Modelling Dual-Processes in a Connectionist Network

Toby Johnson (T.Johnson@exeter.ac.uk)
IPL McLaren (i.p.l@mclaren@exeter.ac.uk)

Psychology, University of Exeter, Exeter, UK, EX4 4QG

Abstract

This paper presents a connectionist network simulation of Livesey and McLaren's (2009) results. In that paper they showed that participants with post-discrimination gradients that were initially peak shifted became monotonic as they were exposed to the full range of test stimuli. While the authors suggest that this is the result of rule-based processes 'taking over' responding, we show how a connectionist network with an attentional parameter and realistic activation functions for the input can simulate both the peak shifted and monotonic gradients. Although we do not infer that the monotonic gradient obtained in peak shift paradigms is entirely the result of associative, rather than propositional processes, we suggest that perhaps it is a change in the allocation of attention, in conjunction with the underlying representational structures used for the stimuli that facilitates rule induction in this case.

Keywords: Associative Learning; Connectionist Network; Peak Shift; Dual-Process; Single Process

Introduction

Dual-Process theorists assume that human associative learning relies upon two sets of processes (see McLaren et al., 2014; 2019). One set, aptly named 'associative processes', are those also found in non-human animals, and were the initial focus of learning theory. This resulted in the algorithms developed by Rescorla and Wagner (1972) and Mackintosh (1975), which provided the foundations for the development of more sophisticated models of associative learning that now describe both the representation of stimuli, as well as the learning algorithms that govern association formation (e.g., McLaren & Mackintosh, 2002). The other set of processes are more complex, higher-level cognitive processes, that are unique to humans. These 'propositional processes' are conscious and result in articulable rules about the relationship between stimuli. Another approach, espoused by single-process theorists (see Mitchell, de Houwer, & Lovibond, 2009) assumes that human associative learning can be explained entirely via propositional processes. But what both approaches agree on is the need for rule-based processes to explain at least some learning and behavior in humans.

Connectionist theorists, however, have a long history of analyzing psychological domains thought to require multiple processes or rules and showing that in fact they can be modelled in terms of networks using just one set of basic processes in a quite simple architecture. An excellent example of this approach can be found in the Seidenberg and McClelland (1989) model of reading aloud developed in the last century, which was able to produce effects that, up until then, had been imputed to a dual process account of reading using separate lexically-based and grapheme-phoneme conversion routes. In fact, this use of one network to replace a more complex set of routes/processes goes all the way back to theorists such as Spence (1937) who showed how phenomena such as transposition which seemed to imply relational learning would, in fact, emerge from a very simple associative model. The example that we will consider here is one that Spence also dealt with at the time, peak shift, but now we will show how this type of approach can be extended to also explain the monotonic gradients typically obtained when people have become aware of the rule governing responding.

Peak Shift

The peak shift effect (e.g., Hanson, 1957; Livesey & McLaren, 2009) has provided what has been taken to be evidence for the existence of associative processes in humans. This effect occurs after discrimination training with two similar stimuli that vary along a dimension (e.g., colored rectangles, varying from green to blue). When tested with several stimuli that vary more extremely along the same dimension, participants who have been unable to induce a rule during training, show a peak shift. That is, they are most accurate not to the training stimuli, but to stimuli located slightly farther along the dimension, away from the opposite training stimulus. Accuracy then reduces as the testing stimulus moves even farther along the dimension. For participants who have induced a rule to aid with discrimination, such as 'if the stimulus is green then it belongs to category A, if it is blue then it belongs to category B', then participants peak accuracy will typically be to the bluest and the greenest stimuli they are presented with. As such, these participants show monotonically increasing accuracy as the stimuli moves farther along the dimension.

Quite apart from this dissociation in the pattern of responding between people who have induced a rule and those who have not, peak shift is reliably seen in pigeons (e.g., Hanson, 1957), and there is evidence that learning in pigeons is the result of associative, rather than propositional processes (e.g., Meier, Lea, & McLaren, 2016). Furthermore, it can be easily modelled by associative networks that assume elemental representations of stimuli (e.g., McLaren & Mackintosh, 2002).

