Discrimination and Generalization in Pattern Categorization: A Case for Elemental Associative Learning

E. J. Livesey (el253@cam.ac.uk)
P. J. C. Broadhurst (pjcb3@cam.ac.uk)
I. P. L. McLaren (iplm2@cam.ac.uk)
Department of Experimental Psychology, University of Cambridge
Downing Street, Cambridge CB2 3EB. UK.

Abstract

Peak shift effects produced with complex pattern stimuli are influenced by the level of spatial variability over successive presentations of the same training stimulus. This effect, now demonstrated in humans and animals, poses a significant problem for past models of peak shift, which assume that the spatial location of the stimulus components will not influence post-discrimination generalization. Here we present a model based on contemporary elemental associative theory which can accurately account for the effect, and furthermore does so without any of the assumptions of dimensionality on which previous models of peak shift depend. In so doing, we hope to further illuminate the nature of dimensionality, i.e. what makes a dimension a dimension.

Peak shift with 'icon' stimuli

Since Hanson's first report of the effect in 1959, peak shift has been demonstrated in many forms, using human and non-human subjects, and using a variety of stimuli (see for instance Ghirlanda & Enquist, 2003). Peak shift is observed over a series of stimuli with properties that lie along a dimension or have some systematic ordinal relationship. In conditioning experiments, subjects are typically trained to discriminate between an S+, a stimulus to which responses are reinforced, and a very similar S- to which no reinforcement is given for responding. When tested across several stimuli, subjects' peak response rate often occurs not for S+ but for a similar stimulus further along the dimension away from S- (Hanson, 1959). Likewise, in human categorization experiments, if subjects are trained to discriminate between two very similar stimuli and are then given a generalization test, accuracy is sometimes found to be highest for stimuli either side of the training stimuli rather than the training stimuli themselves. Peak shift is conventionally associated with very simple dimensional designs employing, for instance, lights of differing wavelength or tilted lines of differing angle. However, peak shift has also been shown with much more complex stimuli, the properties of which are ordered along an *artificial* dimension. For instance, patterns of small abstract shapes commonly referred to as 'icons' have been used to produce peak shift effects in both animal conditioning and human categorization (Oakeshott, 2002; Wills and Mackintosh, 1998).

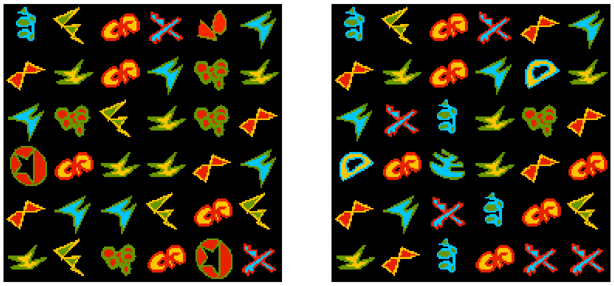

Figure 1. Example of 'icon' stimuli used in training.

Typically in these experiments, the frequency of occurrence of each icon is systematically varied from one stimulus to the next. This is done in such a way to give the resulting patterns an ordinal relationship with one another, in effect mimicking points along a dimension. Shifting from one stimulus to the next along this artificial dimension requires changing some of the icons while holding others constant. Figure 1, for instance shows a pair of training stimuli, under the conditions used for the current experiment and model, which share two thirds of their icons. The full dimensional design and frequency of occurrence of the icons in each stimulus can be seen in Table 1. The patterns of icons in Figure 1 represent stimuli 7 and 9 on an artificial dimension of 15 stimuli. Participants must learn to press a left key for stimulus 7 (S_L) and a right key for stimulus 9 (S_R). Livesey (2004) found that under some conditions, when tested on all 15 stimulus positions, participants categorize the patterns with highest accuracy at positions that are 3 steps to either side of the training stimuli (i.e. stimuli 4 and 12), and thus demonstrate a significant peak shift. However, the shape of the post-discrimination generalization gradient appears to be greatly influenced by the level of spatial variation over successive presentations of each stimulus, an effect first demonstrated with pigeons by Oakeshott (2002). The icon experiments that have produced peak shift effects have typically allowed the position of each icon within the pattern to vary randomly from one trial to the next, and in many cases have also allowed the exact number of each type of icon to vary from one trial to the next in such a way that the frequency of occurrence averages to a given proportion over many trials. In contrast, when the position of each icon and the frequency of occurrence of each icon type are both fixed, there appears to be either no evidence of peak shift at all, with a relatively linear generalization gradient (Oakeshott, 2002), or a diminished peak shift, with a similar generalization gradient but occurring over a smaller range along the dimension (Livesey, 2004). Figure 2 shows the results from 3 conditions run by Livesey (2004), demonstrating the difference in resultant post-discrimination generalization gradients after 'fixed' and 'varied' training.

