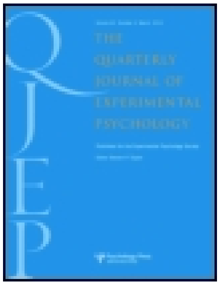

The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Array training in a categorization task

Donald Homa [a], Derek Powell [a] & Ryan Ferguson [a]
[a] Department of Psychology, Arizona State University, Tempe, AZ, USA
Accepted author version posted online: 24 Apr 2013.Published online: 28 May 2013.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2013
http://dx.doi.org/10.1080/17470218.2013.790909

Routledge
Taylor & Francis Group

Array training in a categorization task

Donald Homa, Derek Powell, and Ryan Ferguson

Department of Psychology, Arizona State University, Tempe, AZ, USA

Two components of categorization, within-category commonalities and between-category distinctiveness, were investigated in a categorization task. Subjects learned three prototype categories composed of moderately high distortions, by observing arrays containing patterns that belonged either to a common prototype category or to three different categories; a third group learned patterns presented one at a time, mirroring the standard paradigm. Following 6 learning blocks, subjects transferred to old patterns and new patterns at low-, medium-, and high-level distortions of the category prototype. The results showed that array training facilitated learning, especially when patterns in the array belonged to the same category. Transfer results showed a strong gradient effect across pattern distortion level for all conditions, with the highest performance obtained following array training on different category patterns and worst in the control condition. Interestingly, the old training patterns were classified worse than new low and no better than medium distortions. Neither this ordering nor the steepness of the gradient across prototype similarity for each condition could be predicted by the generalized context model. A prototype model better captured the steep gradient and ordinal pattern of results, although the overall fits were only slightly better than the exemplar model. The crucial role played by category commonalities and distinctiveness on categorical representations is addressed.

Keywords: Categories; Concepts; Abstraction; Modelling; Array.

The vast majority of studies on human categorization teach concepts by example, rather than definition, and typically by the presentation of the instances one at a time. This paradigm, which has changed little since the seminal publication of Hull's (1920) monograph, has proved to be productive—variables that shape categories have been identified (e.g., Homa, 1984), and a myriad of formal, quantitative models have been developed to capture these phenomena (e.g., Busemeyer & Diederich, 2010). For example, variables such as category size (the number of instances used to define a category), number of categories to be learned, category similarity, pattern distortion, and instance frequency, when manipulated in the learning phase, can be shown to dramatically affect subsequent transfer. This supports the view that concepts are dynamically modified by the kinds of experiences that define them. A major concern has been the development of formal, quantitative models that capture the resulting database.

The present study departs from the standard paradigm by varying how many instances, and of what type, occurred on a typical learning trial. The learning phase was modified to include the simultaneous presentation of a subset of the training patterns on each trial. In addition, the simultaneous presentation was further divided into two conditions. In one condition, the subset of instances contained members from the same

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

We would like to thank Gwen Vitt and Li Heing for assistance in the collection of data.

category. In another condition, the subset of instances contained instances that belonged to the three different categories. Therefore, in the array-same condition, the subject was able to inspect three patterns from the same category, which afforded the obvious advantage of noting what these instances had in common. In the array-different condition, the subject could note how the three categories differ from each other, thereby highlighting category distinctions. Comparison between these two methods was contrasted with the standard condition, here called the sequential condition, in which the same patterns were presented in learning but one at a time.

The rationale for this modification borrows from two earlier theories of categorization, Gibson's distinctive feature theory (Gibson & Gibson, 1955) and the prototype abstraction theory of Posner and Keele (1968, 1970). According to distinctive feature theory, categories are learned by the discovery of features that permit discrimination of one category from another. According to the original prototype theory of Posner and Keele, categories become defined by identification of the features or components common to the category. This latter perspective mirrors the view of James (1890) that a concept becomes apparent by abstracting the invariant component from the changing contexts in which the object appears. An attempt to integrate the distinctive features and prototype views was provided by Homa and Chambliss (1975). These authors manipulated both category size and number of categories to be learned, followed by a common transfer test. Variations in category size should selectively influence identification of what is common to a category, whereas number of categories to be initially learned in acquisition should aid in the isolation of distinctive features. Subjects initially learned two, four, or six categories, where each category was defined by either three or six instances. The results showed that each variable differentially affected later transfer, with the caveat that these variables also interacted; category size became important only when more than two categories had to be discriminated. The results were explained by a qualitative feature model in which category experience helped to isolate features that were common, distinctive, and idiosyncratic, with the variables of category size and number of categories to be discriminated functioning to isolate the various feature types.

In the present study, the presentation of an array containing instances from the same category was hypothesized to facilitate both learning and later transfer, relative to the standard sequential presentation, since commonalities to a category should be readily promoted by this manipulation. That is, simultaneous viewing of three instances reduces memory demands, relative to sequential presentation, since commonalities in the latter require accurate retention of previous instances. The simultaneous presentation of instances from different categories was hypothesized to be facilitative as well, because simultaneous viewing of three patterns from different categories should better promote knowledge of what makes the categories different relative to sequential presentation. We also anticipated that the array-same-category condition would result in the most rapid learning of all, since distinctive training cannot become effective until these features are identified as such and also common to that category.

The present study shares similarities with the free sorting task (e.g., Ahn & Medin, 1992; Regehr & Brooks, 1995). For example, in the Regehr and Brooks (1995) experiment, the entire ensemble of stimuli was simultaneously available, and subjects were instructed to sort the stimuli into two "natural" groups. Most subjects formed groups by seeking out an invariant component, suggesting that subjects have a preference for single-dimension sorts. A similar bias was also obtained by Ashby, Queller, and Berretty (1999) who explored the learning of concepts in the absence of feedback. More recently, Pothos and Close (2008) demonstrated, in a free sorting task, that classifications using one versus two dimensions could be predicted by the optimal set of similarity measurements across the entire stimulus space. However, differences between the present paradigm and the free sorting task are worth noting: (a) Subjects in the present task only saw random subsets of three patterns from the full ensemble of learning patterns; (b) a learning phase was used in which feedback was provided that was consistent with prototype constraints; and (c) whatever commonalities existed within patterns of the same category in the present study were based on similarity of patterns or pattern components rather than invariant components. Nonetheless, these studies support the hypothesis that subjects readily search for commonalities among patterns, an outcome that should favour the learning of arrays containing patterns from the same category as opposed to arrays containing patterns from different categories.