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Livesey & McLaren (2009)

Livesey and McLaren (2009) conducted an extended peak shift experiment in which participants were trained to categorize, with feedback, two similar stimuli that varied only in their hue. At test, they were required to categorize a wider range of stimuli (12 in Experiment One and 6 in Experiment Two), varying more along the hue dimension. The authors reported that, on average, responding across the test phase was not peak shifted. There was a significant rise in accuracy, but not a significant fall. However, they showed that participants post-discrimination gradients changed from peak shifted at the beginning of the test phase, to monotonically increasing at the end. They suggested that some associative strength had been accrued during the training phase, but not enough to allow for rule induction in some participants. Only when participants were shown the full range of stimuli during test were they able to articulate the difference between the stimuli (that they varied in hue). This explained why responding changed from associatively based, resulting in peak shift at the start, to rule-based, resulting in the monotonic gradient by the end. Moreover, in experiment 2, they reported that this change in the post-discrimination gradients only occurred in those who were unable to induce a rule during training, or who induced the wrong rule (see Figure 1). For those who noticed the difference in hue during training, responding followed the monotonic gradient throughout the entirety of the test phase.

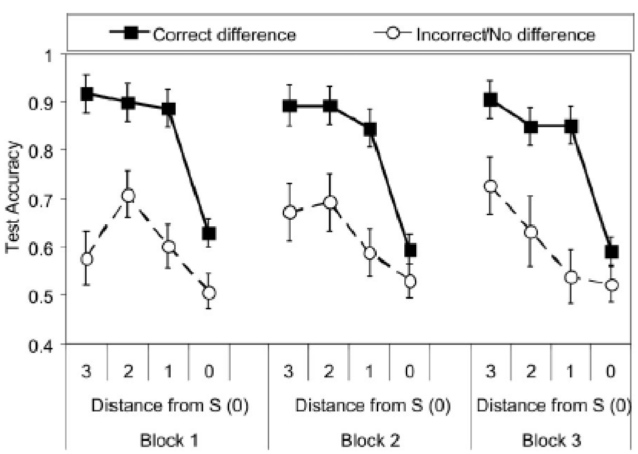

Figure 1: Graph from Livesey & McLaren (2009) experiment 2, showing how the post-discrimination gradient changed for those who reported noticing no difference (hollow circles) between the training stimuli from peak shifted at the beginning of testing to monotonic at the end.

Historically, not a great deal of further analysis has been seen as necessary for the monotonic gradient. If participants have induced a rule such as ‘if blue then respond A, if green then respond B’, then their behavior has been taken as accounted for. But what this fails to explain is how the rule itself has been induced, and indeed how it controls responding. During testing, participants received no feedback on their responses, yet responding to stimuli at the extreme ends of the dimension became more accurate. This suggests that whatever associative strength had accrued as a result of training is influencing rule induction. If these two processes were completely separate, then the hypothesis testing involved in rule induction would not yield more accurate responding during testing, as there is no feedback.

One way in which associative processes may be influencing rule induction is through participants observing their own responses. Indeed, Livesey and McLaren (2009) offered the same argument. If participants become aware that they are pressing a particular response key to a stimulus, then they may begin trying to understand why. It would then not be long until they become aware of the subtle differences between the stimuli (i.e., hue) and focus their attention on that dimension. Our thought was that this change in attention can be modelled using the same associative network that can simulate peak shift. Previous networks, such as Kruschke’s (1992) ALCOVE (itself using an algorithm that owes much to Mackintosh, 1975), have suggested that attentional changes can affect learning and performance, and have incorporated this in their model.

Here we will present a new model that can simulate the pattern of results found in Livesey & McLaren (2009). We have not attempted to fit the model to the data, and as such the overall accuracies are not identical to those reported by the authors. However, we will show how an associative network can simulate peak shift and then generate the monotonic gradient, through a simple mechanism that acts to both increase the gain on the input, and in doing so varies the representational structure of the stimuli under test. We theorize that this captures the shift in attention and consequent change in the underlying representational structure with which stimuli are encoded.

The Model

Summary of the model

The model we describe in the present paper is a feedforward-backpropagation connectionist network with a winner-take-all (WTA) system for determining responses (Rumelhart, Hinton, & Williams, 1986; Wills & McLaren, 1997). Both the input and hidden layers comprises 10 units. The output layer and WTA system comprises two units. All units in the input layer are connected to all those in the hidden layer, and all those in the hidden layer are connected to each unit in the output layer. These two units reflect the two different key presses that could be used to classify stimuli, and were each connected to their respective WTA units, which determined the final output for that particular trial. Connections between the input and hidden, as well as the hidden and output layers were modified using the delta rule and backpropagation, allowing the network to learn. Our model assumes elemental representation of stimuli, and thus for any given stimulus, several input units will be activated. Figure 2 shows the basic architecture of the model. The details and unique formal equations of the model will be described below.

