

This article was downloaded by:[Universitätsbibliothek]
On: 7 February 2008
Access Details: [subscription number 788880525]
Publisher: Psychology Press
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:
<http://www.informaworld.com/smpp/title-content=t716100704>

The modulating influence of category size on the classification of exception patterns

Donald Homa^a; Michael J. Proulx^b; Mark Blair^c

^a Arizona State University, Tempe, AZ, USA

^b Heinrich Heine University Duesseldorf, Duesseldorf, Germany

^c Simon Fraser University, BC, Vancouver, Canada

First Published on: 23 May 2007

To cite this Article: Homa, Donald, Proulx, Michael J. and Blair, Mark (2007) 'The modulating influence of category size on the classification of exception patterns', The Quarterly Journal of Experimental Psychology, 61:3, 425 - 443

To link to this article: DOI: 10.1080/17470210701238883

URL: <http://dx.doi.org/10.1080/17470210701238883>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

The modulating influence of category size on the classification of exception patterns

Donald Homa

Arizona State University, Tempe, AZ, USA

Michael J. Proulx

Heinrich Heine University Duesseldorf, Duesseldorf, Germany

Mark Blair

Simon Fraser University, Vancouver, BC, Canada

Generalization gradients to exception patterns and the category prototype were investigated in two experiments. In Experiment 1, participants first learned categories of large size that contained a single exception pattern, followed by a transfer test containing new instances that had a manipulated similarity relationship to the exception or a nonexception training pattern as well as distortions of the prototype. The results demonstrated transfer gradients tracked the prototype category rather than the feedback category of the exception category. In Experiment 2, transfer performance was investigated for categories varying in size (5, 10, 20), partially crossed with the number of exception patterns (1, 2, 4). Here, the generalization gradients tracked the feedback category of the training instance when category size was small but tracked the prototype category when category size was large. The benefits of increased category size still emerged, even with proportionality of exception patterns held constant. These, and other outcomes, were consistent with a mixed model of classification, in which exemplar influences were dominant with small-sized categories and/or high error rates, and prototype influences were dominant with larger sized categories.

A current debate, and the focus of the present study, is concerned with the memorial representation of concepts. This dispute centres around two dominant views of representation, exemplar based and abstraction based. Exemplar views (e.g., Lamberts, 1994; Medin & Schaffer, 1978; Nosofsky, 1988, 1991; Nosofsky & Johansen, 2000; Nosofsky & Zaki, 2002) hold that our

ideas and concepts are derived from the memorial encodings of particular instances. In abstraction-based theories (e.g., Homa, Dunbar, & Nohre, 1991; Knapp & Anderson, 1984; Minda & Smith, 2001; Smith & Minda, 1998), conceptual representations include summary or integrated interpretations of related experiences.

Correspondence should be addressed to Donald Homa, Department of Psychology, Arizona State University, Tempe, AZ 85287, USA. E-mail: donhoma@asu.edu

We would like to thank the Department of Chemistry at Indiana University for providing us with the source deck and write-up of STEPIT 7.4 (copyright 1975 by J. P. Chandler), and T. P. Roo for his comments on an earlier version of this manuscript.

This debate remains unresolved today. Attempts to separate the influences of specific exemplars from abstractions have been repeatedly parried, leading some researchers to conclude that the distinction may be hopeless (Barsalou, 1989). Examples of argument and counterargument include the forgetting rates of exemplars and prototypes (Medin & Schaffer, 1978; Posner & Keele, 1970), the shape of generalization gradients (Busemeyer, Dewey, & Medin, 1984; Homa, Sterling, & Trepel, 1981), and the importance of exemplar frequency on classification (Homa et al., 1991; Shin & Nosofsky, 1992). In the majority of cases, exemplar theorists have attempted to demonstrate that their model, with suitable parameter adjustments, could mimic supposedly inconsistent findings, although reinterpretation of results supporting exemplar-based models of classification has also occurred (Smith & Minda, 2000).

Nevertheless, the present study demonstrates that a general and powerful class of exemplar-based models of classification, loosely constrained by the form of its processing assumptions, can make predictions different from a prototype model, and that selected results clearly argue for a prototype interpretation and against an exemplar one. Our general thesis is that a role for each type of influence—exemplar and prototype—is probably necessary to explain category performance generally, but that exemplar and prototype influences operate in different segments of the hypothetical performance space shaped by learning variables (Homa, 1984).

The present study differs from previous research by focusing on the shape of generalization gradients to particular learning patterns and to the category prototype where the learning pattern either violated or not the prototype class for a pattern. A prototype category refers to the set of exemplars generated from a prototype, typically by a distortion algorithm (e.g., Blair & Homa, 2001; Posner, Goldsmith, & Welton, 1967). The distortion algorithm can generate an infinity of patterns, and rarely is a pattern more similar to a prototype other than its generating prototype, either objectively (Homa, 1978; Posner et al.,

1967) or psychologically (Homa, Rhoads, & Chambliss, 1979).

A pattern that violates its prototype class is called an exception pattern because it receives, from feedback in learning, a category label different from the remaining patterns of its prototype category; these patterns typically violate linear separability as well (e.g., Blair & Homa, 2001). The distinction between prototype and feedback categories is illustrated in Figure 1 for three prototype forms (arbitrarily called P1, P2, and P3). Members of each prototype category are enclosed within an ellipse and are each generated by a statistical algorithm applied to the same prototype. The members of each feedback category are contained within a common, shaded region. Thus, a feedback category refers to the collection of patterns, regardless of prototype, that receive a common feedback category name. In Figure 1, all instances of prototype category P2 are called Category C, except for P2-5, which is an exception pattern. In the present study, a feedback category always contained one or more exception patterns.

A question of critical importance in the present study is whether the participant will assign new patterns at manipulated levels of similarity to exception (e.g., P2-5) and nonexception (e.g., P2-2) learning patterns into the prototype category (e.g., Category C for pattern P2-2) or into the feedback category (Category B for pattern

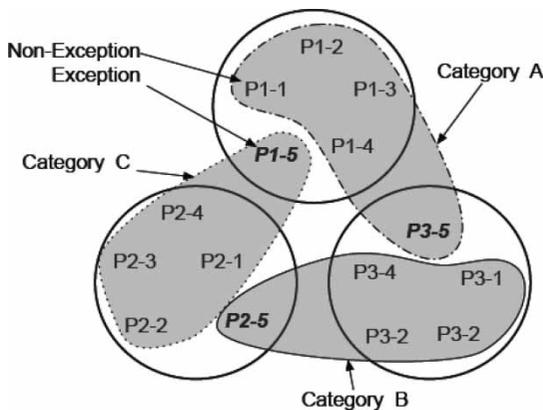


Figure 1. *Prototype and feedback categories.*

P2-5), and whether these tendencies are modulated in precise ways by category size and number of exception patterns in a category.

The reason for focusing on generalization gradients to exception patterns is because exemplar and prototype models make different predictions. Exemplar models predict a sharp or steep gradient across exemplar similarity, with this effect modulated slightly by increasing category size in learning. Exemplar models also predict that transfer patterns will track the category associated with the feedback of its most similar learning pattern, even when that pattern is an exception pattern. In contrast, a prototype model must predict a flat gradient across old–new exemplar similarity, with classification tracking the prototype category and not the feedback category.

The present study also investigated generalization gradients to the category prototype. Prototype models generally predict a sharp gradient across decreasing similarity from the category prototype, whereas exemplar models predict a flatter gradient (Smith, 2002; Smith & Minda, 2002). This prediction arises because prototype models predict that instances similar to the prototype must be classified more accurately than more distorted instances. In contrast, similarity to the category prototype is irrelevant to the predictions of exemplar models unless exemplar similarity is correlated with prototype similarity. In sum, exemplar models predict sharp gradients across old–new exemplar similarity and relatively shallow gradients across similarity to the category prototype whereas prototype models predict exactly the opposite.

