Please cite as:

McLaren, I.P.L**.**, Carpenter, K., Civile, C., Milton, F. McLaren, R., Zhao, D., Ku, Y. and Verbruggen, F. (in press). Categorisation and Perceptual Learning: Why tDCS to Left DLPFC Enhances Generalisation. To appear in *Associative Learning and Cognition, Homage to Prof. N.J. Mackintosh*, Trobablon, J.B. and Chamizo, V.D. (Eds.), University of Barcelona.

Categorisation and Perceptual Learning:
Why tDCS to Left DLPFC Enhances Generalisation

I.P.L McLaren[1], K. Carpenter[1], C. Civile[1], R.McLaren[1], D. Zhao[2], Y. Ku[2], F. Milton[1], and F. Verbruggen[1]

[1]School of Psychology, College of Life and Environmental Sciences,

University of Exeter, UK

[2]School of Psychology and Cognitive Science,

East China Normal University, Shanghai, China

All correspondence concerning this chapter should be addressed to:

I.P.L. McLaren

Washington Singer Laboratories

University of Exeter

Perry Road

Exeter EX4 4QG

i.p.l.mclaren@exeter.ac.uk

Abstract

In the 27 years that have passed since the McLaren, Kaye and Mackintosh (MKM) model of perceptual learning was first proposed, it has undergone considerable theoretical development and been subject to extensive empirical test. But we would argue that the basic principles of the theory remain as valid today as they were in 1989. One of these principles was that salience modulation of stimulus representations based on prediction error was a key component of latent inhibition and perceptual learning. It was this modification of what was otherwise a fairly basic adaptation of the model for categorisation proposed by McCleland and Rumelhart (M&R) that transformed a system that would exhibit enhanced generalisation as category learning progressed, into one that would instead offer an improved capacity for discrimination between exemplars as a consequence of experience with the category. This modification has only been tested indirectly up until now, by looking at the predictions that flow from it and then comparing them to animal and human discrimination following stimulus pre-exposure. In this chapter we test this principle more directly, by using tDCS to disrupt the modulation of salience by prediction error, and show that when this is done, people exhibit the enhanced generalisation predicted by the standard M&R model. We conclude that our results provide further support for the MKM approach to stimulus representation.

How we learn to distinguish between things is one of the basic questions for cognitive psychology. This paper focuses on two aspects of the mechanisms that allow us to do this. Categorisation in this paper refers to our ability to classify stimuli as members of one category or another as a result of trial and error training with members (exemplars) of the categories in question. Perceptual learning here refers to our enhanced ability to discriminate between certain stimuli as a consequence of experience with them or stimuli like them. Taken together, these two phenomena play a crucial role in learning to correctly identify stimuli as members of a particular class, and not confuse one stimulus with another similar one.

There are many theories and models of categorisation, and quite a few theories and models of perceptual learning. One of the few models that addresses both was originally proposed by McLaren, Kaye and Mackintosh (1989, henceforth MKM) in part as a response to and development of McClelland and Rumelhart's (1985, henceforth M&R) connectionist model of categorisation. It is this model that motivated the experiments discussed here, and, given the model-driven nature of our enquiry, we begin with a brief introduction to these models and the experimental paradigms we will use in this paper. We then go on to discuss how recent work using tDCS (trans-cranial Direct Current Stimulation) raises the possibility of influencing the error signal that drives learning and performance in the MKM model so as to change participant's ability to distinguish between stimuli as a consequence of their experience with them. Our paper is an exploration of this possibility, and our results suggest both that perceptual learning and categorisation can be strongly influenced by anodal tDCS to frontal regions of the brain, and that a theory of perceptual learning and categorisation that relies on use of error-based modulation of the salience of the representations of stimulus input provides a good fit to the data we obtain using this preparation. We end by discussing the implications of these results for phenomena such as face processing.

Background

Two Models

McClelland and Rumelhart's seminal 1985 paper used the delta rule, an error correcting learning algorithm closely related to Rescorla-Wagner, (Rescorla and Wagner, 1972) in a connectionist network employing distributed stimulus representations to model categorisation. We cannot do the model full justice here, but it was also noteworthy for its use of non-linear activation

functions and a weight decay mechanism to help it produce both prototype and exemplar effects in what was effectively a single-layer (in that it has a single layer of modifiable weights) connectionist model. It did have one feature, however, that seemed to some of us problematic. This was that the learning algorithm coupled with the activation function inevitably led to units that were most frequently co-activated becoming more active as a consequence. This gave these units greater salience in later learning, and so it would be the units representing the more prototypical elements of a stimulus that would tend to form the strongest links to other units representing category membership.