Subjects viewed 27 patterns on each learning block, nine from each of three categories. In the sequential control condition, this required, of course, 27 trials per training block. In the simultaneous conditions, the 27 different patterns were presented in nine arrays per block. On subsequent learning blocks, the subsets of three were randomized, resulting in unique arrays on each training block. Following six learning blocks, all subjects were transferred to a common set of patterns, including old, new, and prototype patterns. Each subject made a double response to each transfer pattern, first judging whether the pattern was old or new and then identifying to which category the pattern belonged.

The transfer test also included random patterns —patterns generated from prototypes different from those used in learning and therefore unrelated to the categories in the learning phase. On the transfer test, the subject was given the option to assign a pattern to one of the three learned categories or to a "junk" category (a manipulation successfully used previously, e.g., Homa, Burruel, & Field, 1987; Homa, Hout, Milliken, & Milliken, 2011). The rationale for including random patterns in the transfer phase was because array training with patterns from different categories was hypothesized to best highlight category distinctiveness. As a consequence, we anticipated that array training with different category patterns would result in the highest correct assignment of random patterns into the junk category.

Finally, the transfer results lend themselves to formal modelling. We selected, as training patterns, stimuli that were moderately high-level distortions of the prototype. These moderately high-level distortion patterns—forms composed of nine connected dots in a 50×50 grid—have the property that they are neither very similar to each other nor to other category patterns, regardless of their similarity relationship to the category prototype (Homa, 1978; Homa, Proulx, & Blair, 2008), a property verified by the multidimensional scaling analyses presented later. Because transfer patterns included new instances that were low, medium, and high distortions of each training prototype, prototype models of classification (e.g., Homa et al., 2008; Smith & Minda, 2002) must predict a steep gradient across old–new prototype similarity, regardless of condition. In contrast, new transfer patterns had a weak and inconsistent similarity relationship to the old (training) patterns. The upshot of this, detailed later, is that exemplar-based models of classification (e.g., Nosofsky, 1988; Nosofsky & Johansen, 2000) must predict a diminished gradient across old–new prototype similarity for each condition.[1]

Method

Participants

The participants were 74 undergraduate students enrolled in an introductory psychology course at Arizona State University who received class credit for participation. Participants were randomly assigned into the three learning conditions, with the sole constraint that each condition had approximately the same proportion of males and females. The subjects were between 18 and 24 years of age, with a median age of 19. The data for subjects were deleted for showing no learning improvement

[1] Fine-grained modelling of whether array training sharpened or broadened category boundaries, as might be revealed by changes in dimensional weights, are precluded with these stimuli, at least at the current time. The reason is that the functional dimensions of ill-defined patterns are obscure (Neisser, 1967). The overriding advantage of our stimuli, detailed in the discussion, is that they permit manipulation of pairwise similarities not possible with simpler stimuli composed of well-defined dimensions.

across the six training blocks, two from the sequential (control) condition, and four each from the same-category and different-category array conditions. All analyses were based on 64 participants, 21 in the sequential and same-category array condition, and 22 in the different-category array condition.[2]

Materials and apparatus

Stimuli were connected-dot patterns formed from nine dots arranged in a 50 × 50 matrix, similar to those used in previous research (Homa, 1978). Initially, three patterns were randomly generated to serve as the category prototypes for the three categories. The exemplars for each category were formed from distortions of these three prototype patterns, based on statistical decision rules. All training patterns were moderate–high distortions of the category prototype, with an average vertex displacement of 4.03 units/vertex from its prototype and a range of 3.60–4.25. During the learning phase, each category was represented by nine different exemplars. These 27 patterns were each displayed once during each learning block. The learning phase consisted of six learning blocks.

For each category, 45 additional new category patterns were generated for use in the transfer phase. Of these, five were low-level distortions, five were medium distortions, and five were high distortions for each of the three categories. Low-level distortions had an average distance moved per vertex of 1.20 units; the medium- and high-level distortions had an average distance moved per vertex of 2.80 and 4.60 units, respectively (Posner, Goldsmith, & Welton, 1967).[3] In general, the low-level distortions look similar to each other, whereas the high-level distortions share little obvious similarity to each other. Fifteen additional "junk" patterns were created based on distortions from other prototypes and had an average distance moved per vertex to any of the learning patterns of 10–15 units. The transfer set also contained each category prototype, resulting in a transfer set of 90 different patterns.

Table 1. *Mean objective distance and estimated multidimensional distances for selected pattern pairs*

Pattern pair	*Obj dist*	*MDS*
P to M/H	4.03	*0.638*
P to L	1.30	0.206
P to M	3.04	0.481
P to H	4.99	0.789
MH to MH	5.44	*0.861*
MH to L	4.19	*0.663*
MH to M	4.82	*0.763*
MH to H	6.12	*0.968*
L to L	1.82	0.289
L to M	3.24	0.513
L to H	5.09	0.805
M to M	4.12	0.652
M to H	5.65	0.894
H to H	6.63	1.049

Note: Obj dist = objective distance. MDS (multidimensional scaling) values in italics were estimated from the linear relationship between MDS and objective distance, where P = prototype, L, M, and H refer to low, medium, and high distortions from the prototype, respectively, and MH refers to patterns defined as medium–high pattern distortions.

