Transposition and Generalization on an Artificial Dimension

J.E.M. Locke

M.B. Suret

I.P.L. McLaren (iplm2@cus.cam.ac.uk)
Department of Experimental Psychology; Downing Street
Cambridge, CB2 3EB UK

Abstract

In two experiments we demonstrate that a model based on generalization gives a good fit to data obtained when subjects have to transfer learning based on one discrimination to another on the same dimension. Experiment 1 demonstrates an effect of transposition-based transfer from pre-training to the subsequently trained target discriminations. Experiment 2 shows that this effect is unlikely to be an artifact caused by subjects' sensitivity to changes in stimulus-response assignments.

Introduction

McLaren and Suret (2000) were the first to report an effect of transfer along a continuum (Lawrence, 1952) with an artificial dimension constructed by morphing in equal intervals between two similar faces. In a later paper Suret and McLaren (2002) demonstrated that the transfer effects they found with this dimension could be explained in terms of the generalization gradients set up by discrimination training. They developed a connectionist model of discrimination learning which is a modification of the McLaren and Mackintosh (2000, 2002) model and was able to give a satisfactory fit to the data by exploiting the similarity relationships between faces on the dimension to produce a generalization gradient of the required type. This was achieved by using Gaussian patterns of activation over an ordered set of units representing the dimension to code for a given stimulus on that dimension. Training a discrimination involved associating the units activated by a stimulus with its outcome via the delta rule, and this in turn produced a gradient of generalization to a particular outcome across the dimension. The experiments reported in this paper were designed to further test this model of discrimination learning, and, in particular, to assess the validity of this generalization gradient approach to transfer between discriminations on a dimension. In order to do this, we moved from experiments based on transfer along a continuum to those based on the phenomenon of transposition.

Kohler (1918) was one of the first people to demonstrate transposition in animals. In this test, chickens were trained on a two stimulus discrimination task between a dark card (S-) and a light card (S+). In the test phase they had to choose between the original stimulus (S+) and a new stimulus, an even lighter card (S'), and it was found that they showed a preference for responding to the new, lighter stimulus (S').

The standard associative account of this phenomenon appeals to the notion of generalization, making it a promising preparation for our purposes. Spence's (1936) theory of discrimination learning in animals assumes that if a response to a particular stimulus is followed by reward, the excitatory value of that stimulus is strengthened. Equally, if a response to a stimulus is not followed by reward, the inhibitory value of that stimulus is increased. The two add in algebraic fashion, to result in a final net value of the stimulus, which governs the animal's response. Spence (1937) later argued that when animals are presented with a discrimination between two stimuli from the same dimension there will be some generalization between them. Therefore, the excitatory tendency to respond to S+ will also be elicited by S- but to a smaller degree, and likewise the inhibition associated with S- will also be associated with S+ but to a smaller degree. The level of responding to a given stimulus is then determined by the difference between the excitation and inhibition at that point along the dimension.

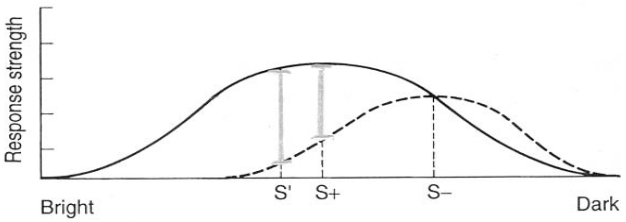

Figure 1: Excitatory and inhibitory gradients developed during training on S+ vs. S-.

This is well illustrated by Figure 1, which shows the gradients of excitation and inhibition developed around S+ and S- respectively. At the position of the novel stimu-

lus, S', the difference between the two generalization gradients is greater than for S+ therefore predicting greater responding to this novel stimulus and hence transposition.

A more modern, elemental explanation of transposition is illustrated in Figure 2. On this approach, a stimulus is represented by a set of activated elements or units, a distributed representation. Variation along a stimulus dimension such as brightness will, for the most part, be represented by different elements corresponding to different values on the dimension, rather than the activation level of a single element being the primary indicator of value on the dimension. Each element has a 'tuning curve' such that it responds most strongly to a certain value on that dimension and this response drops off fairly rapidly with distance from this value. Many elements will be active when any stimulus with value on that dimension is present, the coding is via a pattern of activation. Learning will proceed via association between the elements activated by a stimulus and other units representing reward.