Table 1. Dimensional design used in the current experiment and in Livesey (2004).

Stimulus no. & distance from nearest S		Frequency of occurrence of icons A to X in each stimulus																								
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
6	1	1	2	4	5	6	6	5	4	2	1															
5	2		1	2	4	5	6	6	5	4	2	1														
4	3			1	2	4	5	6	6	5	4	2	1													
3	4				1	2	4	5	6	6	5	4	2	1												
2	5					1	2	4	5	6	6	5	4	2	1											
1	6						1	2	4	5	6	6	5	4	2	1										
0	7							1	2	4	5	6	6	5	4	2	1									S_L
	8								1	2	4	5	6	6	5	4	2	1								
0	9									1	2	4	5	6	6	5	4	2	1							S_R
1	10										1	2	4	5	6	6	5	4	2	1						
2	11											1	2	4	5	6	6	5	4	2	1					
3	12												1	2	4	5	6	6	5	4	2	1				
4	13													1	2	4	5	6	6	5	4	2	1			
5	14														1	2	4	5	6	6	5	4	2	1		
6	15															1	2	4	5	6	6	5	4	2	1	

To compare the curvature of the post-discrimination generalization gradients produced under fixed and varied conditions, Livesey (2004) used two contrast scores calculated as a test of non-linearity. Both gradients were negatively accelerating approaching the training stimuli S and resembled a weak peak shift effect. In this case, the primary interest was the scale over which the curvature of the two gradients was occurring as it was predicted that the fixed gradient would peak earlier and decline more rapidly (that is, manifest on a compressed scale). Quadratic contrasts of the (1, -2, 1) form were calculated over points 0, 3, and 6 steps from S and again for points 0, 2, and 4 steps from S, referred to respectively as 'stretched' and 'compressed' contrasts. The significant interaction between these contrasts suggested that the curvature of the two gradients did indeed differ in scale and shape.

There is therefore evidence from human and pigeon research that spatial variability of the icon patterns affects the nature of the post-discrimination generalization gradient. This is significant because none of the models previously used to simulate peak shift along an artificial dimension can adequately account for such an effect. Peak shift experiments with icon stimuli have been cited as evidence supporting elemental associative learning theories of generalization and discrimination (for instance Blough, 1975). However, the modeling that has stemmed from this research generally makes important assumptions about the underlying dimensionality of the paradigm. This may take the form of a graded activation function over several units thought to be 'tuned' to similar stimulus properties, a strategy which has been very successful for modeling peak shift along a physical dimension. It would seem risky, however, to rely on such an assumption where the dimension itself is an artificial construct. In any case, past models of peak shift along an artificial dimension have generally assumed that the relative position of the icons should make little or no difference to the overall pattern of generalization.