Stimulus presentation and data recording were done using E-Prime v1.2 software. Stimuli, feedback, and instructions were presented on the screen, while participants made their responses using the keyboard. The stimuli were generated at a resolution of 300 × 300 pixels and were presented on a display with a resolution of 1,024 × 768 pixels. Feedback was displayed as text in black, 32-point Arial font. Stimuli were dark blue (RGB 000, 000, 128) with lines 1 pixel wide. Both feedback and stimuli were displayed on a pure white background.

Procedure

Participants were seated at a computer workstation and were first given a brief set of instructions

[2] Statistical outcomes were not altered by the removal of these subjects.

[3] The mean distortion of a pattern is generated by a program that takes as a value its distortion level—for example, 1.20 for a low-level distortion, generating a potential infinite number of patterns at that value. The distortion level of any pattern from its prototype is computed as the mean of the Euclidean distances of each point from the corresponding value of its prototype. The resulting sample of stimuli at a particular distortion level has a mean value that typically differs slightly from this mean. For the present study, the specific mean values are shown in Table 1.

regarding informed consent and procedure for completing the experiment. The experiment was divided into two phases, a learning phase and a transfer phase. In the learning phase, participants were told that they would see a number of patterns that belonged to three groups, A, B, and C, and that their task was to learn which patterns belonged to the three groups.

Participants in the array-different-category condition (array-DC) were presented with three stimuli at once. These stimuli were arranged in an evenly spaced horizontal row, centred at a height 284 pixels below the top of the screen. The patterns were evenly spaced across the screen such that there were 31 pixels between each image and its nearest neighbour and, for the left and rightmost patterns, the sides of the screen. Each of the stimuli on screen belonged to a different group, such that there was always one stimulus from each category in any given array. Participants were made aware of this fact during their instructions and were instructed to observe all three of the patterns before making their category judgements. Presentation of the stimuli was randomized such that exemplars from each category could appear in any of the three screen positions. The assignment of category names (A, B, C) to the three different prototypes was randomized for each subject. Participants made their responses using the keyboard, responding from left to right. After all three responses, the correct responses were displayed below their corresponding patterns. This feedback was displayed for 3 s, after which the next set of stimuli was displayed. For the initial two blocks of training, each array was available for 7.5 s regardless of how rapidly the participant responded.[4] For Blocks 3–6, presentation time was reduced to 6 s.

Participants in the array-same-category condition (array-SC) were also shown three patterns at once, and these stimuli were again arranged in an evenly spaced horizontal row, centred at a height 284 pixels below the top of the screen. Similarly, the patterns were again evenly spaced across the screen such that there were 31 pixels between each image and its nearest neighbour and, for the left and rightmost patterns, the sides of the screen. In this condition, however, all three of the stimuli belonged to the same category. Participants were informed of this fact in their instructions, and they were only required to give one response. They were again encouraged to examine each of the patterns before responding. After their response, the correct response was displayed beneath the centre stimuli for 3 s, after which the next set of stimuli was displayed. As was the case for the array-DC condition, the array was shown for 7.5 s on Blocks 1–2 and was reduced to 6 s on Blocks 3–6, resulting in an average of 2.5 and 2.0 s per pattern, respectively, in the array conditions. An example of a same-category array and a different-category array is shown in Figure 1.

Participants in the sequential (SEQ) condition were shown only one pattern at a time. This pattern was centred on the screen and displayed at the same size and resolution as the patterns in the simultaneous conditions. Participants made their response to the pattern, after which the correct response was displayed below the pattern for one second. Each array was presented for study for 2.5 s for Blocks 1 and 2 and for 2 s on Blocks 3–6. As a consequence, each pattern, regardless of condition, was shown for 2.5 s on Blocks 1 and 2 and for 2 s on Blocks 3–6.

[4] In an initial pilot experiment, we discovered that subjects in the array-same condition often responded rapidly in the learning phase, thereby diminishing later performance on the recognition test. For example, in the pilot study, the mean latency in learning was about 2 s for the sequential and array-different-category conditions and less than one second in the array-same-category conditions, a disparity that was manifested across all training blocks. As a result, we opted to force all subjects to have available roughly equal processing time for each stimulus by keeping the display visible for a time that was comparable across conditions. Whether level of processing in the conditions was equated by this manipulation is unclear (a concern mentioned by one of the reviewers), since we cannot assume that objective time per pattern was equal to subjective processing time per pattern, although any disadvantage of our manipulation should have occurred in the same-category array condition. An additional disadvantage of equating mean time per stimulus in learning is that response latencies in learning are rendered less meaningful since responses probably include a proportion of completed but delayed responding. As a result, latencies are not reported in the present study.

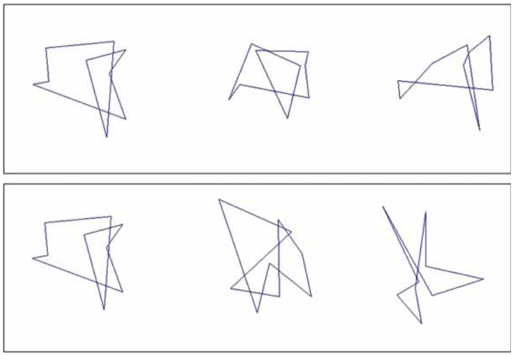

Figure 1. *Examples of stimulus arrays used in array learning. Top row shows patterns from the same category; bottom row shows patterns from different categories.*

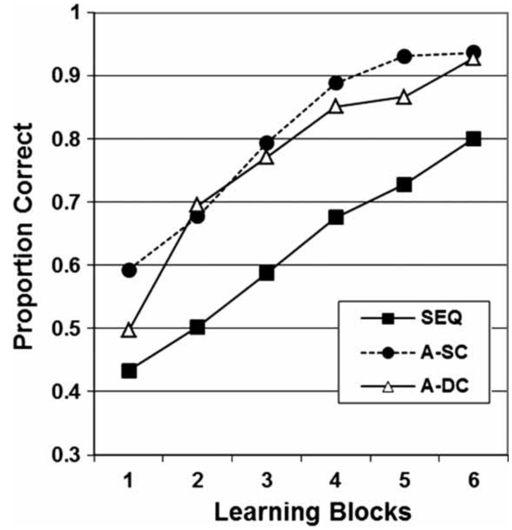

Figure 2. *Mean proportion correct learning across blocks as a function of condition (SEQ = sequential; A-SC = array-same-category; A-DC = array-different-category).*

The transfer phase was the same for all three conditions and consisted of the sequential display of all 27 learning patterns, the 45 additional patterns, the 15 junk patterns, and the three prototype patterns, for a total of 90 patterns. Participants were instructed to make both a recognition judgement and a category judgement for each pattern. The pattern was displayed as in the sequential condition along with instructions above it. First, "Is this pattern old?" and, after the initial response of yes or no, indicated by Y or N on the keyboard, the instructions "categorize this pattern" appeared above the pattern. After the participants' response, the next pattern was shown, again with the recognition instructions. No feedback was given during the transfer phase.