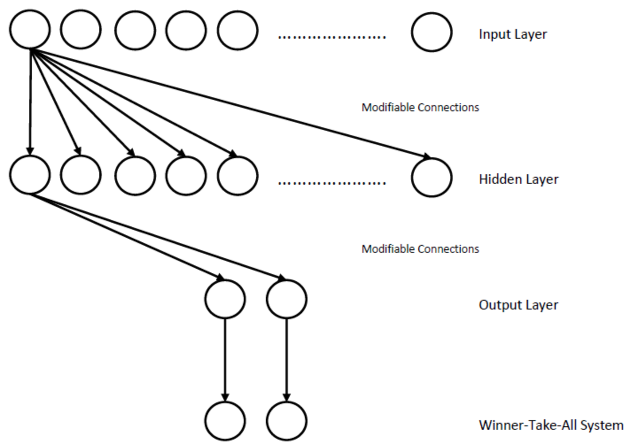

Figure 2: The basic architecture of the connectionist network outlined in this paper.

Stimulus representation & attentional mechanism

Stimuli were represented as a pattern of activation across the input units that followed a Gaussian distribution. The 10 input units corresponded to the dimension (i.e., hue), and for any given stimulus, several would be activated. Thus, the two training stimuli activated overlapping sets of units, however the magnitude of activation differed such that the peak activation for one stimulus may have been at input unit 4, but for the other at unit 5. In reality, for any given stimulus, peak activation may not necessarily have been located perfectly on an input unit, but rather could have been located 'between' units. Most importantly, stimuli located closer to each other on the dimension activated many more overlapping units compared to stimuli located farther apart, but the pattern of activation for each stimulus across the 10 input units was unique.

$$I_i = ve^{-k\left(i-\frac{hue}{l}\right)^2} \tag{1}$$

Equation 1 describes the Gaussian function that produces the initial input for a given input unit, I_i. Here v is a constant that determines the maximum amplitude of the gradient. The value of this constant has very important implications with regards to Equation 2 (below), as it essentially acts as an attentional parameter in two different ways. First it controls the strength of the input, with higher numbers resulting in higher peak amplitudes. Secondly, it interacts with Equation 2, to alter the shape of the Gaussian function, irrespective of k. This was the only parameter that was changed to produce our results. k determines the broadness of the gradient, with lower numbers resulting in broader gradients, hue refers to the hue of the stimulus, and l is a constant that determines how the changes in hue affect the overall position of the gradient across the input units. A low value will result in similar stimuli activating less overlapping units, whereas a high value will result in similar stimuli activating more overlapping units. This input is then passed through an activation function to produce the final activation, A_i for any given input unit.

$$A_i = p\left(\frac{I_i}{(I_i + t)}\right) \tag{2}$$

Equation 2 is derived from the initial input to a unit, I_i. That is, each unit's final activation is calculated from its own initial input via this function. Here, p is a constant that determines the final maximum activation of the unit and t is a constant that affects the shape of the final activation gradient. Equation 2 interacts with Equation 1 such that units with a low input are influenced much more greatly compared to units with already high inputs. Differences across units with an initial input lower than t will remain more similar in their magnitude after calculation of the final activation compared to differences across units with an initial input higher than t, which will be limited by p.

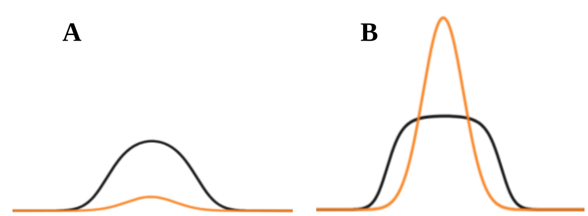

Figure 3: Diagrams showing the Gaussian pattern of activation after the initial input function (equation 1: orange line) and the final activation function (equation 2: black line). Figure 3A shows the difference between initial input and final activation for a set of units with a low initial input. Figure 3B shows the difference between initial and final activation for a set of units with a high input, arising from attending more closely on the stimulus (and in particular to the dimension that is deemed relevant). The constant v determines the peak amplitude for the initial input.

