The Associability Theory vs. The Strategic Re-Coding Theory: The Reverse Transfer Along a Continuum Effect in Human Discrimination Learning

Mizue Tachi (MT334@hermes.cam.ac.uk)

R.P. McLaren (RPM31@hermes.cam.ac.uk)

I.P.L. McLaren (IPLM2@cam.ac.uk)
Department of Experimental Psychology, Downing Street
Cambridge, CB2 3EB, UK

Abstract

The Reverse TAC effect occurs when learning of a hard discrimination is facilitated by pre-training on an easier discrimination on the same dimension, even though the response assignments used in pre-training on that easy discrimination are reversed when shifting to the harder discrimination. The Reverse TAC effect has been demonstrated in both animal (Mackintosh and Little, 1970) and human experiments (McLaren and Suret, 2003). We consider two explanations for this effect; one is the associability-based hypothesis and the other is the strategic re-coding-based hypothesis. Associability theory gives an account based on a combination of associability processes and generalization. Associability is a learning rate parameter determined by the relative predictability of an outcome based on its relationship with the target stimulus. If the target stimulus is a relatively good predictor of the outcome (compared to other stimuli present), then this will tend to maintain, or increase associability. This then controls the rate of learning in a simple, associatively-based discrimination learning process. On the other hand, the strategic re-coding theory explains the Reverse TAC effect by appealing to a combination of the adoption of some cognitively derived strategy with simple discrimination learning and emphasizes the interaction between cognitive and associative processes.

To assess these hypotheses, we carried out two experiments focusing on the effect of varying the amount of training. Our conclusion is that the strategic re-coding hypothesis is the less plausible account of the Reverse TAC effect. Instead, our empirical data suggested that if subjects' initial motivation was relatively low and they faced a situation where their performance deteriorated (through no fault of their own) over an extended period of time, then the associability of the to-be-discriminated stimuli would decrease, resulting in a deterioration in their performance.

Introduction

How do people discriminate objects in the environment, and what mechanisms underlie this process? Perhaps the most plausible explanation for these skills may be found among associative learning theories. The basic idea of these theories is that learning progresses via the development of associations between target stimuli and subsequent events. Our strategy was to look at some specific phenomena in learning in order to make progress in understanding the mechanisms involved.

Some years ago Mackintosh and Little (1970) demonstrated what were then controversial results in a pigeon experiment; that pre-training on an easy discrimination could facilitate acquisition of a harder discrimination using stimuli drawn from the same dimension, even when the response assignments used in pre-training were reversed when shifting to the harder discrimination. Thus, if we denote four stimuli on a dimension by A, B, C, D, such that A vs. D is the easy discrimination, and B vs. C the hard one, the finding is that pre-training on A+ D- will facilitate acquisition of B- C+. This is the basic Reverse TAC effect which was the focus of this research. Recently, McLaren and Suret (2003) were able to replicate Mackintosh and Little's (1970) finding in humans. This effect, the Reverse TAC effect, was successfully explained in Mackintosh and Little's 1970 paper by appealing to the notion of associability. McLaren and Suret (2003) also adopted Mackintosh and Little(1970)'s idea by way of explanation for their human empirical data, assuming that the same, associatively-based processes, were operating in pigeons and humans.

The basic mechanism underpinning the TAC effect is often taken to be simple generalisation of the associations acquired between stimuli and outcomes during pre-training to those acquired during training, a process that undoubtedly contributes to standard demonstrations of Transfer Along a Continuum (TAC) (Lawrence, 1952). In TAC, the procedure is to pre-train A+ D-, then shift to B+ C-. Both associability processes and generalisation may be expected to assist in acquisition of the more difficult discrimination in this case, resulting in more rapid learning compared to controls trained on B+ C- for the same total number of trials. It is the finding of more rapid acquisition of B- C+ after pre-training on A+ D-, however, that demonstrates an effect of associability that cannot be explained by (indeed is contrary to) simple generalisation. Associability theory, such as that formally proposed by Mackintosh (1975), can, however, deal with this, and Mackintosh and Little (1970) provided an elegant explanation for the Reverse TAC. An associability theory essentially requires that a learning rate parameter (that may, perhaps, also influence performance) be postulated whose magnitude is governed by an organism's prior history with the stimulus and outcome under consideration. Large values of this parameter promote rapid learning, small values slow learning. Application of this approach to Mackintosh and Little (1970)'s results involves noting that the easy discrimination used in pre-training is (not surprisingly) rapidly acquired. This is taken to maintain a high level of associability for the features of the stimuli A and D that support the discrimination between them, in line with the idea (Mackintosh, 1975) that relatively good predictors of an outcome will have high associability. In contrast, the non-predictive features (those shared by both stimuli or irrelevant to the discrimination such as left / right position) will tend to experience a drop in associability. The effect is that the organism's learning is focused on the predictive elements of the discrimination. Then, on transfer to the hard discrimination, because B will be more similar to A than D, and C more similar to D than A, it will be exactly the elements of B and C that best distinguish between them that will have the highest associabilities, promoting learning of the discrimination.