Results

Learning data

Figure 2 shows mean accuracy on each learning block, shown separately for each condition. A repeated measures analysis of variance (ANOVA) revealed a main effect of learning blocks, $F(5, 305) = 80.68$, $MSE = 12.01$, $\eta^2 = .569$, $p < .001$, and condition, $F(2, 61) = 14.76$, $MSE = 58.20$, $\eta^2 = .326$, both ps $< .001$. Generally, performance improved by about 40% from the initial to the terminal block, with performance on the two array presentations (same category and different category) exceeding performance on the sequential condition ($p < .05$, Bonferroni).[5] The Block × Condition interaction was not significant, $p > .20$.

Transfer: Classification

Figure 3 shows the mean classification accuracy on the transfer test for each of the conditions. Overall, classification was highest on the category prototype (.833), followed by new–low (.802), old training patterns (.719), and new–medium (.681), and lowest for new–high distortions (.481). An analysis on the new transfer patterns revealed a significant effect of distortion level, $F(2, 122) = 33.23$, $MSE = 284.99$,

[5] Although not the focus of the present study, we anticipated that the array-same-category patterns would have an advantage early in learning. Differences in learning rates, however, are compromised by potentially different guessing rates and strategies for the three conditions. Nonetheless, an analysis of the initial block revealed that the array-same-category condition resulted in a significantly higher proportion of correct responses than the sequential condition ($p < .05$) and a marginally significant advantage compared to the array-different-category condition ($p < .07$). On an individual level, over 70% of subjects in the array-same-category condition correctly identified over 50% of the category members on the initial trial versus only 36% of the subjects in the array-different-category condition.

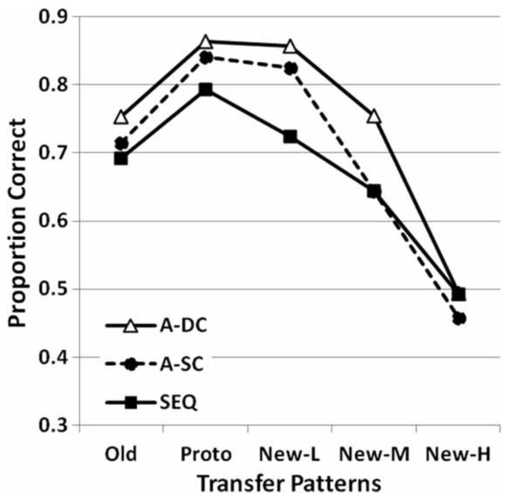

Figure 3. *Mean classification accuracy of transfer patterns (old, new–low, new–medium, new–high, prototype; proto = prototype; L = low; M = medium; H = high) as a function of learning condition (SEQ = sequential; A-SC = array-same-category; A-DC = array-different-category).*

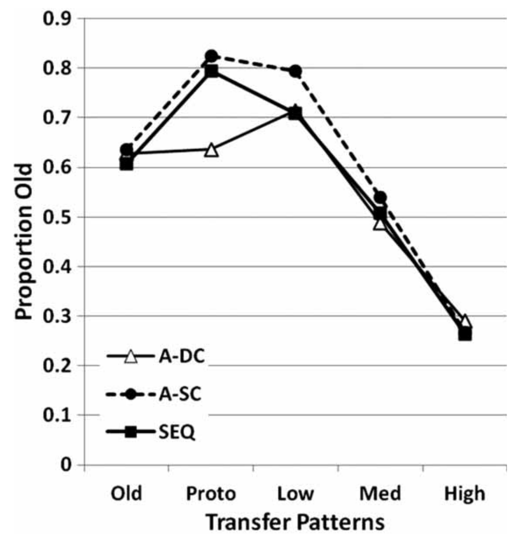

Figure 4. *Mean likelihood of calling a pattern old, as a function of pattern type (old, new–low, new–medium, new–high, prototype; proto = prototype; med = medium) and condition (SEQ = sequential; A-SC = array-same-category; A-DC = array-different-category).*

$\eta^2 = .353$, $p < .001$, and a significant Condition × Distortion level interaction, $F(4,\ 122) = 2.63$, $MSE = 291.77$, $\eta^2 = .079$, $p < .05$; the main effect of condition was not significant, $p > .10$. The interaction was due to the substantial 10–14% difference in classification favouring the array conditions versus sequential condition on the low-level distortions (Bonferroni, $p < .05$ in each case), the 11% advantage of the array-different-category condition compared to the sequential and array-same-category condition on the medium-level distortions (Bonferroni, $p < .05$), and the nonsignificant difference when patterns were high-level distortions.

Pairwise test revealed that classification accuracy for pattern types was similar for the three conditions, generally revealed as prototype = low > old ≥ medium > high ($p < .05$ contrasts). The exceptions were low = old in the sequential condition and old > medium in the array-same-category condition.

Correct classification of the unrelated patterns into the junk category was moderately accurate and did not differ among the conditions, $p > .20$, SEQ = .530, array-SC = .629, array-DC = .567. A compositional analysis (Homa et al., 1987) was performed on each condition. In this analysis, purity measures what proportion of items assigned to a category is correct.[6] This analysis revealed that purity was highest for the array-DC condition (.730), intermediate for the array-SC condition (.700), and lowest for the SEQ condition (.667).