A few studies have explored the shape of generalization gradients to the category prototype with formal models (e.g., Smith & Minda, 2002). However, no one has explored generalization gradients across old–new similarity when categories contained exceptions. Nosofsky and Zaki (2002) did investigate the transfer to an exception, training pattern and concluded that their exemplar-based model of classification could explain the resulting performance. However, this conclusion is tempered because only a single and high level of exemplar similarity was investigated, and category size was held constant and at a small size.

Experiment 1 explored the shape of gradients to exception and nonexception training patterns when categories were very large (45 different instances per category) and contained a single exception. Indeed, the point of Experiment 1 was to create a category representation that we believed could only be fitted by a substantial, prototype influence. Since previous studies have argued that abstraction of a prototype emerges only following the learning of categories of large size (Homa et al., 1991), we predicted that similarity to the category prototype would be significant, but that old–new exemplar similarity would be irrelevant, regardless of whether the similarity was to an exception or nonexception pattern. Formally, this outcome should be fitted only by a model with a substantial prototype contribution but not an exemplar model of classification.

In Experiment 2, categories of intermediate size were used, and the number of exception patterns per category was varied in a partial factorial design. Increasing the number of exception patterns per category also increases the proportion of exception patterns for that category when category size is held constant. By varying category size as well, an analysis of the shape of category gradients tied to exception and nonexception patterns under three contrasts was possible: When the number of exception patterns per category was held constant, and category size was increased; when the number of exception patterns per category was varied, and category size was held constant; and when the rate of exception patterns per category was held constant, and category size was varied. We anticipated that exemplar models would better predict transfer performance when the training categories were either small in size or contained multiple exception patterns.

Model predictions

In keeping with current exemplar models of classification (e.g., Nosofsky & Zaki, 2002), assume that the distance between any two patterns, i and j , has been computed in a multidimensional space, that is:

$$d_{ij} = \sum [(x_{ik} - x_{jk})^2]^{1/2} \quad 1$$

where, x_{ik} and x_{jk} are the values of patterns i and j on dimension k .¹ Typically, the similarity between patterns is an exponential function of their distance in multidimensional space:

$$s_{ij} = \exp(-cd_{ij}) \quad 2$$

The parameter c functions as a scaling (sensitivity) parameter (e.g., Nosofsky & Johansen, 2000).

The classification of a pattern into one of three categories, A, B, or C, is determined by computing the summed similarity of pattern i to all members of Category A, Category B, and Category C and then determining the ratio of evidence for each category—for example, for classification of pattern i into Category A, the formula is:

$$P(R_A|S_i) = \frac{\sum_{j \in C_A} s_{ij}}{\sum_{j \in C_A} s_{ik} + \sum_{j \in C_B} s_{ik} + \sum_{j \in C_C} s_{ik}} \quad 3$$

In a similar manner (e.g., Smith & Minda, 1998), the classification evidence due to a prototype influence is determined by the overall similarity of pattern i to each of the three prototypes, P_A , P_B , and P_C —that is, for assignment of pattern i into Category A:

$$P(R_A|S_i) = \frac{s_{iPA}}{s_{iPA} + s_{iPB} + s_{iPC}} \quad 4$$

We again assume that similarity of pattern i to Prototype A, B, or C is exponentially related to the multidimensional distance separating them, for example:

$$s_{iPA} = \exp(-gd_{iPA}) \quad 5$$

The only additional assumption is that the sensitivity parameter for the category prototype, g , is allowed to take on a value different from the sensitivity parameter, c , for exemplar similarity.

The mixed model is a weighted average of these two components. If the exemplar contribution (E) to classification is represented as in Equation 3, and the prototype contribution (P) is represented as in Equation 4, then the mixed model becomes:

$$P(R_A|S_i) = \beta P + (1 - \beta)E, \quad 0 \leq \beta \leq 1 \quad 6$$

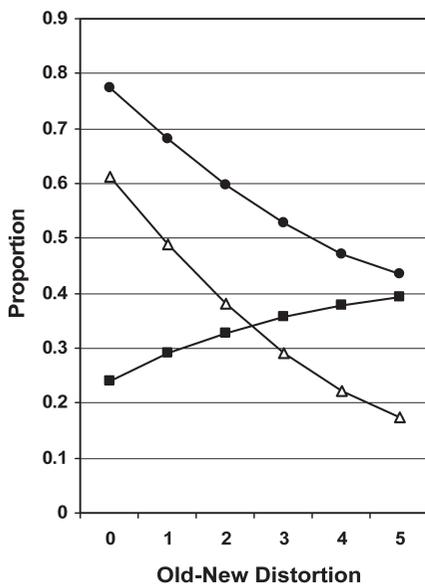
The mixed model, therefore, has three parameters: the sensitivity parameters c and g and the prototype/exemplar contribution, indexed by β . When $\beta = 0$, the mixed model reduces to the generalized context model (e.g., Nosofsky & Zaki, 2002), and classification is determined solely by distances among the instances; when $\beta = 1$, the mixed model reduces to a pure prototype model, and exemplar distances are irrelevant. Intermediate values of β reflect a mixed model in which both prototype and exemplar influences determine classification. A critical prediction is that β be near zero for categories of small size and systematically increase with increasing category size. The prototype contribution should also decrease with the number of exceptions per category.

Large values of β produce flat gradients across old/new similarity to training patterns and sharp gradients for similarity to the category prototype; small values of β generate sharp gradients across old/new similarity and flatter gradients across distortions from the prototype. These gradients, especially those across old/new similarity, are further modified by category size. An example of predicted rates of classification into prototype and feedback categories for patterns at manipulated levels of similarity to exception and nonexception patterns is shown in Figure 2 for categories of size 5 and 20.² The gradient NE-P

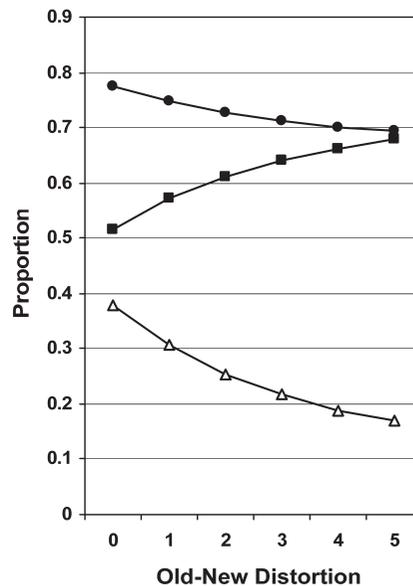
¹ The stimuli used in each experiment are distorted polygon-like patterns (Homa, 1978). Distances for these stimuli are better described by a Euclidean ($r = 2$) than city-block ($r = 1$) metric. This is consistent with previous modelling and multidimensional scaling of these stimuli (e.g., Homa et al., 1979). Also, participants invariably describe the distorted forms used here in terms of higher order perceptual properties rather than in terms of individual dimensions.

² Predicted gradients were generated using the distances obtained from a multidimensional scaling of selected patterns used in learning and transfer and by selecting values of c and g that closely matched those found in the quantitative fits reported in this manuscript ($c = g = 2.50$). Selecting different values of c and g does not substantially change the shapes of these functions and does not measurably alter the predictions contained in the Introduction.