The stimuli in the hard discrimination (S+ vs. S-) originally trained are close on the dimension such that there is a large degree of overlap between the patterns of activation that represent them, resulting in considerable generalization between them. The elements that are most active, and so dominate learning, are not those that best discriminate between the stimuli. If a new stimulus (S') is added on the dimension above S+, the most active units when this stimulus is presented are those that best discriminate between S+ and S-. Therefore, in a choice between S+ and S' the units representing S' give rise to a stronger net association with reinforcement, than the units representing S+, meaning that S' will be chosen over S+.

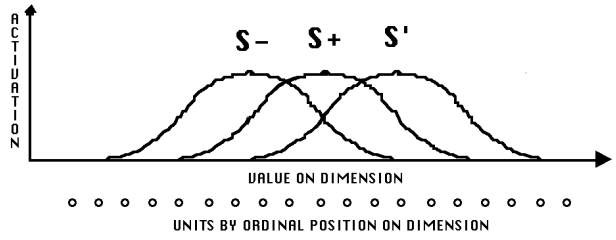

Figure 2: Stimulus representation on a dimension.

Experiment 1

Our first experiment tests the prediction made by both Spence and elemental theories that generalization on a dimension should lead to transposition. We investigated learning and performance on an artificial, morphed face dimension, by pre-training an initial discrimination, and then transferring to training on another discrimination in a manner either consistent or inconsistent with the effects of transposition outlined above. Addressing the issue of generalization in this way had the added advantage of allowing us to pit the generalization explanation of discrimination learning and transfer against an alternative position that stresses the importance of associating particular stimuli to particular responses. Generalization is then seen as a later computation based on these learned associations, and is thus a secondary consequence rather than the primary result of discrimination learning.

Stimuli and Apparatus

The stimuli used in all phases of the experiment were pictures of faces constructed from two similar black and white passport photographs. These images were morphed into one another and the experimental stimuli were taken from this dimension. Two original photographs are used as poles of the dimension, and then ten equal intervals are set to create the nine intermediate stimuli used for the experiments. The dimensional nature of the stimuli allows an investigation of how generalization and discrimination between the stimuli occurs under a variety of manipulations. An example of one of the dimensions is shown in Figure 3, but the experiment involved learning about four different face pair dimensions concurrently, with the assignment of the face dimensions to the conditions of the experiment counterbalanced appropriately. In this, and subsequent experiments the faces at positions 3 and 9 on a dimension constitute an easy discrimination whilst those at 5 and 7 a hard discrimination. These stimuli were presented on an Apple Macintosh computer running Real Basic. They were 3.5 cm by 4.5 cm and subjects sat approximately 50 cm from the screen.

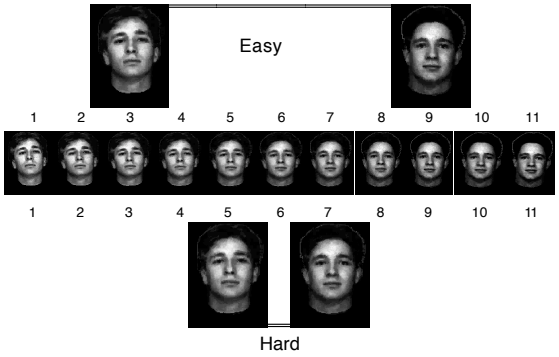

Figure 3: One of the four morphed face dimensions used in these experiments.

Subjects and Design

Subjects were 32 undergraduates and graduates from the University of Cambridge with an age range of 18 – 30. All subjects received two blocks of pre-training on the hard discrimination between stimulus 5 and stimulus 7, denoted 5+, 7-. The + and – simply show different response assignment, which in this case were either the left or right response keys. Responses were counterbalanced so that a given picture on a given dimension had an equal

chance of being assigned to the left or right key. Pre-training ran concurrently on all four dimensions for a fixed number of trials (forty trials per block, five for each of the eight faces). They then received another training phase (another forty trials in total, five for each face) in which there were four conditions. There were two transposition conditions, one closer along the dimension to one pole (7+, 9-) and one closer to the opposite pole (3+, 5-), and then the equivalent reverse transposition conditions where the key assignments were the opposite way round. This within subjects design meant that subjects experienced every condition, one on a given dimension. The dimension used for each condition varied between subjects. This was followed by a final test phase in which performance following training in each condition was assessed without giving the feedback used in pre-training and training. In this phase each of the eleven faces from the four dimensions is shown 5 times. The data of interest are the responses to the stimuli in this final test phase, in particular the differences in responses between 3 and 5, and 7 and 9 following training in either the transposition or reverse transposition conditions. A significant difference between responses following training on these two conditions would provide evidence for the generalization mechanisms underlying this type of learning. Note that the design means that in the Transposition condition subjects will experience a change in response for the stimuli pre-trained, whereas in the Reverse Transposition condition these stimulus-response assignments are unchanged. Thus a theory that emphasizes acquisition of individual stimulus-response associations will predict that learning will be stronger in the Reverse Transposition condition, in direct opposition to generalization theory which predicts that it is the generalization gradient that matters, and that this will favor the Transposition condition as the gradient from pre-training to training is unchanged.