Transfer: Recognition

Figure 4 shows the likelihood that the transfer items were called "old", shown separately for each condition. The main effect of item was significant, $F(4,\ 244) = 114.81$, $MSE = 393.758$, $\eta^2 = .653$,

[6] Purity differs from the traditional hit rate by adjusting performance to reflect how often incorrect patterns, such as random patterns and patterns from the other categories, are also assigned to a category. For example, if 10 out of 15 patterns are correctly classified into a category, the hit rate would be .667. However, if this subject also erroneously included 5 patterns from the other learned categories as well as 10 random patterns, then only 10 of the 30 patterns assigned to that category would be correct. In this case, the purity value for that category would be .333. Additional details of this type of analysis are contained in Homa et al. (1987).

$p < .001$, as was the Condition × Item interaction, $F(8,\ 244) = 2.25$, $MSE = 393.758$, $\eta^2 = .069$. However, the main effect of condition was not, $p > .20$. Overall, the likelihood of calling an item "old" was highest for the category prototype (.752) and new–low distortions (.739), lowest for the new–high distortions (.273), and intermediate for the old training patterns (.623). The Condition × Item interaction was due to similar likelihoods of calling the various items old except for the category prototype; the sequential and array-SC conditions had (false) recognition rates for the category prototypes of .794 and .825, respectively; this rate was significantly reduced for the array-DC (.636). Subsequent Bonferroni tests revealed that this reduced level of false alarming to the category prototype was significantly lower for the array-DC condition, $p < .05$.

Pairwise test revealed that the likelihood of calling a pattern old mirrored that of the classification results: Prototype = low ≥ old ≥ medium > high ($p < .05$ contrasts). The exceptions were that low = old in the sequential condition, and old = medium in the array-same-category condition.

Modelling of the transfer data

Formal modelling was initially applied to the classification results. The best fitting parameters were then maintained to fit the recognition results.

To model the transfer data in the classification phase, a measure of similarity is needed. Fortunately, we already have a fairly complete set of objective pattern distances and similarity relationships for these patterns at various levels of distortion (Homa et al., 2008). In brief, pattern pairs were rated on a 9-point scale for similarity, where the total pool of patterns contained the category prototype and three examples from each distortion level for each category. These ratings were scaled in dimensions 1–6.

Table 1 shows the objective and multidimensional scaling (MDS) distances for the three-dimensional solutions for the critical pattern comparisons that were used in all modelling.[7]

Initially, two general models were considered, a version of the generalized context model (Nosofsky, 1988; Nosofsky & Johansen, 2000) and a prototype model (Smith & Minda, 2002). In both models, the similarity between any two patterns is an exponential function of MDS distance. Let the multidimensional distance between any two patterns, i and j, be noted as d_{ij}—that is:

$$d_{ij} = \left\{ \sum (x_{ik} - x_{jk})^2 \right\}^{1/2} \tag{1}$$

where, x_{ik} and x_{jk} are the values of patterns i and j on dimension k. Typically, the similarity between patterns i and j, s_{ij}, is an exponential function of their separation in multidimensional space—that is, s_{ij}, the similarity between these patterns is:

$$S_{ij} = \exp(-cd_{ij}) \tag{2}$$

The parameter c typically functions as a scaling (sensitivity) parameter (e.g., Nosofsky & Johansen, 2000) and determines the degree of discriminability among the patterns, with larger values of c reflecting enhanced discriminability.

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

[7] Two cautionary notes are warranted. First, the multidimensional distances were derived from the same prototypes and patterns of the same distortion levels as those applied here. However, the particular patterns used in scaling were different from those used in our learning and transfer. Although some variation in multidimensional distances are inevitable when different patterns of the same distortion level are used, we have found that most patterns cluster tightly around the distances reported here. Second, the model fits did require one estimate not provided by the multidimensional scaling—the distance of a medium–high pattern to the low, medium, and high distortions. The estimated distances were based on the roughly linear relationship between objective within-category distance and multidimensional distance in three dimensions. As is clear from Table 2, the estimated values do fall within the contrasts provided by multidimensional scaling for the remaining patterns. That is, the distance of a medium–high to low, medium, and high distortions is slightly greater than those distances obtained for medium-level distortions to these contrasts and slightly less than those obtained for the high-level distortions. Neither of these concerns should bias the model fits explored here.

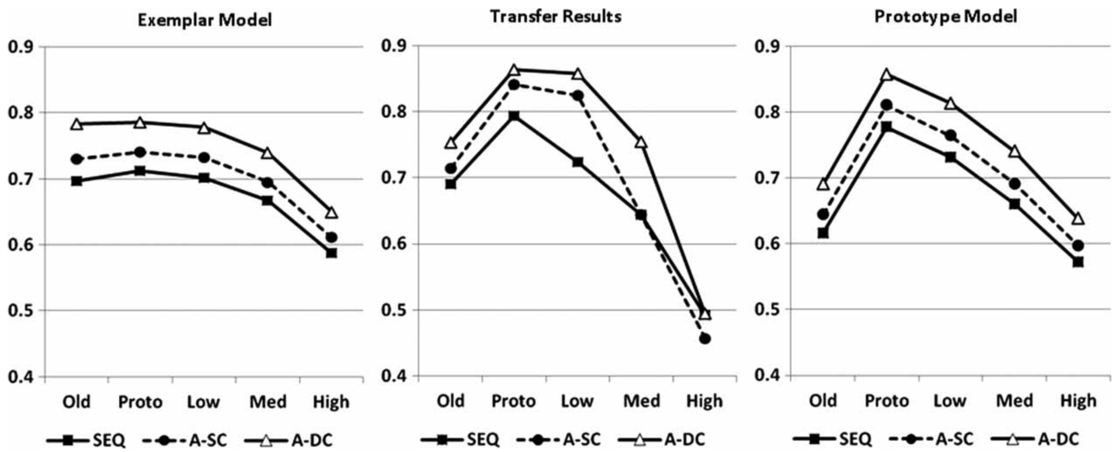

Figure 5. *Best fitting classification results for exemplar (left panel) and prototype (right panel) models for each condition (SEQ = sequential; A-SC = array-same-category; A-DC = array-different-category); results are shown in middle panel; proto = prototype.*

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., Homa et al., 2008; 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,

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

We again assume that similarity of pattern i to prototype A, B, or C is exponentially related to the multidimensional distance between these two—that is, for $i \in A$,

$$S_{iPA} = \exp(-g d_{iPA}) \quad (5)$$

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—that is, c and g are free and independent parameters. As was the case with parameter c, increased values of g should reflect enhanced discriminability among the training prototypes.