Learning & output

Both the hidden and output layer were activated using the logistic activation function, whereby 0.5 represents a resting level, 0.1 represents maximum inhibition and 0.9 represents maximum excitation. The strength of the connections between the input and hidden, and hidden and output layers were modified using the delta rule and backpropagation (Rumelhart, Hinton, & Williams, 1986). For the two output units, representing different key presses, target activation was set to 0.9 and 0.5, respectively, representing one key press, or 0.5 and 0.9, respectively, representing the other key press. The WTA system (Wills & McLaren, 1997) is comprised of two units, connected to their respective output units, that are self-excitatory and mutually inhibitory and used a similar activation function to Equation 2. Each WTA unit takes the activation from its respective output unit and competes with the other WTA unit until one unit's activation reaches a criterion and is labelled a winner. This system has the benefit of labelling one output, or key press, as a winner on each trial, making it clear what the response on that trial is. In this way we are able to obtain data similar to those generated in experiments with human participants.

Simulations

We simulated 100 experiments, with each experiment consisting of 100 simulated participants. All constants, except the main attentional constant, v, remained the same throughout the simulation and for all participants. For Equation 1, the constant determining the shape of the Gaussian gradient for the initial activation was k = 0.8, and the constant determining how changes in hue affect the location of the Gaussian gradient was l = 25.5. For equation 2, the constant determining peak final activation was p = 2, and the constant determining the shape of the Gaussian gradient for the final activation was t = 1. v, the attentional constant was v = 0.2 during training and the first stage of testing and was changed to v = 12, for the second stage of testing.

For each simulated participant, the weights between the input and hidden, and hidden and output layers were initially randomized. Each simulated participant was trained with 10,000 iterations of each of the two training stimuli (using v = 0.2). They were then tested twice each on stimuli that varied more greatly across the dimension. This comprised of the two training stimuli, T, stimuli on each side of the dimension which lay further along the dimension than T, 'Near' (N), and stimuli at either end of the dimension, 'Distant' (D). This testing stage occurred twice, with different constants for v in each (0.2 for the first test and 12 for the second, although due to the various parameters in our model there are likely to be many more). We have not undertaken any formal model fitting, and thus the values we have used for the parameters are not necessarily the only ones, nor the best, in replicating the exact pattern of results. But they will serve to illustrate the principles involved.

For each testing stage, analysis is separate. For each simulated participant, the accuracy for each test stimulus was calculated. A final average accuracy for each stimulus was then calculated across the 100 simulated participants for that particular simulated experiment. This was done for all 100 simulations. As per Livesey and McLaren (2009), we then collapsed across either side of the dimension, around the training stimulus to produce the average accuracy for each test stimulus as a function of its distance away from the training stimulus. Thus, as described above, we have the average for test stimuli at positions, T, N, and D; and although McLaren and Livesey had 3 positions other than training, we reiterate that we are not trying to directly replicate their results, but simply to show that the gradients of responding can be simulated entirely using an associative network, by changing an attentional parameter. Similarly, previous research on peak shift and post-discrimination gradients has collapsed test stimuli into three position, often labelled as Training, Near, and Distant/Far and has successfully reported peak shifted gradients (Wills & Mackintosh, 1998, Jones & McLaren, 1999).

Results & Discussion

As the simulation comprises of, essentially, 10,000 participants, all differences, however small, are significant. However, again we direct the reader to the point above concerning the post-discrimination gradients and patterns of responding. Figure 4 shows the results of the simulation with both values for the constant v. The value of 0.2 was set during the training and the first testing stage, resulting in peak shift, and the value of 12 was set during the second testing stage, resulting in the monotonic gradient.