This conclusion seems secure when applied to Mackintosh and Little (1970)'s data, but what of McLaren and Suret (2003)? Humans are typically considered to have more resources available to them than pigeons, and this introduces the possibility of other explanations for this result. Perhaps the most obvious is suggested by consideration of Hall and Honey's (1989) analysis of acquired distinctiveness in combination with work on verbal discrimination learning (Ekstrand, Wallace and Underwood, 1966). If pre-training is taken to result in good acquisition of the response "left" to A and "right" to D (as these experiments use two keys on a computer keyboard as outcomes) then the stimuli A and D have representations whose distinctiveness has been further enhanced by association to these two different outcomes. Then, on transfer to B- C+, generalisation to B from A will tend to produce the outcome "left" (incorrectly so) and generalisation from D to C "right". This would not aid pigeons, but say that humans notice that they are suddenly getting things consistently wrong, and then cognitively re-configure themselves by adopting the strategy " when I feel like making a left response I'll press the right key, and when I feel it's the right key I'll press the left key". There is no need for an appeal to associability in this explanation, just a combination of the correct strategy or task set and simple generalisation, much as in classic demonstrations of verbal discrimination learning (Underwood, Jesse and Ekstrand, 1964).

Now that there are two possible explanations in play for what may, in short, be termed the Reverse TAC effect, the question naturally arises of how to distinguish between them. Both theories assume that discrimination learning will proceed by development of associations between target stimuli and allocated outcomes in pre-training, and discrimination of the target stimuli (harder discrimination) in training is facilitated by generalization from those stimuli (easier discrimination) employed in pre-training.

Here, the previously cited work on verbal discrimination learning is our guide. It strongly suggests that use of a strategy that requires the production of a response opposite to the one previously learned can result in good initial performance in a transfer test, but that with extended training performance will not improve and may even decline. The reason for this is as follows; the previously acquired associations play the role of “mediators” for the to-be-learned learned associations during training. But during training itself, these associations will decrease in strength, and the new, opposing associations between stimuli and outcomes will increase in strength. This follows quite straightforwardly as a consequence of the assumption that associatively-based learning continues during training on the hard discrimination. Given that the hypothesis is that subjects are using the strategy of pressing the left key when they are shown stimulus C on the basis that they feel like making a right key response, training will reliably link the occurrence of stimulus C with a left key press, and have the effect of building up the C – Left association and extinguishing the C – Right association. Over trials this will degrade, and eventually eradicate the basis for the subject’s decision to press the left key! The prediction, then, is that more training after the reversal will certainly not help performance if this hypothesis is correct. An associability analysis, on the other hand, should be relatively unaffected, since this theory does not assume any cognitive mechanisms such as strategic re-coding of learned associations, so there should be only direct connections between the target stimuli and the currently trained outcomes (i.e., there is no use of old associations to implement new associations). In Experiment 1 we test these predictions by using 2 blocks as well as 1 block of training after the standard pre-training regime used in previous experiments. Previous pilot experiments looking at 1 and 2 blocks of training separately had indicated that we might expect the effect predicted by the strategic re-coding hypothesis (Tachi, 2004), but Experiment 1 is the first to test for the effect of varying the amount of training within a single experiment.

Experiment 1

Stimuli and Apparatus

The stimuli used in Experiment 1 were 44 monochromatic pictures of faces, which were specially designed to create four discriminable dimensions. They were grouped into four subsets (two male and two female) that consisted of 11 faces each. Each subset was on an artificial dimension defined by two original passport photos of university undergraduates at the poles and nine intermediate morphed pictures, which were created using a standard morphing software package, Morph, between these poles. All stimuli were assigned the numbers from 1 to 44 indicating their order on each dimension (i.e. 1-11, 12-22, etc.).