Figure 5 shows the best least square fits for the transfer results on the classification test. For each comparison, the best fitting exemplar model is shown in the left panel, the best fitting prototype model is shown in the right panel, and the actual results are shown in the middle panel.[8]

Mean deviations between obtained and predicted values favoured the prototype fits for each condition: The upper portion of Table 2 shows the mean absolute deviation between observed and predicted classification performance, the magnitude of the gradient across new distortion level, and the magnitude of the gradient when the prototype is included in the gradient. For the sequential, array-same-category, and array-different-category conditions, the deviation between observed and predicted values was .046, .083, and .072, respectively, for the exemplar model; for the prototype model, the corresponding values were .039, .069, and .055, respectively. More importantly, the exemplar-based model cannot adequately predict

[8] Fits were based on condition means for each item type.

Table 2. *Exemplar and prototype model fits to classification and recognition results, as a function of training conditions*

		Sequential			*Array-SC*			*Array-DC*		
Transfer	*Contrasts*	*Obs*	*Exem*	*Proto*	*Obs*	*Exem*	*Proto*	*Obs*	*Exem*	*Proto*
Classification	Mean abs dev	—	.046	.039	—	.083	.069	—	.072	.055
	Gradient: low–high	.232	.114	.159	.368	.121	.167	.364	.128	.175
	Gradient: proto–high	.302	.124	.205	.384	.130	.213	.370	.136	.219
	Old–new low	−.033	−.005	−.115	−.111	−.002	−.120	−.135	+.005	−.123
	Exemplar c		1.659			1.812			2.076	
	Prototype g		1.213			1.338			1.553	
Recognition	Mean abs dev	—	.126	.119	—	.138	.130	—	.091	.096
	Gradient: low–high	.445	.081	.114	.527	.090	.128	.424	.112	.164
	Gradient: proto–high	.531	.088	.157	.558	.098	.176	.347	.122	.225
	Old–new low	−.101	+.006	−.086	−.159	−.002	−.097	−.087	+.006	−.125
	Exemplar k_e		2.903			2.164			2.131	
	Prototype k_p		0.668			0.526			0.562	

Note: Abs dev = absolute deviation; proto = prototype; obs = observed; exem = exemplar; SC = same category; DC = different category.

the shape of the transfer results. In particular, the exemplar model fails for three reasons: (a) It underpredicts the magnitude of the generalization gradient across pattern distortion level; (b) it predicts that new patterns cannot be classified better than the old instances; and (c) it underpredicts performance on the category prototype. Some of these concerns obtained for exemplar-based models of classification are not new (Homa et al., 2008; Homa, Sterling, & Trepel, 1981), but they are especially visible in the present study. The reason for this shortcoming is that the old training patterns—themselves moderately high-level distortions of the category prototype—are not very similar either to other training patterns or to new patterns at any level of distortion of the category prototype. In contrast, the new patterns of low, medium, and high distortion have a strong (and increasingly reduced) similarity relationship to the category prototype. Finally, parameters assumed to reflect discriminability following training (c for exemplar, g for prototype) were greatest in the array-different-category condition and smallest for the sequential condition, an outcome consistent with the hypothesis that category distinctions should be enhanced by array-different-category training.

Exemplar-based models of classification oftentimes invoke an additional parameter that assesses whether judgements are made probabilistically or deterministically (Nosofsky & Johansen, 2000). The additional of this parameter (γ) did little to improve fits and does not alter the main conclusions.[9]

Modelling of recognition

Modelling of the recognition results maintained the parameters obtained for classification fits to the prototype and exemplar model. The sole difference was the estimation of a recognition threshold parameter, optimized for the both exemplar and prototype models. The simplest assumption is that the subject calls a pattern old if its summed similarity to all patterns exceeds some criterion (Nosofsky & Zaki, 1998). Specifically, for the exemplar model, the probability that pattern i is called "old" is determined by summing the similarity of this pattern to all study items in

[9] Addition of a γ parameter, used to assess whether the subject responded in a deterministic or probabilistic manner, did little to improve the fits, decreasing the mean absolute deviations by .002, .004, and .010 for the sequential, array-SC, and array-DC conditions, respectively. More critically, the addition of this parameter did little to improve the underprediction of the gradient across distortion level or the underprediction of the category prototype.