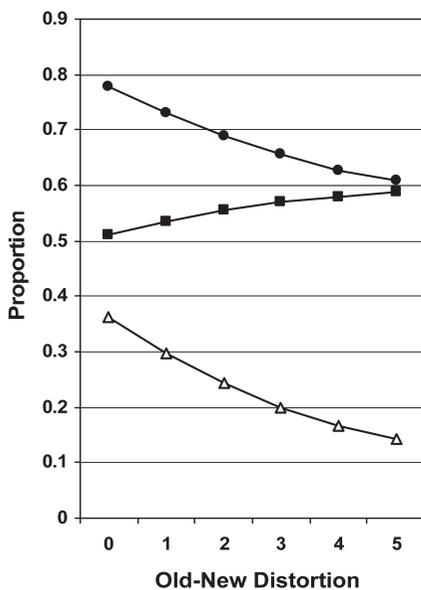
Category 5, $\beta = .00$



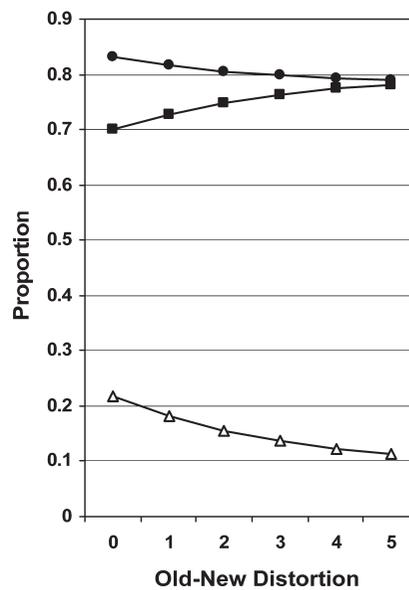
Category 20, $\beta = .00$



Category 5, $\beta = .50$



Category 20, $\beta = .50$



—●— NE-P —■— E-P —△— E-FB

—●— NE-P —■— E-P —△— E-FB

Figure 2. Predicted classification gradients across old–new similarity into prototype (P) and feedback (FB) categories as a function of β and category size. New patterns are related to either an exception (E) or a nonexception (NE) training pattern.

refers to classification rates into the prototype category (P) for new instances related to a nonexception training pattern (NE). The E-FB gradient refers to classification rates into the feedback category (FB) for new instances related to an exception training pattern. The third gradient is E-P, which indexes the rate of classification into the prototype category for new instances related to an exception pattern.

When category size is small, and a prototype influence is absent (top left panel), similar gradients for NE-P and E-FB are predicted, and classification tracks the exception pattern only slightly below the gradient for nonexception patterns. When category size is large, and the prototype component is substantial (bottom right panel), the NE-P and E-FB gradients separate substantially, and each gradient is nearly flat across old–new similarity. Here, the impact of an exception pattern is nearly absent, and the participant sorts into the prototype category—the gradient E-P nearly matches the NE-P gradient, and classification into the prototype category is largely unaffected by old–new similarity and whether the most similar training pattern was associated with the exception or nonexception pattern.

The similarity between the panels in Figure 2 for the Category 5, $\beta = .50$, and the Category 20, $\beta = .00$, motivated a fourth model prediction, addressed and evaluated in the Discussion—that participants who train on large category sizes might base their category assignments on a substantial prototype contribution and a restricted set of training patterns retrieved at the time of transfer. The retrieval of a subset, rather than the full set, of stored exemplars in a category judgment task was recently demonstrated by Storms, De Boeck, and Ruts (2000). Separating these two extremes—whether the participant consults

the full set of stored patterns with no prototype contribution versus the retrieval of a restricted set of patterns combined with a strong prototype contribution—is based on the steepness of the gradients across old–new distortion for exception and nonexception patterns as well as the shape of the gradient to the category prototype.

The transfer test also included new patterns unrelated to any particular training pattern that were also high-level distortions from the category prototype. The inclusion of unrelated new patterns provides a converging check for prototype influences—if exemplar influences are zero, these patterns should be classified into the prototype category at rates equal to those for the new patterns (also high distortions) that were at a manipulated distance from a particular training pattern.

EXPERIMENT 1

The stimuli were abstract forms generated from different prototypes. The distortion of each pattern from its prototype (and from other forms) can be precisely measured (e.g., Homa, 1978). Participants first learned three categories containing 45 instances each, 1 of which was an exception pattern, followed by a transfer test that contained old, related new, unrelated new, and the category prototype for each category. The related new instances had a manipulated similarity relationship (1.0–5.0 unit distortion) to an exception or nonexception training pattern while maintaining a low-similarity relationship (5.0 unit distortion) to the prototype.³

Method

Participants

A total of 15 Arizona State University undergraduates served as participants.

³ Because we were ultimately interested in deriving similarity values for quantitative fits, a second group of 24 participants provided pairwise similarity ratings to selected patterns from the same prototypes as those used in Experiment 1. These ratings were then subjected to a nonmetric multidimensional scaling analysis (Kruskal, 1964; Shepard, 1962), whose distances were then used to fit the quantitative predictions. The stimulus pool consisted of 30 different patterns, 10 each from each of three prototypes. The full matrix of 435 interpattern ratings was then multidimensional scaled (Kruskal, 1964; Shepard, 1962) in Dimensions 1–6, using a program called KYST. The stress function declined from about .35 in 1 dimension to less than .10 in 6 dimensions (stress formula 1). The 3-dimensional solution was deemed reasonable (stress = .148) and was selected for all subsequent analyses.

Materials and apparatus

Members of three form categories were generated in the manner described previously (e.g., 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 arbitrarily designated as the category prototype; different members of this category are then generated by statistically moving each of the dots of the prototype. These particular prototypes were selected because prior multidimensional scaling revealed that, within the context of 20 prototypes, these three prototypes were roughly equidistant from each other and moderately distant from the centroid of the space.

The amount of dot displacement determines the distortion level of a pattern. A mixture of low-, medium-, and high-level distortions from the prototype was used in the learning phase. For low-level, medium-level, and high-level distortions, each dot is displaced, on the average, about 1.20, 2.80, and 4.60 units, respectively, from each corresponding dot of the prototype (Posner et al., 1967); patterns belonging to different prototype categories have their dots displaced, on the average, by 10–15 units (Homa, 1978). The degree of distortion between learning and selected transfer patterns was systematically varied by displacing each dot of a new pattern exactly 1.0, 2.0, 3.0, 4.0, or 5.0 units of a high-level distortion used in the learning phase. These transfer patterns were themselves high-level distortions of the category prototype and were checked to verify that these patterns were neither closer nor further from the category prototype following exemplar distortion. These patterns are termed related new to distinguish them from other new instances, also high-level distortions from the category prototype, but not specifically related to any particular training instance (unrelated new). These latter patterns varied between 5.0–8.0 units from any particular training instance.

The patterns were drawn by a Cal-Comp plotter and were converted to slides. During the learning and transfer phase, the slides were

presented by a Kodak carousel projector (Model 851) onto a flat white wall. Each participant viewed the stimuli from a seated position at a distance of about 5 feet.

Learning phase

On a given study/test block, the stimuli were shown in random order, one at a time, for 5 s, accompanied with corrective feedback. These same stimuli were then shown, without feedback, in a different random order, and the participant classified each instance by writing his response onto a prepared test sheet.

Participants performed three learning sessions of approximately 1 hr in length over the course of 3 days. Each session consisted of 45 different stimuli arranged in three random orders, presented in five alternating study/test blocks. Of these 45 patterns per session, 15 were from each of the three categories, with 5 low-level, 5 medium-level, and 5 high-level distortions. The location of the exception patterns was counterbalanced, such that it occurred in Session 1 for one category, Session 2 for the second category, and Session 3 for the third category.

Transfer

Participants received the transfer task on the day following completion of the learning phase. A total of 98 stimuli was presented, 26 from each of the three learned categories and 20 foils. The 26 stimuli from each category consisted of the category prototype, 5 old patterns (1 low, 1 medium, and 3 high distortions, including the exception pattern), 15 new patterns that were high-level distortions from the prototype and which had a manipulated similarity relationship to a high-level distortion training pattern, and 5 new patterns that were high-level distortions from the prototype and not specifically related to any particular old pattern. Of the 15 new patterns related to a particular old pattern, 5 were related to an exception pattern (a pattern from a different prototype), 1 at each of 5 distances from the pattern, and 10 to a nonexception pattern (2 patterns at each of 5 distances from the training pattern). The range of distortion of a pattern

from a particular old varied from 1.0 units/vertex to 5.0 units/vertex, which roughly spans the range from a low-level (1.20 units/vertex) to a high-level (4.60 units/vertex) distortion. These distortions from a training pattern were checked (and modified, if necessary) so that they maintained a high-level distortion relationship to the category prototype. The end result was that the related new patterns had a similarity relationship to a particular old instance, in which the old instance was either an exception or a nonexception training pattern.

