been known for some time that small category sizes produce shallow generalization gradients, whereas large category sizes produce steep generalization gradients (Homa, 1978). But even the effect of category size, and its role in the shaping of generalization gradients, can be muted when only two categories are learned (Homa & Chambliss, 1975), presumably because the subject need learn only one category, and the second is learned by default; learning more than two categories, or providing an optional third “none” category, reestablishes the gradient (Homa & Hibbs, 1978). The integrity of a database in human categorization experiments is further compromised by the single-category paradigm, where the default “noncategory” is undefined and likely shaped by the composition of the transfer set and the decisional whims of the subject. Selection of a single category for study while holding category size constant and then withholding feedback during a transfer task composed of patterns having, sometimes, peculiar characteristics—the major characteristics of the single-category paradigm—can lead only to empirical mischief and theoretical cost. If researchers wish to study generalization gradients, tied to either the category prototype or selected training instances, then considerable attention should be given at the outset to how the categories are to be defined to the subject and the suitability of the paradigm.

References


Received January 2, 2010
Revision received August 31, 2010
Accepted September 7, 2010

---

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://notify.apa.org/ and you will be notified by e-mail when issues of interest to you become available!