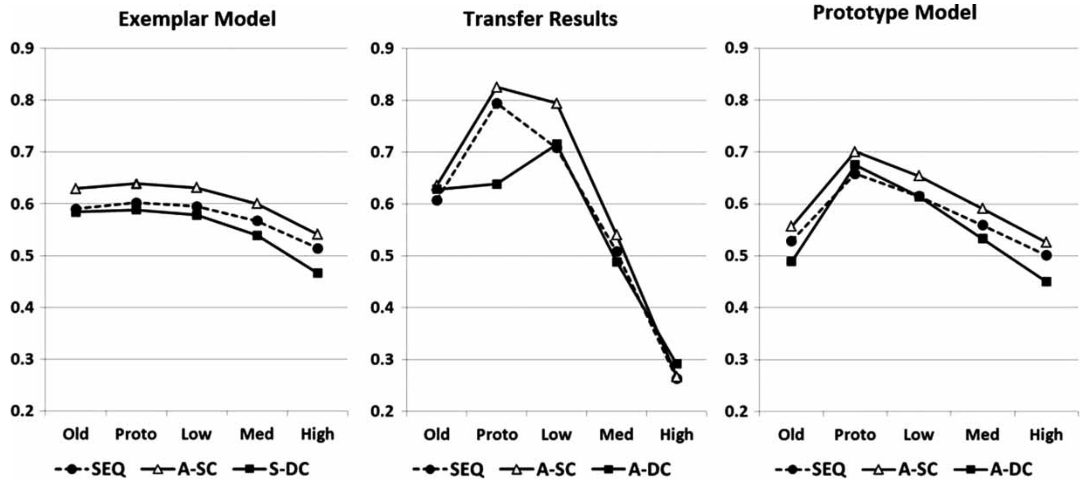

Figure 6. *Best fitting recognition results for exemplar (left panel) and prototype (right panel) models for each condition (SEQ = sequential; A-SC = array-same-category; A-DC = array-different-category); results are shown in middle panel; proto = prototype.*

Categories A, B, and C and then noting whether it exceeds some criterion, k—for example:

$$P(\text{“old”} \mid S_i) = \frac{\sum S_{ij}}{\left(\sum S_{ij} + k_e\right)} \qquad (6)$$

where k_e is the recognition threshold for the exemplar model.

A similar expression is used for the prototype model, except that the summed similarity is to all three prototypes, and the recognition threshold is k_p.

Figure 6 mirrors the display shown in Figure 5, except the probability of calling a pattern old is shown. The least square fit of the exemplar model is shown in the left panel, the prototype model is shown in the right panel, and the actual results are shown in the middle. The bottom portion of Table 2 shows selected contrasts for each model as well as the estimated criterion parameters. Although deviations between observed and predicted were greater than those for the classification results (averaging about .10 for each model), the exemplar model again fared worse in predicting the shape of the generalization gradient and the patterning of oldness proportions for old, prototype and new, low distortions. In particular, the exemplar model again failed to predict the steep gradient between the prototype and high distortions. Overall, the exemplar model predicted only 23% of the actual recognition gradient whereas the prototype model predicted 42% of this gradient.

Discussion

The present study introduced a novel modification to the typical category paradigm by presenting multiple patterns in an array format in learning. The rationale was that commonalities within a category or distinctions among categories might be highlighted, depending upon how the array was structured, compared to the standard paradigm of singly presenting patterns. The results provided mild support for these hypotheses—learning on the initial block was significantly speeded by array training when the patterns belonged to the same category compared to the sequential presentation of patterns and marginally superior to array training that contained patterns from different categories. In contrast, transfer performance favoured array training again, but when the training patterns were drawn from different categories. This latter result suggests that category distinctiveness may be more important than within-category commonalities, at least given the constraints of the present experiment. What seems likely is that both factors contribute to categorical knowledge, with the caveat that within-category may be more readily available early in learning. Taken together, support was found for both category distinctiveness training (Gibson & Gibson, 1955) and identification of components common to a category (e.g., Posner & Keele, 1968, 1970). The results of the present study also lend support to previous research that argued for the importance of category shaping variables to category knowledge, with category size selectively influencing the identification of category commonalities and number of categories to be learned enhancing knowledge of category distinctiveness (Homa & Chambliss, 1975).

We are currently exploring whether array training may afford an optimal way to teach expertise. Tanaka, Curran, and Sheinberg (2005) argued that expertise in the identification of wading birds and owls might be accomplished via extensive training at a particular level of the taxonomy. A reasonable expectation, given the results of the present study, is that array training with examples from the same category may be optimal for the isolation of common properties needed to make judgements *higher* within the taxonomy—for example, basic-level category judgements. However, learning of finer grained distinctions at the subordinate or even deeper levels may be profitably gained via array training with stimuli belonging to different categories. This would be consonant with the finding that learning is speeded by array training with members from the same category, but later transfer could profit more by training with members from different categories.

Finally, formal modelling generally supported a prototype model over an exemplar model of classification. The major shortcoming of the exemplar model (e.g., Nosofsky, 1988; Shin & Nosofsky, 1992) was the failure to adequately predict three results: (a) the steep gradient across similarity to the category prototype obtained in all three conditions for both classification and recognition; (b) the substantially higher classification of new low-level distortions than the training patterns; and (c) the higher classification accuracy and higher false-alarm rates for the category prototype in each condition than the training patterns. Generally, the prototype model better fitted each of these concerns, although even the prototype model slightly underpredicted the magnitude of the gradient across pattern distortion level. More important, the prototype model better fitted the ordinal relationship obtained for the classification and recognition of the old, new, and prototype patterns than did the exemplar model (Wills & Pathos, 2012). In particular, the exemplar model predicted similar classification accuracy and oldness ratings for the old, prototype, and low-level distortions; the prototype model did not. The results clearly reflected the latter outcome.

The reason for the poor exemplar fit to the classification is probably due to two factors: The training patterns were moderately high-level distortions from the category prototype and had, as a consequence, a weak similarity relationship to each other; and these training patterns had a minimal and fairly uniform similarity relationship to the transfer patterns, regardless of their distortion level to the category prototype. In contrast, the new transfer patterns had a lawful and substantial similarity relationship to the category prototype. This relationship is readily seen in Figure 7, which

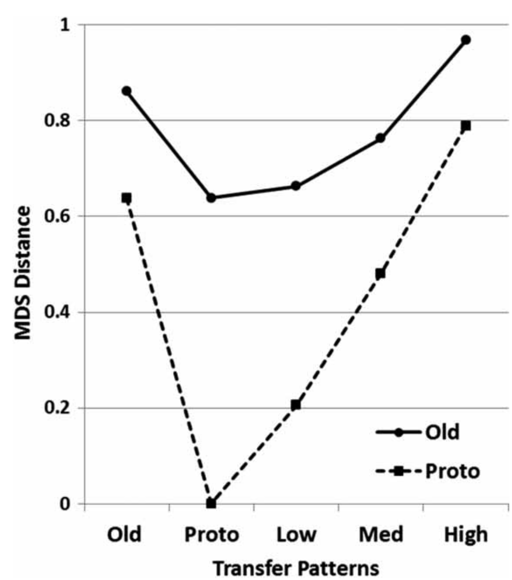

Figure 7. *Mean MDS (multidimensional scaling) distances for the training patterns and the category prototype to transfer patterns of different distortion level to the category prototype; proto = prototype.*

shows the MDS distances between training patterns or the prototype to the old patterns and new patterns at all levels of distortion.