No feedback was provided on the transfer task. Participants were told that some patterns belonged to none of the three learned categories and that they should assign those patterns to a miscellaneous “none” category.⁴

Results

Learning

All participants showed substantial learning, with mean error rates decreasing progressively from the initial (.278) to the terminal (.030) learning trial.

Transfer

Figure 3 shows the assignments into the prototype category for instances related to nonexception (left panel) and exception (right panel) patterns as a function of old–new distortion (0.0–5.0). Also shown are the model fits for exemplar, prototype, and mixed models (discussed shortly). An analysis of classification performance revealed that the main effect of item type (exception vs. nonexception) was significant, $F(1, 14) = 9.33$, $\eta^2 = .400$, $p < .01$, with classification of nonexception items exceeding that of exception items (.929 vs. .818). An analysis of old–new distortion (0.00–5.00) restricted to the nonexception patterns revealed that the decline in assignments with increasing distortion was statistically significant,

$F(5, 70) = 3.43$, $\eta^2 = .196$, $p < .05$, due entirely to the most extreme level (5.0) of old–new distortion; performance across the distortion range from 0.00–4.00 produced no significant decline. A similar analysis of old–new distortion for the exception patterns was not significant, $F(5, 70) = 0.71$, $p > .20$.

These outcomes were best predicted by the mixed and prototype models, with the exemplar model faring the worst for these contrasts. The major deficiency of the exemplar model was that it predicted a strong gradient across old–new similarity, especially for the exception items. This model also underpredicted performance for exception items.

Assignments into the feedback category (the category containing the exception) for instances related to the exception patterns is shown in Figure 4 as a function of old–new distortion. Again, each of the model predictions is shown as well. In general, these assignments decreased with increasing old–new distortion, $F(5, 70) = 2.09$, $\eta^2 = .149$, $p = .07$, an outcome best captured by the mixed model. The prototype model predicted a stable and low rate of assignments across old–new distortion, generally underpredicting performance. The exemplar model poorly captured this outcome, substantially overestimating these assignments, especially when old–new distortion was small.

An analysis of assignment of new patterns into the prototype category revealed that related nonexception patterns were classified better ($NE = .929$) than related exception ($E = .818$) or unrelated patterns ($U = .800$), $F(2, 28) = 4.70$, $\eta^2 = .251$, $p < .02$. In general, each model did an adequate job, although the prototype model must predict equivalent performance on each of the new pattern types, since old–new exemplar distortion should be irrelevant for this model (all patterns were equidistant from the prototype).

⁴The quantitative model fitted to the data included classification into the none category, mirroring the form of Busemeyer et al. (1984). Therefore, predicted classification into the prototype, feedback, other, and none category summed to 1.00. However, the use of a “none” category in Experiment 1 proved to be uninformative, with less than 5% of all category patterns classified into this category. The “none” option was dropped in Experiment 2.

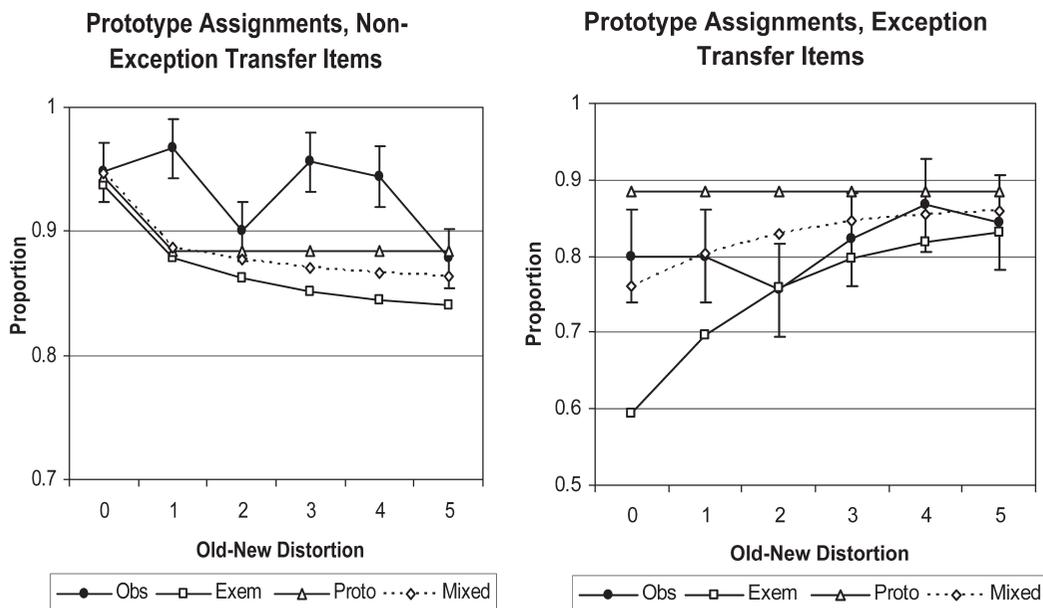


Figure 3. Mean proportion classification into the prototype category for new patterns at various distortions from a nonexception training pattern (left panel) or exception training pattern (right panel), Experiment 1. Models predictions are also shown (error bars refer to standard error of the mean).

A final analysis assessed the classification of training patterns that were low-, medium-, and high-level distortions of the prototype. This analysis revealed no significance difference among these types, $F(2, 28) = 2.36$, $\eta^2 = .144$, $p = .11$, but ceiling effects (all prototype distortions were old patterns and were classified with a 95% accuracy) probably minimized the likelihood of finding significance for these patterns.

Quantitative fits to a mixed model

The important item similarities obtained from the multidimensional scaling task used to fit quantitative models are summarized in Table 1.

All model fits were applied to the full response matrix, including assignments of related and unrelated new patterns into the prototype, feedback, other category (the category containing neither the prototype pattern or the exception), and "none" category. Analyses were done for three model variations: the full mixed model, which accorded a weight to prototype (β) and exemplar

Feedback Assignments, Exception Transfer Items

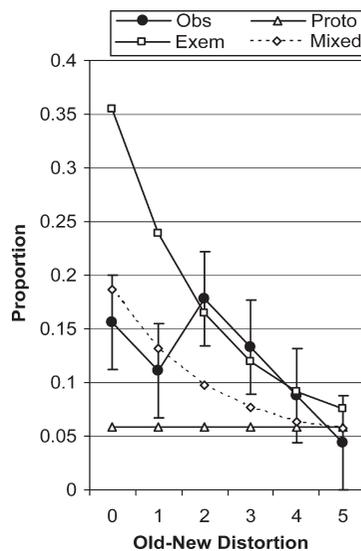


Figure 4. Mean proportion of classification into the feedback category for new patterns at various distortions from an exception training pattern, with accompanying model predictions, Experiment 1 (error bars refer to standard error of the mean).

Table 1. *Derived Euclidean distances for selected pattern contrasts*

| <i>Pattern</i> | <i>Low</i> | <i>Med</i> | <i>High</i> |
|----------------|------------|------------|-------------------|
| Prototype | 0.24 | 0.45 | 0.82 |
| Low | 0.26 | 0.50 | 0.76 |
| Medium | | 0.72 | 0.86 |
| High | | | 1.07 |
| Random | 1.60 | 1.60 | 1.60 |
| New-1.0 | | | 0.18 ^a |
| New-2.0 | | | 0.35 ^a |
| New-3.0 | | | 0.52 ^a |
| New-4.0 | | | 0.70 ^a |
| New-5.0 | | | 0.87 ^a |

Note: All distances were derived from the three-dimensional solution space. Random patterns refer to patterns from alternative (between) category distances.