If we take the dimension as spanning stimuli 1-11, pictures 3 and 9 from each dimension were always shown when subjects were engaging in the “Easy” task, and pictures 5 and 7 were shown as the Hard stimuli (see Figure 1 for an example).

On test, all 44 stimuli were shown to the subjects to measure not only discrimination between the critical stimuli at 5 and 7 on each dimension, but also to look at generalization along each dimension. Although designated as the “Easy” discrimination, stimuli 3 and 9 on each dimension were still relatively similar to one another, and learning the discrimination may have been quite difficult. We hoped that this would make the subjects more likely to employ some associative process rather than relying on rules to distinguish between these stimuli.

The experiment-running programme was written in REALbasic and run on an Apple Macintosh computer. The stimuli were shown as greyscale images 3.5cm by 4.5cm. This experiment was conducted in a moderately illuminated and quiet room, away from any external noise.

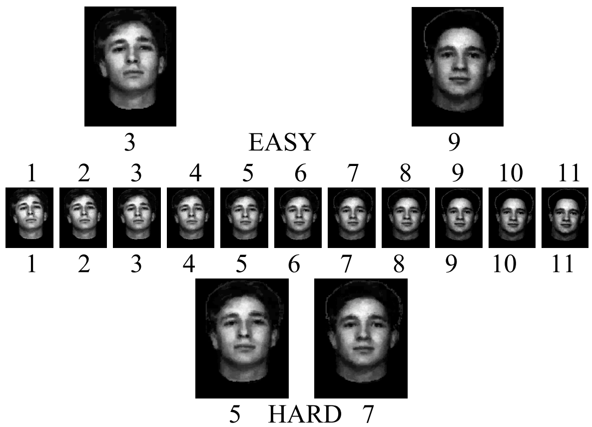

Figure 1: The morphed face dimension.

Subjects and Design

The 32 subjects in this experiment were undergraduates and graduates from the University of Cambridge. They were randomly allocated to four conditions, ‘Easy Reversed 1’, ‘Easy Reversed 2’, ‘Hard 1’ and ‘Hard 2.’ The factors differentiating these groups were the contents of pre-training (easy vs. hard stimuli) and the amount of training after the pre-training (1 or 2 blocks). All subjects received three blocks of pre-training on the four face dimensions for a fixed number of trials (40 trials per block, five presentations on each of the two faces used from each dimension). In the Easy Reversed conditions, subjects were trained on the easy discriminations between stimuli 3 and 9 with opposite response assignments to those used in later training, denoted as 3+ 9-. In the Hard conditions subjects were pre-trained on the hard discrimination between stimuli 5 and 7 with congruent response assignments to those used in the training phase, denoted as 5- 7+. The + and – simply indicated different response assignments; either the ‘X’(left) or ‘.’(right) key press, and these assignments were counterbalanced across subjects. After the pre-training phase, subjects received either 1 or 2 blocks of training on the Hard stimuli (i.e. 5- 7+) for all four dimensions (40 trials per block, five presentations per stimuli). This was followed by a test phase that consisted of five blocks of all 44 stimuli (all 11 stimuli from each of the four dimensions) presented without feedback, so that performance across each dimension could be assessed. The data of interest were the responses to the stimuli, especially the trained stimuli (5 and 7), in the final test phase. Significant differences in these responses would indicate acquisition of the trained discriminations.