These particular properties, critical to the present study, cannot occur in most experiments that employ binary patterns that vary along a small set of dimensions and require classification into two categories. For example, the heavily researched 5/4 paradigm introduced by Medin and Schaffer (1978) and the source of extensive investigation by later researchers (e.g., Minda & Smith, 2002; Zaki, Nosofsky, Stanton, & Cohen, 2003) can generate a stimulus population of 16 different stimuli. Because of the limited stimulus pool, it is impossible to generate training patterns that are high-level distortions from the category prototype that are also uniformly distant (or nearly so) from novel patterns at manipulated levels of similarity from the category prototype. In fact, we know of no study using binary-valued patterns that fit the properties considered critical here.

A limitation of the modelling fits in the present study should be mentioned. Fine-grained modelling of whether array training sharpened or broadened category boundaries, as might be revealed by changes in dimensional weights, cannot be made. This analysis is precluded with the stimuli used in the present study, at least at the present time, because the functional dimensions of ill-defined patterns are obscure and not readily identifiable (Neisser, 1967). The type of simple modelling provided here is possible because pairwise Euclidean distances in *N*-dimensional space among ill-defined patterns are readily obtained and unaffected by dimensional rotation. In addition, pairwise distances are minimally changed, and the ordinal rank-ordering of distances for each pattern pair are largely unaffected by the number of scaled dimensions in an MDS space. However, the identification of critical dimensions, a preliminary requirement for analyses that explore dimensional weighting (perhaps based on the array manipulations used here) is considerably more difficult. Without an objective way to identify the dimensions (e.g., Nosofsky, 1987), each rotation within a dimensional solution would produce a different set of weights. Consequently, we cannot dismiss the possibility that exemplar-based models of classification that use dimensional weighting for same- and different-category array conditions might better capture the steep gradients across pattern distortion levels obtained in the present study.

Nonetheless, the overriding advantage of using the ill-defined patterns used in the present study is that they permit manipulation of pairwise similarities not possible with simpler stimuli composed of well-defined dimensions. More critically, the critical structural properties explored here—training on moderately high-level distortions of a category that are neither very similar to each other nor to patterns at low, medium, and high distortions of the category—are manifested in many if not most natural categories. Young and Hamer (1994) provide numerous examples of multidimensional scaling of natural categories. It is a simple matter to identify category members at various distances from the centroid (prototype) while maintaining roughly equidistance relations to other moderately high-distortion (less typical) members. For example, in a multidimensional space, robin, eagle and swan are increasingly distant from the centroid of the bird category, and these members are approximately equidistant to other high distortions like penguin and ostrich. Similarly, dog, wolf, and bear are increasingly distant from the centroid of four-footed animals and, again, roughly equidistant to other high distortions like zebra and mouse (Homa & Silver, 1976). To investigate exemplar and prototype models, it is necessary to evaluate similarity relationships that are critical to these theories, an alternative often obviated by the use of categories composed of binary-valued stimuli that are assigned to either of two categories.

In conclusion, the present study demonstrates that array training can provide useful insights into the role of common and distinctive properties that critically shape our concepts. One clear advantage of array presentation is that memory demands are reduced compared to sequential presentation, affording the opportunity to visually search for commonalities or category distinctions, depending on how the array is structured. Should this procedure provide a method for the optimization of learning and retention of concepts, there remains a host of variables that might be profitably investigated, including array size, time of inspection, quality of the exemplars, and number of categories to be learned. These variables were held constant in the present study. These manipulations might also reveal whether the various models of classification currently popular might best fit different segments of the resulting data space shaped by these manipulations.

Original manuscript received 12 April 2012
Accepted revision received 11 March 2013
First published online 16 May 2013

REFERENCES

Ahn, W. K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science*, *16,* 81–121.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61,* 1178–1199.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Thousand Oaks, CA: Sage Publications.

Gibson, J. J., & Gibson, E. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62,* 32–41.

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* (Vol. 18). New York, NY: Academic Press.

Homa, D., Burruel, L., & Field, D. (1987). The changing composition of abstracted categories under manipulations of decisional change, choice difficulty, and category size. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13,* 401–412.

Homa, D., & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory*, *104*(4), 351–359.

Homa, D., Hout, M., Milliken, L., & Milliken, A. M. (2011). Bogus concerns about the false prototype enhancement effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 368–377.

Homa, D., Proulx, M. J., & Blair, M. (2008). The modulating influence of category size on the classification of exception patterns. *Quarterly Journal of Experimental Psychology*, *61,* 425–443.

Homa, D., & Silver, R. (1976). Triadic decision making in lexical memory. *Memory & Cognition*, *4,* 532–540.

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.

Hull, C. I. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs*, *28*(Whole No. 123).

James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt & Company.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85,* 207–238.

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 275–292.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13,* 87–108.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of multiple-system phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 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*(4), 247–255.

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. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77,* 353–363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83* (2), 304–308.

Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition*, *107,* 581–602.

Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21,* 347–363.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*(3), 278–304.

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. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28,* 800–811.

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16,* 145–151.

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Review*, *138,* 102–125.

Young, F. W., & Hamer, R. M. (1994). *Theory and applications of multidimensional scaling*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1160–1173.