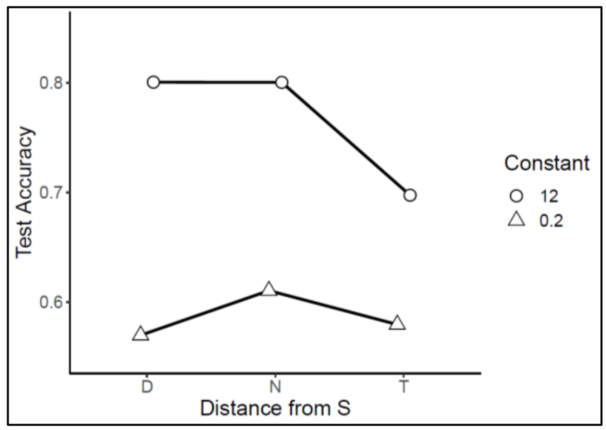

Figure 4: Graph showing the simulation results using the model outlined in this paper. All simulated participants were trained with the constant v = 0.2. The hollow triangles show the results for testing using the constant v = 0.2, and the hollow circles show the results for testing using v = 12. Error bars indicating standard error would normally be included but we have omitted them as they are so small.

What is important here is that when our constant is set to 0.2 (remembering that all of the simulated participants were trained with this constant also), the simulation produces a distinctly peak shifted gradient, similar to that produced by the 'Incorrect/No Difference' participants, in Block one of Livesey and McLaren's (2009) test phase. However, when our attentional input parameter is set to 12, during the final stage of testing, the simulation produces a monotonic gradient, similar to that produced by the participants in Block 3 of Livesey and McLaren's test phase. Admittedly, our simulation does not produce the linearly increasing function that Livesey and McLaren found, but it is not uncommon to find only a small numerical, and statistically insignificant, increase between extreme stimuli (see Jones & McLaren, 1999). Moreover, In Livesey and McLaren, we can see how participants overall accuracy does not increase when their gradient of responding becomes more monotonic, at least not to the same extent as in our simulation. However, this is perhaps to be expected as testing can be construed as a period of extinction, and our model does not implement this aspect of the procedures used.

Livesey and McLaren (2009) suggested that for participants who reported an incorrect rule or did not notice a difference between the training stimuli, responding at test was initially governed by the associative strength that had accrued due to more basic associative processes. There is good evidence to suggest that peak shift is the result of these processes, and it is commonly modelled using elemental representation, as we have done so here (see McLaren & Mackintosh, 2002 for a more detailed account, but see also Lee, Hayes, & Lovibond, 2018 for an alternative account). The authors go on to state that the change in post-discrimination gradients between the first and final block of testing, becoming monotonic, is the result of these participants being exposed to the full range of stimuli, which vary more greatly in hue, facilitating rule induction as this becomes an easily perceivable difference in the stimuli.

However, whilst it is straightforward to suggest that the monotonic gradient is due to propositional, rule-based processes, this account does not necessarily answer the question of how these processes begin to take over, or, and perhaps more importantly, how they control responding.

As mentioned in the beginning, if participants begin hypothesis testing during the testing stage, a period in which there is no feedback, then, in the absence of any associatively generated expectations, there is no reason that accuracy should improve over trials, yet it does. Our suggestion is that they are receiving information from something that demonstrates to them that one hypothesis, or rule (i.e., that blue belongs to category A and green belongs to category B) is more likely to yield the correct answer, compared to another (i.e., that blue belongs to category B and green belongs to category A) and that it is the associative system that provides this information. Whilst participants are consciously unsure of the particular category with which the stimulus they are presented with belongs to, they may begin to notice a pattern in their own responding. When they are presented with a particular stimulus they tend to want to press the key corresponding to category 'A'. This would prompt participants to attempt to understand why they have been inclined to press that button and they would focus their attention more closely on properties of the stimulus. At the same time, the sheer range of variation in the stimuli that they are now attending to would lead them to become aware of what dimensional variation was occurring. Importantly, this happens quickly during testing, because there is a greater range of stimuli available and once participants become aware of their own responses they are likely to soon be presented with stimuli that vary obviously on a dimension. This is also in keeping with the results of Livesey and McLaren who showed that post-discrimination gradients move from peak shifted to monotonic within three blocks of testing. Overall, we suggest that it is the associative strength accrued by associative processes that serves as the basis for the formation of rules and the switch to propositionally-based responding. This conclusion fits neatly with existing Dual-Process accounts (e.g., McLaren et al., 2019).