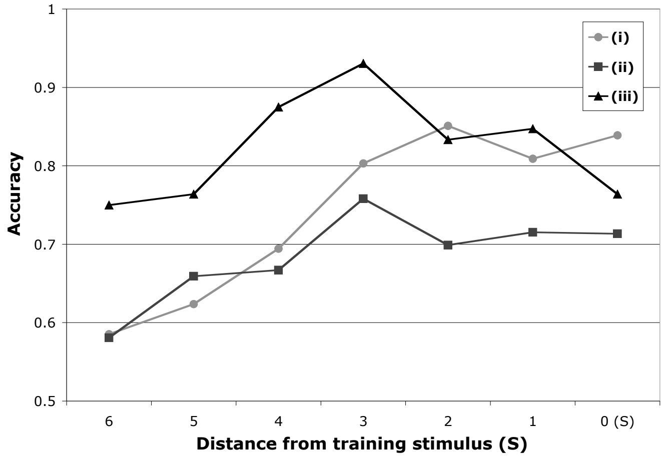

Figure 2. Post-discrimination generalization gradients from Livesey (2004). The three gradients correspond to (i) fixed icon location and frequencies, (ii) varied icon location but fixed frequencies, and (iii) varied location and frequencies.

Modeling the effect of spatial variability.

The model considered here is an implementation of McLaren and Mackintosh's (2000; 2002) general theory of elemental associative learning which has the scope to apply to discrimination and generalization effects such as peak shift. The characteristics that mark this model as a distinct departure from past attempts to simulate peak shift lie not only in the way it treats spatial variation and specificity, but also in the absence of any assumptions of dimensionality or graded activation. Instead, these factors *emerge* from the stimulus properties themselves. The mechanism through which learning proceeds is actually no more complicated than a simple delta rule linking a large number of representational units with two output units. The stimulus is represented by activation of some input units, using a relatively simple form of local coding whereby each icon in each specific location activates a single input node. Thus icon A in position 1 would be represented as activation in one input unit, while icon A in position 2 (or icon B in position 1) would be represented as activation in another. Thus input unit i is fully activated ($A_i = 1$) when that location-specific icon is present and not activated at all ($A_i = 0$) otherwise. In order to model a single subject, one needs 864 input units (24 different icons are used and the stimulus has 36 different locations in which an icon can appear).

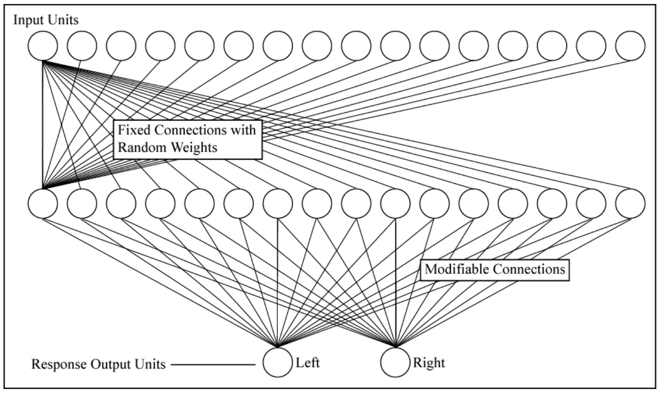

Figure 3. A schematic representation of a model based on McLaren and Mackintosh (2000; 2002)

This method of coding the input units was chosen mainly as a matter of convenience. Of course one might expect a high degree of generalization from, say, icon A in position 1 to icon A in position 2, particularly with a stimulus of this size and complexity. There are several ways one could go about modifying the inputs to express this generalization. Instead of taking this approach, here generalization can occur through the pattern of random weights that connect the input units to a second layer of hidden units. These weights are fixed throughout the experiment and allow a high degree of random variation, taking on one of two forms, depending on which second-layer unit is involved. A proportion of the second layer units have positive connections specific to a particular icon type. Positive weights varying randomly from 0 to +1 connect these units with each of the 36 location-specific inputs that represent a particular icon type, while the connections to all other input units have small, random negative weights. The rest of the second layer units are connected via weights which vary randomly between -1 and +1, irrespective of which icon type and which location the connected input is representing.