^aDistances were estimated.

($1 - \beta$) influences, a pure exemplar model ($\beta = .00$), and a pure prototype model ($\beta = 1.00$).

Quantitative fits revealed that a prototype contribution to classification was substantial ($\beta = .517$). The mean absolute difference between observed and predicted values for the entire response matrix⁵ was .033, .041, and .045 for the mixed, exemplar, and prototype models, respectively. An analysis revealed that the model fits varied significantly, $F(2, 130) = 4.08$, $p < .02$, with the mixed model differing from both the exemplar and prototype models, which did not differ from each other ($p > .05$).

Discussion

Experiment 1 demonstrated that assignment into the prototype category was significantly higher for patterns related to the nonexception patterns than for those related to exception patterns, but that old-new distortion from a training pattern was not a potent variable. For new patterns related to the nonexception items, performance was stable across the entire range of old-new distortion except at the highest level. Similarly, old-new similarity for exception items was not a significant predictor of performance.

Although an exemplar model can predict higher performance on nonexception than on exception patterns, this model predicted neither the level of performance, nor the shape of gradients, of patterns related to exception patterns. Specifically, the exemplar model underpredicted assignments into the prototype category of new patterns related to exception and nonexception training patterns, while overpredicting the steepness of these gradients. In addition, the exemplar model overpredicted both the rate of assignment and the steepness of the gradient of related patterns into the feedback category. In fact, old-new exemplar similarity was not a strong influence on classification in this experiment, an outcome predicted by prototype models of classification. However, a pure prototype model erroneously predicted no difference in classification rates of new patterns that were either unrelated or related to particular training patterns. The only model that captured most of these outcomes was a mixed model with a strong, but not sole, prototype influence. The major deficiency of the mixed model was that it underpredicted classification of new patterns related to the nonexception patterns (as did all models).

In effect, Experiment 1 successfully demonstrated that one can create categories that have a substantial prototype influence and a markedly reduced exemplar influence. Training on categories of very large size produces concepts that are reasonably immune to the influence of an exception pattern.

EXPERIMENT 2

Unfortunately, the shape of the gradient to the category prototype could not be assessed due to ceiling effects, presumably because these patterns were old patterns at the time of test. Experiment 2 addressed this concern by using prototype distortions that were new at the time of transfer, while further investigating the shape of exemplar

⁵ Parameter values were obtained by a program called STEPIT, which finds local minima of a smooth function in several parameters.

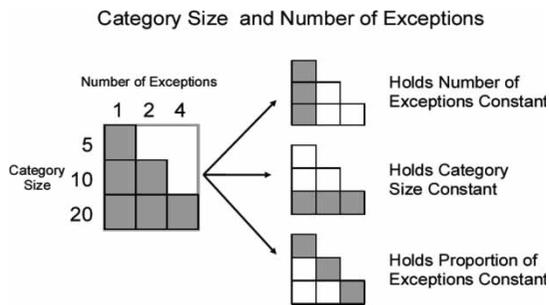


Figure 5. Design of Experiment 2, showing the conditions involving category size (5, 10, 20) and number of exception patterns per category (1, 2, 4).

gradients when categories were of smaller size and which contained one or more exceptions. A mixed model should predict a strong exemplar influence when categories are of small size, or when multiple exceptions per category exist, and an increasing prototype influence as category size increases.

The basic design is shown in Figure 5. A vertical slice through the design box holds constant the number of exception patterns per category while varying category size. Obviously the proportion of exception patterns per category changes as well (for categories of size 5, 10, and 20, these rates were 20%, 10%, and 5%, respectively). A horizontal slice holds category size constant, while varying the number of exception patterns per category. Again, proportionality of exceptions per category varies as well; for category sizes of 10, these rates were 10% and 20%; for category sizes of 20, the corresponding rates were 5%, 10%, and 20%, respectively. Finally, the diagonal slice varies category size and number of exceptions per category, while holding proportionality of exceptions per category constant (20%). This contrast permits an assessment of whether category size alone can override a high but constant rate of exceptions.

Method

Participants

A total of 90 Arizona State University undergraduates were assigned equally to the six

combinations of category size and number of exceptions. None of the participants had previously served in any experiment using these forms.

Materials and apparatus

Participants were run individually in a sound-attenuated room, with all patterns shown on a standard computer screen. The patterns were again distorted shapes from prototype forms and appeared as white outlines against a black background (e.g., Blair & Homa, 2001). All responses were made by keyboard selections.

Procedure

Each participant was told, in part, that a pattern would be presented that belonged to Category A, B, or C, and that the category assignment should be made by striking the A, B, or C key. Following the category judgement, feedback appeared for 1 s as an upper-case letter (e.g., B) located to the right and slightly above the pattern. The pattern and feedback appeared simultaneously for 1 s, the screen went blank for 1 s, and the next pattern was shown. Each block contained 15, 30, or 60 different patterns (for participants in the category size 5, 10, or 20 condition). Learning terminated after either 15 blocks or a criterion of 10% error rate was reached. There was no visible break between blocks, with the order of patterns randomized on each trial block.

Each of the six conditions determined by Category Size \times Number of Exceptions used three different random assignments of patterns, with different patterns functioning as the exception pattern(s) in each random order. This ensured that results were not due to the peculiar properties associated with a given pattern. Furthermore, the same patterns selected as exceptions were contained within each of the category size conditions. All exceptions were high-level distortions.

All categories were composed of a similar mixture of low-, medium-, and high-level prototype distortions—for example, for categories of size 5, 1 pattern was a low-, 1 was a medium-, and 3 were high-level distortions. For categories

of size 10, there were 2 low-, 2 medium-, and 6 high-level distortions; categories of size 20 contained 4 low-, 4 medium-, and 12 high-level distortions.

The transfer test, given immediately after the completion of learning, contained four types of item: the category prototype, unrelated new patterns, new patterns related to a nonexception stimulus, and new patterns related to a category exception. The related new items were at one of three levels of distortion (1.0, 3.0, 5.0) from a training pattern, roughly corresponding to a low-, medium-, or high-level distortion of that pattern.

The transfer test was identical for all conditions with one change. For those conditions having a single exception pattern per category, the transfer set contained 75 patterns, 25 in each of the three learned categories. The 25 patterns per category were composed of the category prototype, 15 unrelated new (5 patterns at each of three levels of distortion), 6 related new patterns associated with nonexception patterns, and 3 related new patterns associated with the exception pattern.

For the conditions involving two or more exceptions per category, there were 84 transfer patterns, 28 in each of three categories. These additional 9 patterns were new items related to a second, exception pattern of the learning category and were included to broaden sampling and increase statistical power. These 9 patterns were equally sampled from the three categories and existed at each of three levels of distortion. The remaining patterns were otherwise identical to the conditions involving a single exception pattern per category.

The transfer set was presented in a random order to the participant, and feedback was omitted.

Results

Results from the three conditions involving a single exception pattern per category are presented first, followed by the three conditions involving two or four exception patterns per category. Comparison among selected contrasts involving a constant exception rate but different category sizes is presented last.

One exception pattern per category: Learning

Mean proportion error rates for the first and last trial and the number of trials needed to reach criterion are presented in Table 2 for each category size condition. In general, participants had an initial error rate near chance on Trial 1 (between 58%–66%) and a terminal error rate between 15–22%.

Transfer

Figure 6 shows the classification assignments of new instances of different distortion (1.0, 3.0, 5.0) to exception (E) and nonexception (NE) training patterns into the prototype (P) or feedback (FB) categories. The left panel shows this performance for category size 5, the middle panel for category size 10, and the right panel for category size 20 (the dotted lines are model predictions, discussed shortly).