Now we turn to the important theoretical underpinnings of the transition from associative to cognitive processes we have just outlined. How do propositional processes come to actually control responding? This is important because, whilst a rule allows a participant to describe the differences between the stimuli, it doesn't explain the process of classification according to that rule. For example, take the rule, 'blue belongs to category A and green belongs to category B'. What denotes blue and what denotes green? Granted, this may be obvious for stimuli lying at extremes of the dimension, but for those that are not, at what point does green become blue? There must be a category boundary at which point the participant switches from responding "category A" to "category B", yet answers to post-experimental questionnaires rarely detail any such mechanism, leaving a great deal to still be explained.

If we return to the participant who has begun to notice that they are making consistent responses to certain stimuli, and who is now attending more closely to the stimuli they are being presented with, we can see where the input parameters of our network may have an effect. In focusing more closely on the stimulus, and in particular on the dimension that they deem relevant, participants encode more strongly the information about those features in that stimulus, and it is this that corresponds to the change in the constant, v from 0.2 to 12, that we have made in the network simulations, causing the underlying representational structure to change slightly, such that the final Gaussian gradient with which it is represented becomes broader. This in turn then also allows for generalization to a greater range of stimuli in our connectionist network, which results in stimuli at the more extreme values of the dimension being responded to more accurately. It is the control of that increase in attention to the appropriate dimension that is triggered by noticing that there is variation on that dimension, and that this variation seems to be related to responding on the part of the participant. And this, in itself, leads to the new, monotonic pattern of responding which has been reported as evidence of rule-based responding (and which undoubtedly co-varies with ability to report the appropriate rule; Jones & McLaren, 1999; Livesey & McLaren, 2009).

There is an even more extreme version of this hypothesis which makes the change in attention for the relevant dimension automatic rather than controlled in nature. This would posit something like the algorithm used in Mackintosh (1975, see also Sutherland and Mackintosh, 1971) or Kruschke (1992) acting to automatically increase attention to the relevant dimension and thus facilitate the transition from a peak shift pattern of results to a monotonic one. This type of explanation would undoubtedly work when comparing participants who did not get the rule during training to those who did in the first test block, as long as we argue that ability to articulate the rule is now epiphenomenal, and simply enabled by the high level of performance achieved once attention to the relevant dimension is high enough. But this explanation has difficulty with the changing pattern of results across test blocks for the "no rule" participants. In the absence of feedback, the algorithms in question would actually tend to reduce attention to the relevant dimension if this were to be interpreted as an extinction phase, and certainly wouldn't act to increase it further. Given this, the transition to a monotonic pattern of responding would not be generated, and so we would once again have to posit some element of control on the part of these participants. Nevertheless, we can see the possibility of a mechanism of this kind facilitating learning and eventual rule induction.

It could be argued that the inclusion of equation 2 in our model is not necessary, as it simply serves to broaden the generalization gradient. We could simulate a similar effect by foregoing equation 2 and instead only changing parameter k of equation 1, however we argue against this for several reasons. First, this would create similar questions to those we have attempted to answer in this article. What causes the broadening of this stimulus representation and does this occur before or after perceiving the dimensional variation of the stimulus? If we suggest that the broadening is the result of noticing the dimensional variation, then the question is how is this done? How is control implemented to do this? Instead, by including equation 2, we can allow the initial activation resulting from equation 1 to not change in its characteristics, but merely to increase in amplitude, which we believe better reflects the focusing of attention toward the dimensional features of that stimulus, which in turn broadens the final gradient of activation and allows for more encoding of them.

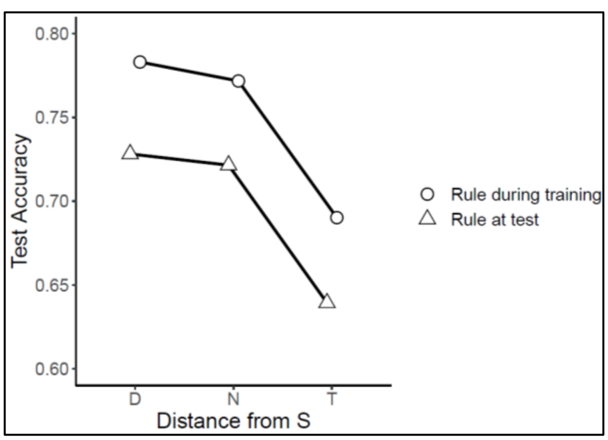

Figure 5: Graph simulating the participants who induced the correct rule either during training (so that they received half the training while $v = 12$) or during test using the model outlined in this paper. Hollow circles indicate participants who induced the rule during training, hollow triangles indicate participants who induced the rule at test. All tests were done with $v = 12$ and after 5000 cycles of training.