Procedure

In both the pre-training and training phases of the experiment, subjects were told that once they pressed the 'G' key, a constant stream of stimuli in the form of faces would appear on the screen. They were told that their task was to sort these stimuli into two categories from the outset. They were to do this by pressing one of two keys ('x" on the left or '.' on the right, the correct key was counterbalanced between subjects) and would receive immediate feedback as to the correctness of the response. If they did not respond within 4 seconds they would be timed out. The subjects were told that the faces were randomly and equally allocated to either left or right key and that their task was to simply find out and remember which ones were 'right' and which 'left'. Stimuli were presented singly in a continuous stream. Each trial started with a '+' fixation point for 1.5 sec, which was then replaced by the face stimuli for a maximum of 4 sec and disappeared once a response or time-out was made. Feedback was then given for 1.5 sec, either 'correct' displayed in the center of the screen or 'wrong' and a beep if an invalid key was pressed. After completing the pre-training and training phases, subjects progressed to the test phase of the experiment. Subjects were told to categorize the stimuli into two categories based on the judgements they had made in the last training phase, but this time no feedback was given. Feedback was replaced by a 1.5 sec pause between the subject's response and the succeeding stimulus.

Results

The results of Experiment 1 are shown in Figure 4. One key, e.g. the left key, is designated the negative category (a press scores –0.5 for that stimulus) and the other, right key, the positive category (scores +0.5) during test. Key assignments were counterbalanced across subjects so that the positive category has equal numbers of left and right responses. The mean key score indicates the average of the key presses across subjects, ranging from –0.5 to +0.5, and would be zero if subjects were indifferent to which stimulus went with a given key.

A three-way ANOVA was carried out on the results of the test phase with three within subjects factors. These were Discrimination (transposition or reverse transposition), Direction (up or down the dimension) and Stimulus (number on the dimension). This gave an $F(10,310) = 1.924$, $p<0.05$ for the main effect of Stimulus, but no other significant effects, in particular it gave a non-significant effect of Discrimination, $F(1,31) = 1.084$, $p>0.05$. The first effect refers to the fact that the stimulus number influences the mean key score, but the lack of significant main effect of Discrimination suggests that there is no significant difference in mean key score between transposition and reverse transposition conditions.

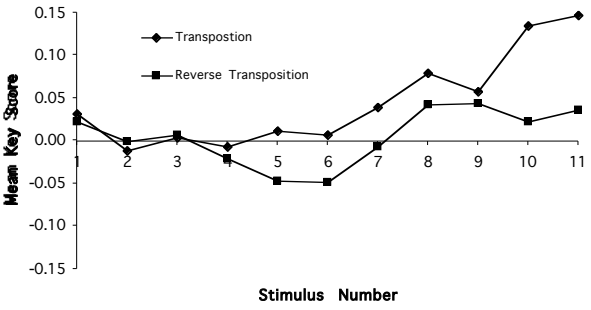

Figure 4: The results of Experiment 1 expressed as mean key score at each point on the dimension.

As a further analysis, regression lines were fitted to the mean performance scores for each stimulus in the two conditions. In each the mean key score was the dependent variable and the stimulus number the independent variable. Single sample t-tests comparing the regression line gradi-

ents to zero yielded a significant result, slope=.013, $t(10) = 3.25$, $p<0.01$, in the Transposition condition, and a non-significant result, slope=.002 $t(10) = 2$, $p>0.05$ in the Reverse Transposition condition. An independent t-test between the two regression line gradients also yielded a significant result, $t(20) = 2.668$, $p<0.02$. From this we can infer that there was a difference in generalization gradients across the dimension in the two conditions. Thus, despite an initial lack of evidence for any difference between the two conditions it does appear that the training discrimination phase has had an effect, resulting in a significant generalization gradient in the Transposition condition but not in the Reverse Transposition condition.