Prototype-based classification: Old–new similarity transfer

An initial analysis investigated classification into the prototype category of new patterns related to exception and nonexception patterns as a function of old–new distortion (1.0, 3.0, 5.0) and category size (5, 10, 20). The main effects of category size, $F(2, 42) = 4.04$, $\eta^2 = .161$, $p < .05$, and type of training pattern (exception, nonexception), $F(1, 42) = 25.14$, $\eta^2 = .374$, $p < .01$, were significant; the main effect of old–new distortion was not, $F(2, 84) = 2.07$, $\eta^2 = .047$, $p > .10$. However, the Old–New Distortion \times Training Pattern Type interaction was significant, $F(2, 84) = 6.58$, $\eta^2 = .135$, $p < .01$. Performance improved

Table 2. Learning performance for each of the conditions

| Condition | Error rate | | Trials to criterion |
|-----------|------------|----------|---------------------|
| | Initial | Terminal | |
| 5(1) | .662 | .222 | 14.53 |
| 10(1) | .576 | .147 | 12.60 |
| 20(1) | .592 | .213 | 12.60 |
| 10(2) | .618 | .362 | 14.60 |
| 20(2) | .591 | .233 | 14.13 |
| 20(4) | .601 | .317 | 14.80 |

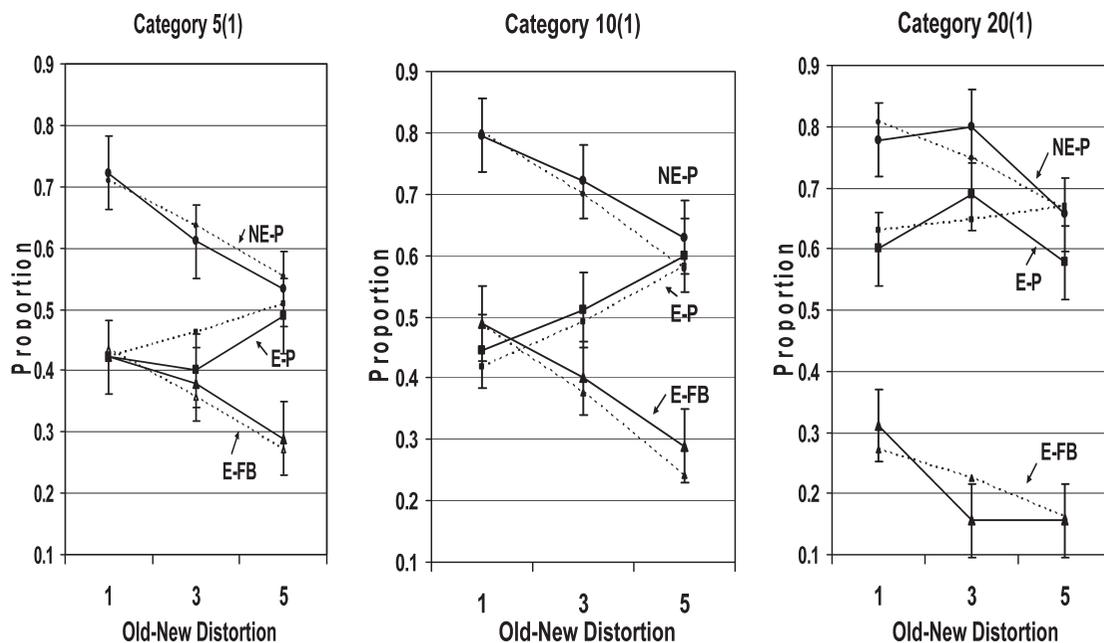


Figure 6. Mean classification assignments of new instances of different distortion (1.0, 3.0, 5.0) to exception (E) and nonexception (NE) training patterns into the prototype (P) or feedback (FB) categories for categories containing a single exception. The left panel shows category size 5, the middle panel category size 10, and the right panel category size 20 (the dotted lines are model predictions), Experiment 2 (error bars refer to standard error of the mean).

with increases in category size (Category 5 = .530, Category 10 = .606, Category 20 = .683), with accuracy greater for patterns related to nonexception (.684) than for those related to exception (.526) patterns. The Old–New \times Type interaction reflected the larger disparity between non-exception and exception patterns when category size was 5 patterns (.622 vs. .437, respectively) than when category size increased (for Category 10, NE = .693, E = .519; for Category 20, NE = .714, E = .622). A noteworthy result, discussed shortly, is that exceptions in categories of size 5 and 10 were classified into the prototype and feedback categories at the same rate when old–new distortion was lowest. However, these probabilities were markedly different when category size was 20.

Prototype similarity

Classification of the prototype distortions (low, medium, high) into the prototype category

increased with increases in category size, $F(2, 42) = 6.76$, $\eta^2 = .244$, $p < .01$, and decreased with increasing pattern distortion, $F(2, 84) = 32.15$, $\eta^2 = .434$, $p < .01$. The interaction between these variables was also significant, $F(4, 84) = 2.90$, $\eta^2 = .121$, $p < .05$, and reflected the greater disparity among the low-, medium-, and high-level distortions for the categories defined by larger numbers of patterns. For each category size, the category prototype was classified best of all patterns (.844), with performance diminishing lawfully with increases in distortion level (low = .806, medium = .738, high = .643).

2 or 4 Erroneous patterns per category: Learning

Learning performance is summarized in the bottom half of Table 2. These conditions were more difficult than those associated with a single exception pattern per category, although the outcomes were lawfully related to presumed task difficulty: The Category 20 condition with 2

exceptions per category, 20(2), had a terminal error rate of .233, versus the 20(4) condition rate of .317, followed by the 10(2) condition of .362.

Transfer

Figure 7 shows the classification assignments of new instances of different distortion (1.0, 3.0, 5.0) to exception (E) and nonexception (NE) training patterns into the prototype (P) or feedback (FB) categories, displayed separately for the three category size conditions.

Prototype-based classification: Old–new similarity

An analysis restricted to classifications into the prototype category as a function pattern type (NE, E) revealed higher rates of assignment into the prototype categories for nonexception patterns than exception patterns, $F(1, 42) = 19.67$, $\eta^2 = .319$, $p < .01$. Simple tests showed that the Category 20 condition with 2 exceptions

per category was superior to the Category 10 condition with 2 exceptions per category, $F(1, 28) = 14.68$, $\eta^2 = .344$, $p = .001$, as well as to the Category 20 condition with 4 exceptions per category, $F(1, 28) = 13.62$, $\eta^2 = .327$, $p < .01$, respectively. The main effect of old–new distortion was not significant, $F(2, 84) = 2.67$, $\eta^2 = .060$, $p = .075$, but it did interact with type of item (NE, E), $F(2, 84) = 21.45$, $\eta^2 = .338$, $p < .001$. The latter interaction reflected the result found earlier, with classification decreasing across old–new distortion for NE–P items and increasing across old–new distortion for E–P items. Again, classification of exception patterns at the lowest level of old–new distortion was classified into the prototype and feedback categories equally often when categories contained 10 patterns (with 2 exceptions) but not when the categories contained 20 patterns (with 2 or 4 exceptions).