Procedure

At the beginning of each phase of this experiment, subjects were given instructions by the experimenter; they were told that they would be shown a series of stimuli in the form of human faces on the screen, and that their task was sorting these pictures into two categories by pressing one of two keys (‘X’ or ‘.’ keys) on the computer keyboard. They would receive immediate feedback on the correctness of their responses. So their task was simply that of finding out and remembering for which stimuli the ‘correct’ response was the ‘X’ key press and for which stimuli the ‘correct’ response was the ‘.’ key press. If they did not respond within a time limit of four seconds they would be timed out and a message would appear on the screen telling them so before the appearance of the next stimulus. Target stimuli were presented one at a time. Each trial started with a fixation cross ‘+’for 1.5 seconds, which was replaced by a face for a maximum of 4 seconds. The picture disappeared once subjects made a response to it or were timed out. Feedback was given for 1.5 seconds with either ‘correct’ or ‘wrong’ displayed in the center of the screen. If the subjects pressed an invalid key, the message ‘invalid key pressed’ would appear. If the subjects responded before the presentation of the stimulus, the message ‘you anticipated the probe’ was displayed before the stimulus appeared on the screen, and the program moved on to the next trial. After completing the three blocks of pre-training, subjects engaged in a totally unrelated paper-based experiment that lasted for approximately 15 minutes, and then they continued the training and test phases of Experiment 1. This procedure was used because the inclusion of a 'gap' between pre-training and training had been found to be the most effective way of obtaining the Reverse TAC effect by Suret and McLaren (2003). On the completion of the pre-training and training phases, subjects progressed to the test phases where stimulus presentation was as before but with no feedback (this was replaced by a 1.5 second pause with a blank screen between the subjects’ response and the next stimulus). Prior to the test phase, subjects were told to categorize the stimuli into two groups (‘X’ or ‘.’) based on the criteria that they had developed in the last phase of training.

Results

The results of Experiment 1 are shown in Figure 2. Following the approach taken in previous studies (i.e. McLaren and Suret (2003)) which used the same stimuli and similar designs looking at other aspects of discrimination learning we collected data from the test phase only. In all statistical analyses presented here the probabilities are two-tailed unless otherwise specified, and were carried out on the data converted into mean response scores. This is a direct index of discrimination performance in our subjects. One key was designated negative (a key press scores -0.5 for that stimulus) and the other key positive (scores +0.5) during test. Hence, the range of the mean response score is from -0.5 to +0.5 and it would be 0 for a given stimulus if subjects did not show any response preference. The critical comparison in this experiment is the discrimination between the stimuli at positions 5 and 7 along each dimension, so we computed a difference score as our final measure for this discrimination. This is the score for all trained dimensions calculated by taking the mean response score for stimulus 5 away from the mean response score for stimulus 7. A positive difference score provides some evidence of learning of the discrimination during the training, perfect learning would be indicated by a score of 1, indifference by a score of 0, with a floor for this index of -1.

An analysis of variance (ANOVA) with pre-training task (easy reversed/hard) and training amount (1block/2blocks) as factors gave an $F(1,28)=3.66$, $p<.05$ (1 tail) indicating an interaction between these factors (no other effects were significant). A one tailed test was deemed appropriate because the earlier pilot work by Tachi (2004) had produced this pattern of results across experiments. Further analysis by means of t-tests revealed that in the Easy Reversed condition, where subjects were pre-trained on the easy discrimination, the group who received 1 block of training after reversal showed significantly better performance (congruent with their training on the hard problem) than those who had 2 blocks of training after reversal ($t(14)=2.78$, $p<.05$). While in the Hard condition, where subjects, in effect, only experienced the training discrimination (Hard task) during the pre-training phase as well, the 2-blocks-trained group seemed to perform better than the 1-block-trained group (but not significantly so $t(14)=0.57$, $p>0.1$). Furthermore, when subjects received 1 block of training, performance in the Easy Reversed group was significantly better than that of the Hard group ($t(14)=2.15$, $p<.05$). However, when subjects had an additional extra block of training, learning in the Easy Reversed group became significantly worse ($t(14)=2.78$, $p<.05$), and this group was now numerically worse than the Hard group that also had two blocks of training (but not significantly so, $t(14)=1.08$, $p>0.1$).

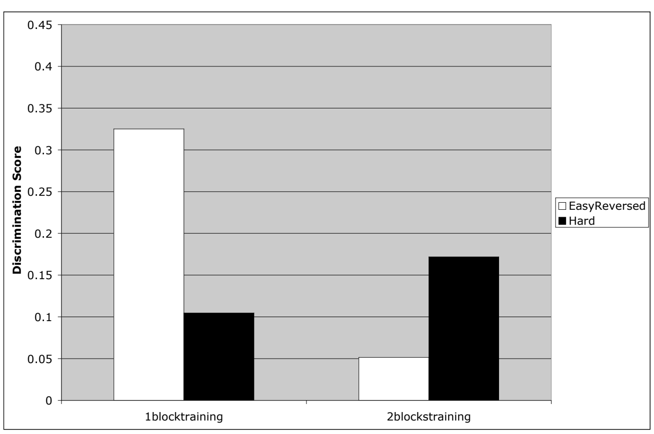

Figure 2: Results for Experiment 1.