There are, of course, other ways in which a participant might begin to respond monotonically during the testing stage, which don't necessarily rely entirely upon the same style of explanation that we have offered here. It may be that soon after noticing some dimensional variation, participants are presented with a stimulus that lies at the extreme end of the dimension. If they assume their response to this is correct, then the shift in responding to a monotonic gradient could occur through associating these stimuli with the category appropriate response. However, such an explanation cannot apply to those who reported the correct rule during the training stage, as during this stage only two stimuli (that lay close to the category boundary) were presented. For these participants, who show a monotonic gradient from the outset of testing, there must be something that follows from their inducing the appropriate rule without experiencing extreme values on the dimension.

Figure 5 shows a comparison of simulations for those who induced the rule during training to those who induced it during test (essentially everyone else) with their performance sampled at the end of the testing period. What we see is an advantage in terms of overall performance for those that induce the rule earlier. This mirrors the advantage observed empirically (see Figure 1, Block 3), which we can expect is amplified by those who "get" the rule during training being the faster learners anyway. We would be the first to admit that the detailed fit to the empirical data is not perfect here, in particular, empirically there is less of a difference on the training stimuli than we have in the simulation, and indeed, performance during test on the training stimuli is rather poor. This is something that may well be amenable to a slight revision in the overlap of our representations of the training stimuli, however, and the fact that the large overall effect can be captured is encouraging.

To summarize, we have developed a connectionist network with a novel attentional control mechanism that serves to change both the overall levels of input a connectionist network would be receiving from stimuli, and also alters the underlying representational structure of that input. Connectionist networks that represent stimuli elementally have no problem simulating peak shift, through representing stimuli as a pattern of activation that varies in a Gaussian manner. Our attentional mechanism allows for that, but also generates the monotonic gradient that is often found in peak shift studies in some participants, which is correlated with the ability to articulate a rule. Using this model, we have been able to simulate a similar pattern of results to those found by Livesey and McLaren (2009), who reported that post-discrimination gradients for participants who were unable to induce a rule during training changed from peak shifted, to monotonic throughout testing as exposure to the full range of test stimuli facilitated rule induction. Although we are not suggesting that associative processes are entirely responsible for the monotonic gradient, we believe that propositional processes may be grounded in their associative counterparts and manifest by exerting control over them.

References

Hanson, H. M. (1957). Discrimination Training Effect on Stimulus Generalization Gradient for Spectrum Stimuli. *Science, 125*, 888-889.

Jones, F., & McLaren, I. P L. (1999). Rules and Associations. In *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society, NJ*, LEA

Kruschke, J. K. (1992). ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. *Psychological Review*, *99*, 22-44.

Lee, J. C., Hayes, B. K., & Lovibond, P. F. (2018). Peak Shift and Rules in Human Generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,*44*(12),1955-1970.

Livesey, E. J., & McLaren, I. P. L. (2009). Discrimination and Generalization Along a Simple Dimension: Peak Shift & Rule-Governed Responding. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(4), 554-565.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276-298.

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

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

McLaren, I. P. L., McAndrew, A., Angerer, K., McLaren, R., Forrest, C., Bowditch, W., … & Verbruggen, F. (2019). Mackintosh lecture-: Association and cognition: Two processes, one system. *Quarterly Journal of Experimental Psychology*, *72*(2), 98-117.

Meier, C., Lea, S. E. G., & McLaren, I. P. L. (2016). Task Switching in Pigeons: Associative Learning or Executive Control? *Journal of Experimental Psychology: Animal Learning and Cogntion*, *42*(2), 163-176.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99) New York, NY: Appleton-Century-Crofts

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323,* 533-536.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568.

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

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning.* New York: Academic Press.

Wills, A. J. & McLaren, I. P. L. (1997). Generalization in Human Category Learning: A Connectionist Account of Differences in Gradient After Discriminative and Non discriminative Training. *Quarterly Journal of Experimental Psychology*, *50*(3), 607-630.

Wills, S., & Mackintosh, N. J. (1998). Peak Shift on an Artificial Dimension. *Quarterly Journal of Experimental Psychology, 51B*, 1-32.