The difference scores for the individual discriminations 3 vs. 5 and 7 vs. 9 trained in the last training phase were calculated by subtracting the mean key score for stimulus 3 from the mean key score for stimulus 5 (5-3), and similarly, the mean key score for stimulus 7 from that of stimulus 9 (9-7). When these difference scores were combined and subjected to analysis it was found that the overall difference scores were not significantly different from zero in either the Transposition or Reverse Transposition condition (Fs<1), suggesting that the discriminations had not been learned to any great extent, though the overall trend was for larger (positive) difference scores in the Transposition condition.

Discussion

The significant difference found between the generalization gradients in the transposition and reverse transposition conditions can be adequately accounted for by Spence's (1937) theory as well as our own. In pre-training, a generalization gradient develops across the dimension by virtue of the association of stimulus representations to their respective keys. In the case of transfer to the transposition discrimination, the shift in stimuli is congruent with the already established generalization gradient, as the differential key response for the new discrimination (i.e. which of the stimuli one should have the greatest tendency to press the left key for) is exactly that already established by this gradient. The reverse transposition condition, on the other hand, goes against the generalization gradient formed in pre-training and requires it to be unlearned.

The observed difference in generalization gradients suggests that there is learning in the transposition condition whilst nothing (overall) is learned in the reverse transposition condition. We would expect, therefore, the individual discriminations to be learnt in the transposition condition but not in the reverse transposition condition. In Experiment 1, however, the discriminations are learnt in neither condition. This is possibly due to a lack of power in the statistical tests used due to the small population size, or simply to the fact that the within subjects design was sufficiently hard and learning sufficiently slow to yield no significant acquisition of the discriminations.

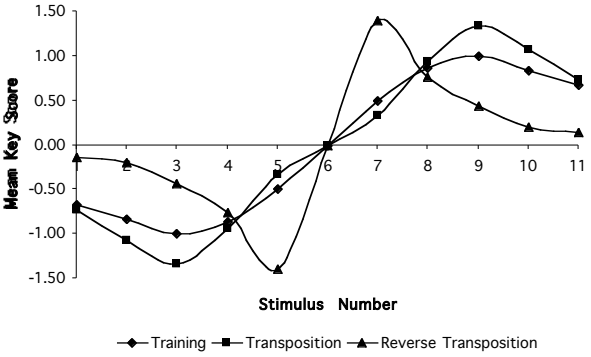

Figure 5: Expected generalization curves following pre-training and training.

Closer examination of how the generalization curves across the dimensions might be affected by training under each condition reveals that traditional accounts of generalization may well be sufficient to explain these results. Figure 5 shows in idealized form the generalization curves that should develop following pre-training on the 5 vs. 7 discrimination and then how this is expected to change following training in the transposition and reverse transposition conditions. In the transposition condition the shape of the generalization curve is maintained, and actually altered very little. In the reverse transposition condition the associative strengths of stimuli 5 and 7 to the respective pre-trained keys are increased, whilst those of 3 and 9 are decreased, resulting in a considerably altered generalization curve.

From Figure 5 we can see how a regression analysis would suggest that there is a lack of a significant generalization gradient following training in the reverse transposition condition, whilst a significant generalization gradient is maintained in the transposition condition. If we consider the discrimination scores expected on the basis of Figure 5 we see that in the transposition condition the difference scores (5 - 3 and 9 - 7) should both be positive, and so their overall summed score (i.e. (5 – 3) + (9 - 7)) should be positive. In the reverse transposition condition both difference scores should be negative and the same corollary applies. We have already seen that the individual discriminations were not learnt in this experiment. This lack of significance for the trained discriminations is perhaps not surprising, given that the subjects in this experiment reported that the within subjects design we had used made learning them a very hard task. We suspect that the reason why we are able to pick up an effect of Condition using the regression analysis is that this allows us to make use of all the data across the dimension, increasing power.

Figure 6 shows the results of a simulation of Experiment 1 using the model reported in Suret and McLaren (2002). The simulation shows a good fit to the trends in the data, in that the slope of a regression line is predicted to be greater for the Transposition condition, as are the difference scores for the discriminations. Note that the predicted discrimination differences in the Reverse Transposition condition are close to zero.