This second layer can be thought of as the "elements" described in the McLaren and Mackintosh model. Importantly, there is no assumption of dimensionality inherent in this system of representation – the elements are activated via a series of random weights with *all* of the activated inputs such that input $i = \sum A_i W_{ij}$ where A_i is the activation of input unit i and W_{ij} is the random fixed weight between input unit i and second layer unit j. If this input is negative then unit j has no activation ($A_j = 0$), otherwise if i is positive then activation of unit j, $A_j = E.i / (D + E.i)$ where E and D are constants. In effect, this means that on presentation of any specific stimulus, roughly half of the second layer units will have an activation of zero as their summed input will be negative, while the activation of the other half will vary between 0 and 1, the distribution of which is affected by the values of the constants D and E.

Each of these second layer units is then connected to two output units, one corresponding to each response (left and right key presses). These weights are initially set to zero and vary according to the delta-rule. The change in weight between second layer unit i and output j, $W_{ij} = S.\Delta.A_i$ where S is a learning rate parameter, A_i is the activation of unit i and Δ is the difference between external input, e, and summed internal input, i. The summed input i is calculated in much the same fashion as for the second layer units ($i = \sum A_i W_{ij}$). During training, the output unit corresponding to the correct response was given an external activation of 1 while the incorrect response output has an external activation of 0.

During the test phase, response probabilities for each test stimulus were calculated using an exponential form of the ratio rule. For instance, the probability of a left response, P(Left Response) = $e^{kL} / (e^{kL} + e^{kR})$ where L and R are the summed inputs to the units corresponding to left response and right responses respectively, and k is a constant held at a single value for all subjects in each group. This ratio rule can be considered a fairly standard method of calculating response probabilities from two competing response tendencies (see for example Wills, Reimers, Stewart, Suret, & McLaren, 2000).

This model provides a very close quantitative fit to the generalization gradients from Livesey (2004) given above, and accurately predicts the difference in curvature of gradients produced under fixed spatial conditions and varied spatial conditions. The crucial feature that allows the model to make such predictions is the manner in which the activation of second layer units can represent highly specific configurations of icons in particular locations and also more general patterns of icon frequency. The interaction between large numbers of these units and the summed error term governing weight change to the output units acts as a selective process, favoring units that extract stable differences between the training stimuli. In the fixed condition, the frequency of occurrence of icons (irrespective of their position within the array) and configurations of icons in specific locations are both stable across multiple presentations of the training stimuli. Second layer units that are highly activated by specific configurations of icons present in one training stimulus and not the other will govern a large component of the discrimination. However, these units will be much less activated by neighboring stimuli as configurations of icons change dramatically in just a few steps along the dimension. In the varied condition, the location of icons does not remain stable across many presentations of the training stimuli. Consequently, the units that are highly activated by predictive icons in many different locations will govern the discrimination. The activation of these units will remain relatively high in neighbouring stimuli, and will produce a broader generalization gradient and a stronger peak shift effect, as shown by Livesey (2004) and Oakeshott (2002).

A further test of the model

An opportunity to test very subtle predictions of the model arose when the current authors identified a methodological issue that needed further examination. In previous fixed conditions used by Oakeshott (2002) and Livesey (2004) the set of test stimuli were created by taking an initial stimulus and then shifting along the dimension, each time retaining the icons that are common to both and replacing the icons that are unique to one stimulus with new icons unique to the other. On each shift, if the number of copies of an icon type was to decrease then the to-be-replaced icon was chosen randomly and irrespective of which icons had been replaced in recent shifts. For instance, if the number of copies of icon K was to be reduced from 6 to 5, then each copy of icon K had a 1 in 6 chance of being replaced. Thus it was quite possible for a specific icon to change twice within a small number of steps, or conversely for an icon to remain in one position across a disproportionately large number of stimuli, as illustrated in Table 2. This adds an unwanted source of variation to the dimension. It means that the location-specific icons on which the solution to the discrimination lies may be readily replaced in test stimuli that lie close by on the dimension.