Discussion

The results from this experiment lend considerable support to the strategic re-coding hypothesis put forward in the introduction. It would seem that further training actually makes subjects worse in the Easy Reversed condition, and this can readily be explained by postulating that the 'feeling' to press a particular key which is then re-coded into the opposite response is degraded by the extra training given in this condition. Another possibility, however, is that subjects in the Easy Reversed condition are becoming increasingly confused as training progresses. They are not warned that response assignments are switched as they move from pre-training to training. Inevitably they make many mistakes at first in training immediately after the reversal has taken place. The group given only one block of training are then moved to test, where mistakes are no longer an issue as there is no feedback, but the group given two blocks of training continue to get feedback and may either become convinced that the task is impossible, and give up by the time they get to test, or the more prolonged experience of relatively low levels of performance may drive the associability of the discriminatory stimulus features down. This may seem a somewhat post-hoc appeal to either motivational or associability factors, but it was noticeable that subjects in the 2 blocks Easy Reversed condition seemed confused about the task at the end of the experiment and some indicated that they thought the task too difficult to do. A simple way to evaluate one of these alternative explanations for our results is to maintain the same experimental design, but ensure that subjects remain highly motivated throughout the experiment. This was achieved by offering bonuses for performance during training and test in Experiment 2. The rationale for this is that the prospect of the bonuses will prevent subjects giving up on the task, but would have little effect on any re-coding that they might do. Indeed, the further inducements might actually encourage re-coding as a strategy for performing the task.

Experiment 2

Subjects and Design

32 subjects in this experiment were taken from the same population as in the previous experiment. They were randomly divided into 4 groups, ‘Easy Reversed 1’, ‘Easy Reversed 2’, ‘Hard 1’ and ‘Hard 2’ as in the previous experiment. All conditions included a time gap between the pre-training and training phases. The experimental design was the same as for Experiment 1, with the exception of the dummy bonus system that we used to motivate subjects during the training phase. In this design, at the beginning of the training phase subjects were instructed that they would earn additional money if they learnt very well, but, in fact, all subjects received a message “Congratulations! Bonus awarded!” at the end of the training no matter how well they learnt. Otherwise procedures were as in Experiment 1.

Results

Figure 3 shows the results of Experiment 2.

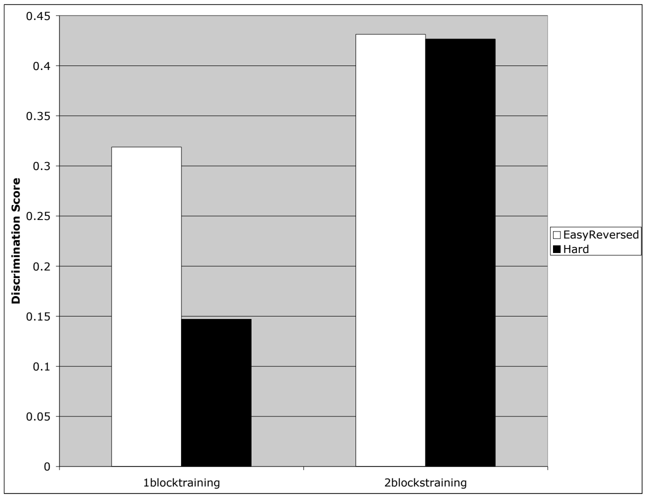

Figure 3: Results for Experiment 2.

A two-way factorial ANOVA with pre-training task (easy reversed/hard) and task amount (1-block/2-blocks) as factors did not indicate any interaction between them this time, $F(1,28)=1.11$, $p>0.1$. There was no longer poor performance in the Easy Reversed 2-block group relative to the Hard 2-block group, rather there was a small non-significant improvement. As we shall see, this result, when taken in combination with those of Experiment 1 poses a significant challenge for the strategic re-coding hypothesis. There was, however, a significant main effect of task amount, $F(1,28)=5.75$, $p<.05$, indicating that two blocks of training were more effective than one.