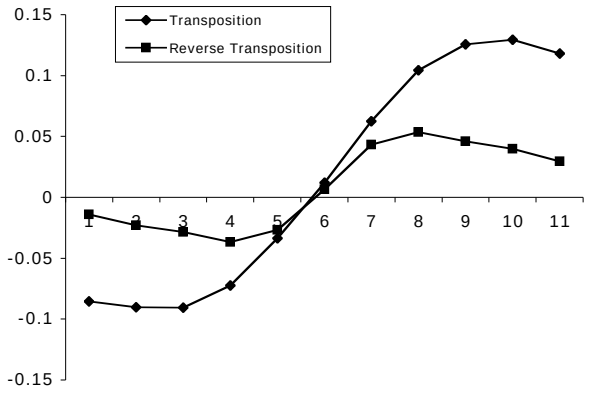

Figure 6: Simulation of Experiment 1.

Experiment 2

Experiment 1 produced results that are consistent with a generalization model of discrimination learning and transfer, albeit the evidence for transposition *per se* was not as strong as one might wish. There is, however, another possible explanation of the results which would jeopardize this conclusion. It may be that subjects are especially sensitive to a change in the response required to a stimulus as they move from pre-training to training phases of the experiment. If this were so, then the stronger learning in the Transposition condition of Experiment 1 would follow as a consequence of the change in response required to stimuli 5 and 7, rather than because of the congruent generalization gradients established in pre-training and training. Experiment 2 addresses this issue by arranging for a change in response assignments to be accompanied by incongruent generalization gradients between pre-training and training phases of the experiment.

In this experiment the stimuli used were the same as in Experiment 1, and 10 new subjects were taken from the same population. They were pre-trained on an easy discrimination (3+, 9-) for all four dimensions for an equal number of trials, (10 per dimension, 40 in total). After this pre-training phase they were then trained on a reverse transposition discrimination (40 trials in total, 5 for each face) either on one side (7-, 9+) on two dimensions, or on the other (3-, 5+) for the remaining two dimensions. The discriminations assigned varied between subjects so that no dimension was trained more on one side than the other.

This was then followed by a test phase as in Experiment 1. Significant differences in responding between the pairs trained in the final training phase (3, 5 and 7, 9) would demonstrate acquisition of the discrimination and based on the generalization model we would expect this not to occur here. This is because the easy pre-training will produce an opposing gradient, and we might expect more learning of the initial easy discrimination than the harder training discriminations. If the change from 3+ and 9- to 3- and 9+ is the important factor then we might expect the discriminations to be learned.

Results

The results of Experiment 2 are shown in Figure 7. Once again the difference score for each discrimination was computed as in Experiment 1. Analysis of the discriminations across subjects revealed that all discriminations were non-significant ($p>0.1$), and showed a trend opposite to that required to learn the training discriminations, though consistent with the gradient that we might expect to be established in pre-training.

There is little doubt, however, that the subjects have learned something. The results for each stimulus shown in Figure 7 demonstrate acquisition of a generalization gradient across the dimension that is the one expected as a result of pre-training on the easy discrimination,.

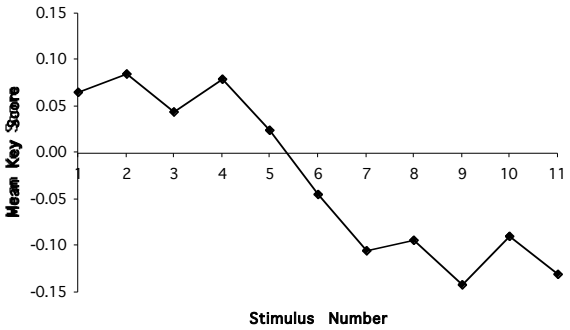

Figure 7: Mean responses by face in Experiment 2.

This impression is confirmed by analysis of the best fitting line to the data points shown in Figure 9. The slope is -.0246, which is significantly different to zero, $t(10) = 2.35$, $p<.05$.

Figure 8 shows simulation results for Experiment 2 which fit well with the trend in the data.

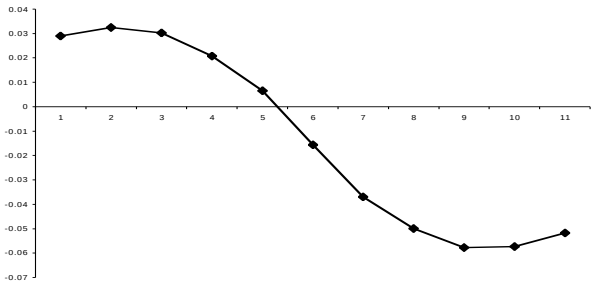

Figure 8: Simulation of Experiment 2.