On a more plausibly constrained dimension, each location-specific icon would be present for a fixed number of adjacent stimuli (6 in the current experiment) so that, moving across the artificial dimension, the first copy of an icon to appear would be the first to be replaced, and so on. Could this subtle difference actually have an effect on the post-discrimination generalization gradient? The model proposed in this paper predicted that indeed it would. Specifically, the model predicts that accuracy should remain relatively high for a greater range of stimuli along the dimension, then decline sharply. In terms of the contrasts used by Livesey (2004), under constrained icon replacement conditions the curvature of the gradient should be more pronounced over the full length of the dimension (larger contrast score over for the stretched contrast) but flatter over the relatively nearby stimuli (smaller contrast score over the compressed dimension). This manipulation should have no effect during training as the training stimuli are too close to each other on the dimension for any icon position to change more than once between S_L and S_R. The result should therefore manifest itself only in post-discrimination generalization across the full range of stimuli. To see why we might expect this manipulation to have this effect, recall that in the fixed condition we assume that configurations of icons play a large part in determining responding. As we move along the dimension, these configurations are disrupted as the icons are changed. In previous fixed condition stimuli, any of the icon types eligible to change could change as we moved from one point on the dimension to the other. In the new, constrained fixed condition this is not so, and as a consequence some icon configurations will definitely persist for longer than would otherwise be expected to be the case. Thus the expectation is that as we move along the dimension from S+ and away from S- say, initially the constrained version stimuli will, on average tend to lose less configurations that have come into existence in S+ but were not present in S-. Similarly they will tend to lose configurations that were present in S- but not in S+. These configurations will, of course, be part of the basis for discrimination between S+ and S-, and this process will tend to lead to maintained responding congruent with that trained to S+ to stimuli containing them. Then as we move further from S+, the configurations found in S+ but not in S- must now be replaced, and responding should fall off rapidly.

Table 2. A possible progression of icons, over a series of test stimuli, for three positions within the stimulus array, under fixed conditions with random icon replacement (i), & fixed conditions with constrained icon replacement (ii).

	(i)	Position No.			(ii)	Position No.		
		1	2	3		1	2	3
Stimulus Number	1	E	J	C		E	J	C
	2	E	J	C		K	J	C
	3	E	J	H		K	J	H
	4	E	J	H		K	J	H
	5	K	J	J		K	J	H
	6	K	J	J		K	J	H
	7	K	J	L		K	P	H
	8	N	J	L		N	P	H
	9	N	J	R		N	P	O

The current experiment tested this prediction. Two very similar fixed conditions were used, one using a fixed method with the same random replacement method as devised by Oakeshott (2002), and one using strict constraints on icon replacement such that each icon in a specific location was present for six (and only six) adjacent stimuli.

Method

Participants and Apparatus 36 undergraduate students from the University of Cambridge participated in the experiment. They were randomly allocated into the two groups 'Random' icon replacement and 'Constrained' icon replacement, and did not receive payment for their participation. Participants were tested individually in a dimly lit room. The experiment was programmed using REALbasic software and run on a Power Macintosh.

Stimuli All stimuli appeared in a rectangular region in the centre of the computer screen measuring 5cm wide by 5cm high, surrounded by a thin white border. Each pattern consisted of an array of 36 icons, 6 icons wide and 6 icons high. The composition of each stimulus, in terms of the number of copies of each icon that it contained, was determined by its position on the dimension, as shown in Table 1. For each subject, 24 icons were randomly assigned a position A to X on the dimension. By way of counterbalancing the icon distributions, each randomly ordered set of icons was allocated to one subject from each group.

For both groups the position of each icon was randomized on the first trial, but then remained constant for all subsequent presentations of that stimulus. For any pair of adjacent stimuli the icons common to both stimuli remained in a fixed location for both stimuli. In the 'random replacement' condition, the location-specific copy of each icon to be replaced was chosen randomly from all the icons of that type present in the stimulus. For the 'constrained replacement' condition, icon replacement was based on the number of adjacent stimuli for which a particular location-specific icon was held constant. Once a particular location in the array was changed, it would change again when (and only when) another six steps along the dimension had been calculated. Thus, with the exception of some icons that appeared in the terminal stimuli 1 and 15, all location-specific icons appeared for 6 adjacent stimuli.