Further analyses showed that after 1-block of training the Easy Reversed group performed better than the Hard group ($t(14)=2.00$, $p<0.05$ 1-tail), however, as already noted, in the 2-block-training condition, there was no significant difference between them. Thus we can argue that the standard Reverse TAC effect has been maintained in Experiment 2, whereas the effect of the extra block of training on the Easy Reversed – Hard difference has disappeared.

Next, we compared results between Experiment 1 and 2. A three-way factorial ANOVA with pre-training task (easy reversed/hard), training amount (1-block/2-blocks) and motivational state (normal/high) as factors gave a main effect of motivational state, $F(1,56)=7.73$, $p<.01$, reflecting the generally higher level of performance seen in Experiment 2. Whilst ANOVA did not indicate an interaction between all three factors, there were significant interactions between pre-training task and training amount, $F(1, 56)=4.48$, $p<.05$, and also training amount and motivation, $F(1, 56)=6.18$, $p<.05$. No other effects were significant. Taking the second interaction, that between training amount and motivation, first; further analyses by means of t-tests on the individual comparisons between experiments revealed (perhaps unsurprisingly) that high motivation facilitated learning. Note, however, that this effect was only significant in the longer training condition. In the conditions where subjects received 2 blocks of training, highly motivated subjects who were trained with the bonus system (Experiment 2) performed significantly better on test than those without the bonus system (Experiment 1) no matter what they had experienced during the pre-training phase ($t(14)=4.02$, $p<.05$) for the Easy Reversed condition, ($t(14)=2.35$, $p<.05$) for the Hard condition). The 1-block training conditions show a different pattern of results since they seemed not to be affected by motivation. There was no significant difference due to different motivational states in any of the conditions. In other words, further comparisons between the 1-block-training conditions in Experiment 2 and other conditions in Experiment 1 implied that, in general, subjects' performance after 1 block of training with the increased motivation was not different from that after normal 1-block-training. Thus, according to these comparisons between Experiment 1 and 2, the increased motivation induced in Experiment 2 facilitates learning in relatively prolonged training situations, but has less influence in the short training conditions. Turning now to the other significant interaction to emerge from this analysis, that between type of pre-training and training amount, the interpretation is relatively straightforward. The Reverse TAC effect is observed after 1 block of training, but not after 2 blocks.

General Discussion

The results in Experiment 1 demonstrated that after 1 block of training subjects pre-trained on the Easy Reversed discrimination showed significantly quicker learning on the training task (Hard discrimination) than those pre-trained on the Hard discrimination throughout, however, their performance suddenly changed for the worse if training was prolonged (an additional block of training). After 2 blocks of training there was no benefit as a result of the Easy Reversed pre-training when compared with the Hard pre-training. What are the implications of these findings for the strategic re-coding and associability accounts of the basic Reverse TAC effect considered in the introduction to this paper?

If the relatively poor performance in the Easy Reversed conditions after 2 blocks of training is simply due to motivational factors, then the associability account is entirely consistent with the data. In this case the explanation of the effect is simply orthogonal to the mechanism for the Reverse TAC effect, and does not in any way constrain our theorising with respect to TAC. Equally, if the effect is a consequence of the reduced associability of stimulus representations then this would add further support for associability processes in human learning. Both these hypotheses could be consistent with the results of Experiment 2. In this experiment the bonus scheme is taken to increase the motivation of subjects during the training and testing phases. Clearly, if the problem in Experiment 1 is that motivation collapses in the course of 2 blocks of training, then measures to increase it should ameliorate the effect, and this is what is observed. The associability-based explanation is more subtle. If we assume that 2 blocks of training allow the associability of the stimulus representations to decline in the Easy Reversed case, then why should an increase in motivation change this? Our hypothesis is that the increased motivation improves the rate at which learning of new associations takes place, without directly affecting the rate at which associability changes are made. If this is so, then the more rapid acquisition of the new associations should lead to the appropriate components of the stimulus representations becoming established as good predictors for the outcomes relative to other components present, and so prevent their associability declining. This explanation hinges on the task being parameterised such that the learning rate is low enough in Experiment 1 for associability in the Easy Reverse condition to decline after 2 blocks of training, but not after 1 block of training. More simulation work accompanied by empirical tests will be needed to see if this is a plausible assumption.