It would be wrong, however, to assume that the non-significant results for the trained discriminations implies that that training had no effect, and that the results are entirely driven by the pre-training of the easy discrimination. The greater power available to us in this experiment allows a split into the two individual discriminations used in training, i.e. 3- vs. 5+ and 7- vs. 9+. Figure 9 shows the results for these two discriminations.

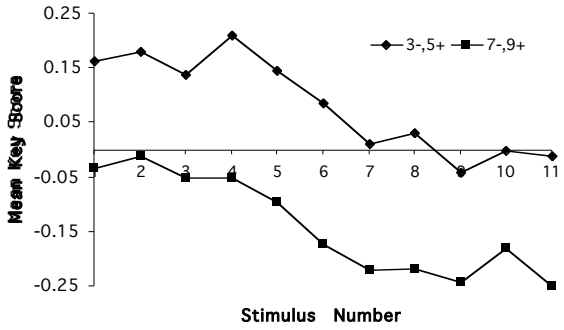

Figure 9: Results for Experiment 2 as a function of trained discrimination.

A comparison of the two discriminations gives a highly significant effect, $F(1,10)=32.9$, $p<.01$. Once again this pattern of results is captured to at least some extent by our simulation of Experiment 2, as shown in Figure 10.

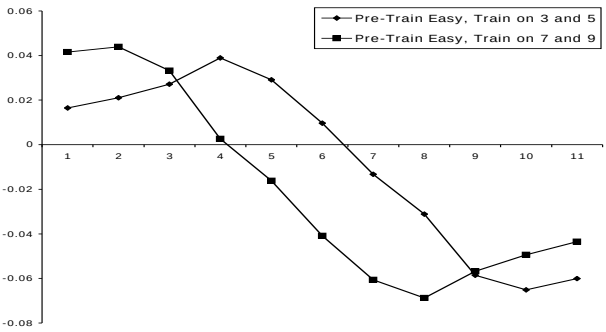

Figure 10: Simulation results for Experiment 2 as a function of trained discrimination.

Discussion

The results of this experiment suggest that a generalization based account of the data is to be preferred. There is little evidence of the subjects having acquired the trained discriminations as the 'change' hypothesis would predict, even though we know that the training had a significant effect on performance. Instead, the generalization model can account for both the gradient established across the dimension and the differential effect of the trained discriminations on this gradient.

General Discussion and Conclusion

The purpose of this paper was to test the generalization model of discrimination learning developed by Suret and McLaren (2002). In general this model gave a good fit to the data, whereas alternative hypotheses pertaining to the specific learning of stimulus-response associations and an enhanced ability to detect changes in stimulus-response assignments receive little support. If subjects learn individual stimulus-response associations (e.g. 5 with left(-)) as a preliminary to some further computation that leads to responding and generalization, then we might expect the Reverse Transposition condition of Experiment 1 to favour this. In this condition the 5- and 7+ mappings are in force for both pre-training and training. This should favour their acquisition, which should aid learning of the discriminations. No evidence in support of this prediction was found in Experiment 1. Appealing to the change in stimulus–response mapping from pre-training to training as another factor that might confound the effect expected in Experiment 1 was not supported by the data from Experiment 2. We conclude that a model of discrimination learning that has generalization built in as a primitive computation is required to explain these data.

References

Kohler, W. (1918). Simple structural functions in chimpanzees and chicken. In W.D.Ellis (1969), *A source book for Gestalt psychology*. London: Routledge & Keegan Paul.

Lawrence, D.H. (1952). The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology, 45*, 511-516.

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

McLaren, I.P.L. and Mackintosh, N.J. (2002). Associative Learning and Elemental Representation: II. Generalization and Discrimination, *Animal Learning and Behaviour 30*, 177-200

McLaren, I.P.L., and Suret, M. (2000). Transfer Along a Continuum: Differentiation or Association? *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society, NJ*, LEA, 340-345

Spence, K.W. (1936). The nature of discrimination learning in animals. *Psychological Review,* 43, 427-449.

Spence, K.W. (1937). The differential response to stimuli varying within a single dimension. *Psychological Review,* 44, 430-444.

Suret, Mark & McLaren, I.P.L. (2002). An Associative Model of Human Learning on an Artificial Dimension. *Proceedings of the 2002 International Joint Conference on Neural Networks, Piscataway: IEEE.* 806-811