The icon stimuli were presented on every second trial, interspersed with filler trials consisting of uniform colored squares that differed slightly in hue. Two very similar shades of green were presented during training, S_L and S_R, and a wider array of colors ranging from blue-green to yellow-green were presented during the transfer test. The primary purpose of the color trials was to negate any effect of immediate contrast between patterns that would make the icon discrimination too easy.

Procedure Participants sat approximately 1m from the computer screen and were given verbal and written instructions pertaining to the training and test phases and the responses they would be required to make. For every trial, participants were required to make either a left or right key response by pressing either 'x' or '.' on the keyboard. During training, feedback appeared after every response in the form of the words "correct" or "wrong" flashed in the centre of the screen. During the test phase no feedback was given. If no response was given within 4 seconds of the stimulus appearing then the trial timed out and the message "no response" appeared on the screen.

As in Table 1, the stimuli at positions 7 and 9 were used as S_L and S_R respectively during the training phase. Trial order was randomized within eight blocks of 12 trials, containing three presentations each of S_L and S_R and three each of the corresponding filler trials, with the condition that trials alternate between icon stimuli and filler stimuli. During the transfer test, all 15 stimulus positions were tested, with an equal number of interleaved color filler trials. Trial order was randomized within blocks of 30 trials, containing one trial of each of the 15 transfer stimuli and 15 filler stimuli. In total there were 180 trials organized in six blocks.

Results and Discussion

4 people were excluded from the results as they failed to acquire the task during training, leaving each group with n = 16 for the purposes of analysis.

During training, both groups acquired the discrimination at roughly the same rate, though accuracy appeared to be slightly higher over all for the constrained group. A repeated measures anova showed a significant effect of training block (F = 15.4, p<.001), reflecting the linear increase in accuracy in both groups, while there was a marginal effect of group (F = 3.4, p = 0.075) and no interaction (F < 1, p>.05).

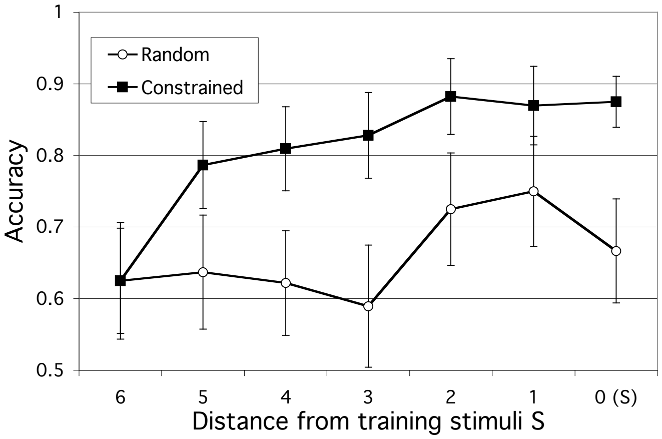

Figure 5. Mean test scores for groups Constrained and Random plotted as a function of distance from the nearest training stimulus (S).

The post-discrimination generalization gradients for both groups are shown in Figure 5. As in training, the overall level of accuracy appears to be higher for the constrained group than the random group, but again it was not reliable (F = 3.464, p = .073). Predictably, there was a main effect of stimulus (F = 5.145, p < .05), though the interaction between stimulus and group did not reach significance (F = 1.835, p > .05). Following the analysis used by Livesey (2004), quadratic trend contrasts were calculated over two sets of three stimuli, in order to test for non-linearity over a compressed range (points 0, 2, and 4) and over a stretched range (points 0, 3, and 6). These test the model prediction that the shape of the post-discrimination generalization gradient will be subtly different in each group. Over the compressed contrast, the difference between groups was in the predicted direction (less negative for the constrained group), though it was not statistically significant ($t = 0.67$, $p > .05$). Over the stretched contrast, the difference was once again in the predicted direction (more negative for the constrained group) and reached one-tailed significance ($t = 1.682$, $p = .05$). An interaction between group and contrast type was also significant ($F = 4.109$, $p < .05$, 1-tailed).