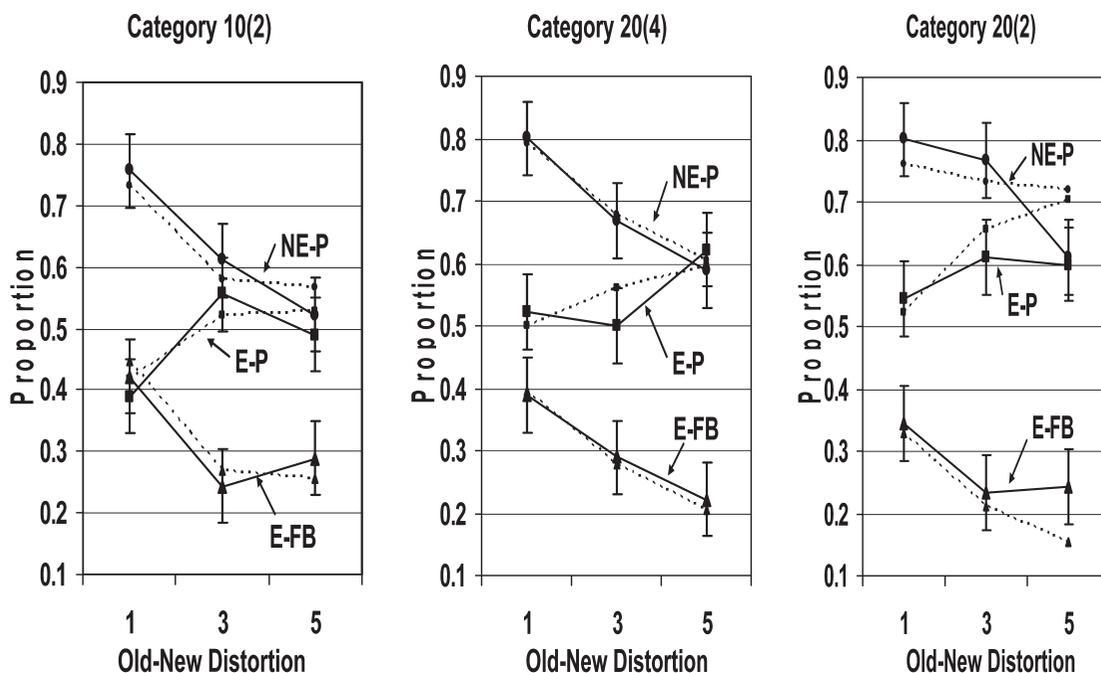


Figure 7. Mean classification assignments of new instances of different distortion (1.0, 3.0, 5.0) to exception (E) and nonexception (NE) training patterns into the prototype (P) or feedback (FB) categories, displayed separately for the three category size conditions, Experiment 2 (error bars refer to standard error of the mean).

Prototype similarity

Classification of unrelated new (low, medium, high) patterns decreased with distortion level (low = .837; medium = .753; high = .641), $F(2, 84) = 40.66$, $\eta^2 = .492$, $p < .001$. The three conditions differed amongst themselves, $F(2, 42) = 3.63$, $\eta^2 = .147$, $p < .05$, with performance ordered with 20(2) highest (.825), 10(2) lowest (.658), and 20(4) intermediate (.748), an ordering that matched that for the related new patterns. A similar ordering was found for the category prototype, with 20(2) = .956, 20(4) = .822, and 10(2) = .708, $F(2, 42) = 2.82$, $\eta^2 = .121$, $p = .067$.

Contrast among 1 vs. 2/4 errors/category conditions

The mean overall correct transfer performance for the six major conditions was ordered as expected, with best performance associated with larger category sizes and smaller manipulated levels of exception rates: 20(1) = .780; 20(2) = .757; 10(1) = .705; 20(4) = .694; 10(2) = .617; 5(1) = .612. A global analysis revealed that the six conditions differed amongst themselves, $F(5, 84) = 4.20$, $\eta^2 = .200$, $p < .01$. In general, differences of 7–8% reached significance at the $p = .05$ level, and, therefore, the six conditions separated themselves into three major groups: 20(1) = 20(2) > 10(1) = 20(4) > 10(2) = 5(1).

A final contrast of the six conditions compared the gradient changes across old–new distortion for new items related to exception and nonexception training patterns. One measure of gradient change is to compute the difference in classification accuracy for new items that were maximally similar (1.0) or maximally different (5.0) from a training pattern. The motivation for this contrast is straightforward—exemplar-based models of classification predict a substantial advantage for new items having a high-level versus low-level of similarity to exception and nonexception training patterns. Contrarily, classifications based on prototype influences predict reduced or minimal gradient changes. Figure 8 shows these contrasts for exception and nonexception patterns, separately for condition.

Two results are apparent: (a) Gradient changes are larger for nonexception than exception patterns; and (b) magnitude of gradient changes is

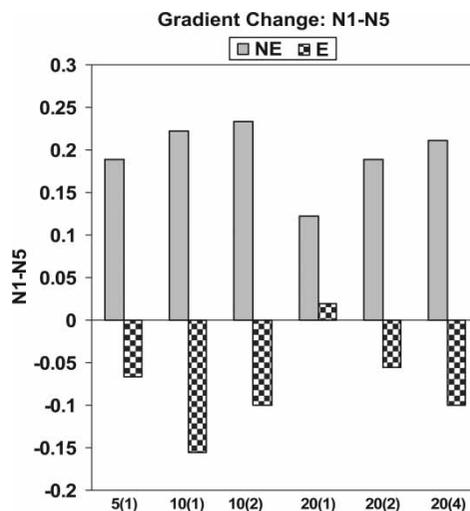


Figure 8. Summary of gradient differences for each of the six conditions, shown separately for nonexception (NE-P) and exception (E-P) patterns, Experiment 2.

generally reduced with increasing category size and decreasing rates of exception patterns.

Model fits

The fit of the general mixed model of classification to the full set of transfer patterns is summarized in the top half of Table 3. Generally, the overall model fit was satisfactory, with the mean absolute difference between observed and predicted values ranging from .01–.03 for each condition. Of most importance are the estimates for the category prototypes. For conditions associated with a single erroneous pattern per category, estimates of β generally increased with category size. For conditions associated with 2 or 4 errors/category, estimates of β decreased overall, but were still modulated by both category size and proportionality of feedback.

Although the exemplar model performed satisfactorily when category size was small, and the rate of exception patterns was high (as indicated by the small values of β), the deficiency of this model became apparent when β was high. For example, when category size was 20 and contained a single exception pattern, the prototype contribution was highest of the six conditions ($\beta = .479$). Figure 9 shows assignments into the prototype

Table 3. Parameters for the complete and partial mixed model, Experiment 2

| Model | Parameters | Learning conditions | | | | | |
|----------|------------|---------------------|-------|-------|-------|-------|-------|
| | | 5(1) | 10(1) | 10(2) | 20(1) | 20(2) | 20(4) |
| Complete | β | .000 | .000 | .000 | .479 | .414 | .185 |
| | c | 2.65 | 2.42 | 2.06 | 3.18 | 3.48 | 3.10 |
| | g | 0.63 | 1.71 | 1.45 | 1.66 | 1.46 | 0.67 |
| | RMSE | .034 | .082 | .045 | .064 | .059 | .047 |
| Partial | β | .000 | .371 | .334 | .635 | .707 | .598 |
| | c | 2.65 | 2.77 | 2.38 | 3.05 | 4.04 | 3.65 |
| | g | 0.63 | 1.51 | 1.36 | 1.93 | 1.99 | 1.91 |
| | RMSE | .034 | .046 | .035 | .062 | .041 | .027 |

Note: Complete model assumes number of patterns retrieved equals training category size; partial model assumes that a subset of training patterns is retrieved. RMSE = root mean square error.

category for the prototype and prototype distortions (low, medium, high). Also shown are the predicted rates for the mixed and exemplar model (including a modified mixed model). Consistent with the conclusions of Smith and Minda (2002), the exemplar model predicts a shallower gradient than obtained, an outcome better realized by the mixed model.