On the face of it, the strategic re-coding hypothesis comes out of this well. It can explain the basic Reverse TAC effect by postulating that subjects in the Easy Reverse condition learn very quickly to press the left key when they feel the urge to press the right key and vice .versa. The loss of the Reverse TAC effect after 2 blocks of training is exactly as the hypothesis predicts. Further training will lead to acquisition of the new stimulus-outcome associations and extinction of the old ones, on the assumption that associative learning continues automatically in these circumstances. This will mean that when confronted with a stimulus subjects will at some stage find themselves in an ambiguous position, unsure of which key they wish to press. Eventually they will have to abandon the very strategy that allowed them to succeed at the task! The strategic re-coding explanation finds itself in some difficulty when considering Experiment 2, however, as it must predict that more training will not improve matters, at least not at first, and that this cannot be avoided simply by increasing motivation. To see this, consider the following attempt to reconcile this hypothesis with the results of Experiment 2 where subjects were relatively highly motivated throughout their training. It would be natural to assume that this results in faster learning during training, and that this might allow subjects to progress through the stages of initial re-coding followed by abandonment of this strategy more rapidly. Even if we postulate that the increased motivation means that subjects learn so fast that the dip in performance after 2 blocks of training observed in Experiment 1 is passed by, then, in this case, the effect seen in Experiment 1 should be seen after 1 block of training rather than 2. It is difficult with this hypothesis to explain the maintained Reverse TAC effect after 1 block of training on the one hand, and the lack of any difference in performance after 2 blocks of training on the other. The result does not conclusively rule out this hypothesis, but does make it less plausible.

Our conclusion is that a line of research that began by promising to support a strategic account of the Reverse TAC effect in fact now calls that hypothesis into question. Thus, the associability theory, in conjunction with some consideration of motivational factors, perhaps provides the best explanation for the results. And this implies that human learning and animal learning have the same associatively-based mechanisms. In order to sustain this account we have to allow that if subjects' motivation is moderate and they are put in a situation where their performance worsens (through no fault of their own) in the absence of any information to help them explain this occurrence, and if this experience is relatively prolonged, then their motivation to do the task will suffer, and their performance collapses as a consequence. Alternatively, it may be that 2 blocks of training after reversal have some more stimulus specific, rather than general motivational effect. Clearly the potential exists for reversal to drive stimulus associability down, as until the previously established associations have been unlearned the stimulus features will predict the wrong outcomes. Our explanation of the Reverse TAC effect postulates that these features have high initial associability, and so will form new associations to the correct training outcomes during this period. It may be that after 1 block of training associability is still high and the new, correct associations are just starting to outweigh the previously established, incorrect associations. But after 2 blocks of training associability has dropped sharply, and if this is reflected in performance as well as learning, then this will negate the growth of the correct associations to these stimulus features, and performance will improve slowly, if at all. It will be a matter for future research to determine which, if either, of these explanations for our results is the correct one.

References

Ekstrand, B.R., Wallace, W.P. and Underwood, B.J. (1966) A Frequency Theory of Verbal Discrimination Learning. *Psychological Review*. 73, 6, 566-578.

Hall, G. and Honey, R.C. (1989) Contextual Effects in Conditioning, Latent Inhibition, and Habituation: Associative and Retrieval Functions of Contextual Cues. *Journal of Experimental Psychology: Animal Behaviour Processes*. 15, 3, 232-241.

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

Mackintosh, N. J. and Little, L. (1970). An analysis of transfer along a continuum. *Canadian Journal of Psychology*, 24, 362-369.

Mackintosh, N.J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 4, 276-298

Suret, M. and McLaren, I. P. L. (2003). Representation and discrimination on an artificial dimension. *The Quarterly Journal of Experimental Psychology*, 56B, 1, 30-42.

Suret, M. (2003) *Salience Modulation and Associability Processes in Transfer along a Continuum.* Doctoral dissertation, Department of Experimental Psychology, University of Cambridge.

Tachi, M. (2004) *How to bring chaos in the complete discrimination? Focusing on critical factors & considering models of discrimination learning*. First year report, Department of Experimental Psychology, University of Cambridge.

Underwood, B.J., Jesse, F. and Ekstrand, B.R. (1964) Knowledge of Rights and Wrongs in Verbal-Discrimination Learning. *Journal of Verbal Learning and Verbal Behaviour*. 3, 183-186.