In addition, it appears that a level of accuracy significantly above chance was retained for a much larger range of test stimuli in the constrained group than the random group. All test scores along the constrained dimension were significantly above chance except for stimuli 6 steps from S ($t = 1.532$, $p > .05$ for position 6, next lowest $t = 4.708$, $p < .001$). In contrast, in the random group only positions 0, 1, and 2 steps from S were significantly above chance (lowest $t = 2.297$, $p<.05$). This may go some way to explaining the unexpected dip in accuracy at position 3 – since positions 3 to 6 differ significantly neither from chance, nor each other, it may be little more than random variation.

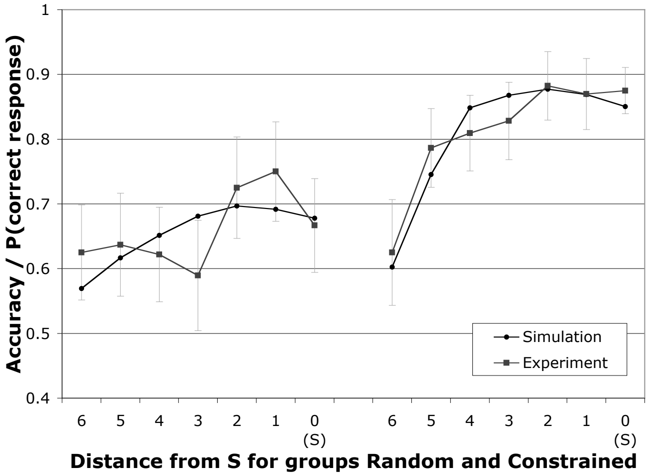

Figure 6. Model predictions fitted to overall accuracy for groups random and constrained respectively.

The experimental results appear to conform to the predictions of the model quite well, the fit is shown in Figure 6. There is some rather pronounced deviation in the random group, particularly at 3 steps from the training stimulus. However, as mentioned, accuracy for points further along the dimension does not increase significantly and it is likely that this is no more than random variation.

Conclusions

The model presented in this paper can account both for the effect of spatial variation in icon placement on the post-discrimination gradients obtained, and for variations in the degree to which our artificial dimension approximates a real one. This makes it the most robust model of its type. Furthermore the model avoids the assumptions of dimensionality on which previous models of peak shift have relied. Instead the icons are treated as location-specific stimulus components in a system of representation that allows dimensional qualities to emerge from a seemingly random pattern of activation of second-layer units. This system is an implementation of the elemental representation described by McLaren and Mackintosh (2000; 2002), and demonstrates that some contemporary elemental learning models are powerful enough to make very subtle and accurate predictions about generalization in human categorization.

References

Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 1*, 3-21.

Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour, 66*, 15-36.

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology, 58*, 321-334.

Livesey, E. J. (2004). *Representation and Discrimination: An analysis of the curvature of post-discrimination generalisation gradients.* Unpublished 1st year report, University of Cambridge, Cambridge.

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior, 28*(3), 211-246.

McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior, 30*(3), 177-200.

Oakeshott, S. M. (2002). *Peak shift: An elemental vs a configural analysis.* Unpublished PhD, University of Cambridge, Cambridge.

Wills, S., & Mackintosh, N. J. (1998). Peak shift on an artificial dimension. *Quarterly Journal of Experimental Psychology Section B- Comparative and Physiological Psychology, 51*(1), 1-32.

Wills, A. J., Reimers, S., Stewart, N., Suret, M., & McLaren, I. P. L. (2000). Tests of the ratio rule in categorization. *Quarterly Journal of Experimental Psychology Section A- Human Experimental Psychology*, 53(4), 983-1011.