The mixed model was also fitted with an adjusted (or functional) category size, shown on the bottom half of Table 3. The adjusted size refers to the number of training patterns hypothetically retrieved at the time of classification necessary to produce an optimal fit to the same data matrix. To determine an optimal fit, category size was allowed to vary across a wide range (but not exceed its objective training size), and the functional category size was defined as that size that produced a smallest root mean square error (RMSE). One surprising finding was that an optimal fit to the mixed model occurred when only 5–7 patterns were retrieved in each of the six conditions, even when the training category size was 20. With the exception of the category size = 5 condition, the improved fits were statistically significant when compared to the full mixed model as well (all p s < .05). Notably, the magnitude of β generally increased with category size

and decreased with the number of exceptions per category: for 5(1), 10(2), 10(1), 20(4), 20(2), and 20(1), $\beta = .00, .33, .37, .60, .71,$ and $.64,$ respectively.⁶

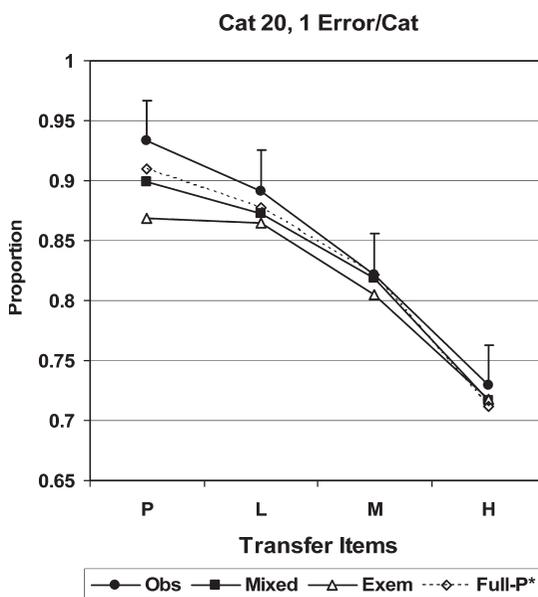


Figure 9. Mean rate of assignment into the prototype category for the prototype and prototype distortions (low, medium, high) for the 20(1) condition, Experiment 2. Also shown are predicted rates for various models (error bars refer to standard error of the mean).

⁶ This analysis was also performed for the results from Experiment 1. Although category size was 45, an optimal fit was obtained with a function category size of 7. The prototype contribution, $\beta = .78,$ was higher than any obtained in Experiment 2.

Discussion

The literature on categorization is replete with the benefits of increasing category size in learning on later transfer (e.g., Homa, 1984). This experiment extends this conclusion for categories of size 5 to 45 when exceptions occur in each category, not just a single exception pattern within a category but as many as four patterns. When 10% of the patterns of a category are exception patterns, the mean overall transfer accuracy was .705 versus .757 for categories of size 10 versus 20, respectively. With an exception rate of 20%, the overall rates of classification were reduced, compared to the 10% error rate, but the benefits of category size still emerged—the corresponding rates, for categories of size 10 versus 20, were .617 and .694.

The mixed model of categorization fitted to the results showed a prototype influence that was minimal for small category sizes and/or high rates of exception patterns and an increasingly large prototype influence with larger category sizes, especially when accompanied with a low rate of exception patterns. Obviously, the mixed model with its additional parameter cannot perform worse than the exemplar or prototype model. However, only a mixed model can simultaneously accommodate all the results found here—the shapes of different gradients tied to exception and nonexception training patterns, the shape of the gradient for patterns related to the prototype, overall differences between patterns related to exception and nonexception patterns, and performance differences for different types of

new patterns, all as a function of category size. Furthermore, the mixed model reveals regularities of exemplar and prototype influences that vary in the manner predicted.⁷

The combination of variables explored in the present study may clarify an apparent inconsistency with the results of Nosofsky and Zaki (2002). Nosofsky and Zaki had participants learn two categories of seven exemplars each, where one pattern in each category was an exception pattern. At the time of transfer, exception neighbours—new patterns that differed from the training pattern by a single value on one of six binary dimensions—were consistently assigned to the category having the exception pattern, leading these authors to conclude: “the exemplar model predicted that items that were similar to the exceptions along attended dimensions would be classified into the exceptions’ category, whereas the mixed-prototype model predicted that such items would be classified into the opposite category” (p. 936).

However, these authors held category size constant (and at a small size), and only a single and high level of old–new similarity was analysed. In the present study, transfer gradients similarly tracked the exception learning pattern when category size was small, and old–new similarity was highest, an outcome that mirrors that of Nosofsky and Zaki (2002). However, when category size was substantially increased, these tendencies reversed. The prototype influence was also largest when category size was maximal.⁸

⁷ Our modelling did not include either differential weighting of the dimensions (obtained from the multidimensional scaling task), as is sometimes done with exemplar modelling, or implementation of the λ response-scaling parameter recently added to the generalized context model. The use of λ , introduced to explain deterministic responding by individual participants, does not always appear in exemplar models (e.g., Zaki & Nosofsky, 2004). Nonetheless, our conclusions are restricted to the models explored and not to other variations that could be considered.

⁸ Zaki and Nosofsky (2004) recently argued that gradient effects obtained with dot pattern stimuli (related to our 9-point forms) may be artificially produced by procedures used to generate distortions. This assertion is false, since their demonstration applies to a 1-category paradigm and cannot be extended to a paradigm in which multiple categories are learned. When participants learn multiple categories, rather than inspecting a single category (as done in their paradigm), correct assignments cannot be based on potential elongation (or other) properties that might produce the gradient (according to their argument) since, presumably, all categories would have these properties. If accuracy of classification cannot be based on this property, then obviously chance performance cannot produce gradients to the prototype. Furthermore, routine inspection of all prototypes and distortions results in the elimination of the occasional peculiar pattern. Finally, we have argued here and elsewhere that prototype gradient effects are manifested most

Finally, the optimal model was a mixed model in which the exemplar contribution to classification involved the retrieval of a small subset of the training patterns. Exemplar-based models of classification are invariably evaluated under the assumption that the full set of training instances enter into a categorical judgement such as recognition or classification (e.g., Nosofsky, 1991; Nosofsky & Zaki, 1998). Rarely has an exemplar model been evaluated in which a subset of training instances is consulted for categorical decisions, presumably because retrieval of less than the full set can only lead to a weaker fit. A recent exception was provided by Storms et al. (2000), who found that typicality judgements and category naming times for natural category members could be fitted by an exemplar model that assumed retrieval of the 10 most frequent exemplars of the category. In the present study, when a mixed model is considered, the focus is less on a single influence (e.g., prototype vs. exemplar) but rather on mutual influences for different segments of the data space. As a consequence, the feasibility of a mixed model in which the exemplar contribution is based on a restricted set of stored instances emerges more readily. Theoretical development of a theory of concepts might profit by disentangling additional conditions or constraints likely to foster strong reliance on abstractions or exemplars, while exploring whether the full set of exemplars is always retrieved when categorical decisions are made.

Original manuscript received 29 April 2006
Accepted revision received 10 December 2006
First published online 23 May 2007

REFERENCES

- Barsalou, L. W. (1989). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. J. Wyer (Eds.), *Content and process specificity in the effects of prior experiences: Advances in social cognition* (Vol. 3). Urbana-Champaign, IL: Lawrence Erlbaum Associates, Inc.
- Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition*, *29*, 1153–1164.
- Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 638–648.
- Homa, D. (1978). Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 407–416.
- Homa, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18). New York: Academic Press.
- Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 444–458.
- Homa, D., Rhoads, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 11–23.
- Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 418–439.

broadly when category size is increased, not generally and less likely when category size is small. In a related vein, the claim of learning during transfer by Zaki and Nosofsky (2004) may be true under some conditions (as when prototypes are repeated in the transfer test, as in their experiment), but their proof is based on an analysis that does not discriminate a response bias interpretation from a learning interpretation. To demonstrate learning within the transfer set, it is necessary to demonstrate performance differences among progressive segments of the transfer test, as in a simple split-half analysis.

- Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 616–637.
- Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–27.
- Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1003–1021.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.
- Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 775–799.
- Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 54–65.
- Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 3–27.
- Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of “multiple-system” phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255.
- Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 924–940.
- Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, *73*, 28–38.
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function: I. *Psychometrika*, *27*, 125–140.
- Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*, 278–304.
- Smith, J. D. (2002). Exemplar theory’s predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, *13*, 437–442.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the midst: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1430.
- Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 3–27.
- Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 800–811.
- Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar based information in natural language categories. *Journal of Memory and Language*, *42*, 51–73.
- Zaki, S. R., & Nosofsky, R. M. (2004). False prototype enhancement effects in dot pattern categorization. *Memory & Cognition*, *32*, 390–389.