This characteristic of the model may not be a problem for categorisation (though we will have more to say about this later) but it is certainly a problem for stimulus representation development as a consequence of experience with a category, i.e. for perceptual learning. McLaren, Leevers and Mackintosh (1994) were the first to show that humans trained to distinguish between two prototype-defined categories of stimuli (in this case chequerboards) were then actually better able to distinguish between two new exemplars drawn from one of these now familiar categories than between two exemplars taken from another entirely novel category that otherwise had a similar prototype-defined structure. The McClelland and Rumelhart model predicts the opposite result because, as illustrated in the lower half of Figure 1, it will be the prototypical features contained within these new exemplars drawn from the familiar category that will be most salient. This will lead to the two exemplars being represented as more rather than less similar as a consequence, because these will tend to be the features shared by the two exemplars.

Figure 1 about here please

Our solution to this problem is shown in Figure 1 (top half), which illustrates how the MKM theory predicts salience will change as a function of experience with exemplars of one category. The crucial difference between this model and that of McClelland and Rumelhart (1985) is that the activation of the units representing stimulus features (or elements as we will often call them) is modulated by their error. Thus, if a unit is relatively unpredicted by other active units, but is externally activated because a feature corresponding to that unit has occurred and been perceived in the environment, then it's activation (i.e. salience for learning purposes) will be high because its error score will also be high. Conversely, if a unit receiving external input is well-predicted by other active units such that its error score is low, then its activation (and salience) will be low. This is exactly the opposite of the effect that occurs in the M&R model, and leads to new exemplars drawn from a

familiar, prototype-defined category being more easily discriminated, because the elements or features they share in common will be relatively low in salience, thus reducing stimulus similarity. The elements on which they differ (which will tend to be those elements that have changed from the prototype) will be relatively salient and this helps in learning to discriminate between them as McLaren, Leevers and Mackintosh (1994) found.

The illustrations in Figure 1 are similar to those of a figure in McLaren (1997) that is also used to explain how experience with exemplars drawn from a prototype-defined category will lead to better within-category discrimination. This 1997 paper, however, deals with one of the first reports of an analogue of the face inversion effect using artificial categories (again chequerboards) rather than faces. McLaren first trained participants to learn (by trial and error) to categorize chequerboard exemplars as belonging to one of two prototype-defined sets. The exemplars were made from the prototypes by randomly changing some of the black and white squares that made up the chequerboard that defined the category prototype, as in McLaren, Leevers and Mackintosh (1994). McLaren (1997) then demonstrated that an inversion effect could be obtained for new exemplars drawn from these now familiar categories, a result since replicated repeatedly by Civile, Zhao, Ku, Elchlepp, Lavric and McLaren (2014). The explanation for this result is that the exposure to the exemplars of the categories participants were trained on initially allows perceptual learning to take place as in McLaren *et al* (1994), and this then improves discrimination and recognition performance to exemplars drawn from those categories that are in the usual upright orientation; but does not help, and as Civile *et al* (2014) argue, actually hinders discrimination and recognition when these exemplars are inverted. This explanation depends on the MKM account of perceptual learning and categorisation, as the M&R model would once again predict the converse result.

There is thus some good evidence for the MKM modification of the M&R model of categorisation. We will use the categorisation followed by discrimination/recognition procedure just discussed later in this paper to test our hypotheses regarding the effects of frontal anodal tDCS stimulation on perceptual learning. But before doing this, we first consider the prior issue of what tDCS might be able to offer in terms of influencing categorisation itself, and how tDCS might affect the type of error-based modulation that is the basis of the MKM model.

tDCS and Categorisation

Our first experiment investigates the effects of tDCS on a standard categorisation task that produces a prototype effect under normal circumstances (Posner and Keele, 1968). This work was inspired by the finding of Ambrus, Zimmer, Kincses, Harza, Kovacs, Paulus and Antal (2011) who provided evidence that tDCS could eliminate the prototype effect. There is other evidence that stimulation of PFC using tDCS can influence categorisation. Lupyan, Mirman, Hamilton and Thompson-Shill (2012) have produced some evidence that stimulation in frontal regions can enhance categorisation, and Kincses, Antal, Nitsche, Bártfai and Paulus (2003) have shown that when tDCS anodal stimulation was delivered over the left PFC (Fp3), probabilistic classification learning (PCL) was improved. Ambrus *et al* (2011), however, found that anodal tDCS, applied to Fp3 during the training phase (and beginning 8 minutes before the training phase started) had a significant and quite different impact on categorisation performance in their version of the prototype distortion task. They obtained a significant decrease in performance accuracy in identifying prototype and low-distortion patterns as category members in the anodal group compared to the sham group. This is a striking aspect of their results as it is contrary to most studies that show increased performance when anodal tDCS is applied to task-relevant cortical areas during task execution (e.g., Fregni, Boggio, Nitsche, Bermpohl, Antal, Feredoes, 2005).

On close inspection, one possible interpretation of Ambrus *et al's* result is that anodal tDCS has reduced learning to the prototype, and increased generalisation to random patterns. This would have the effect of eliminating any prototype effect, and is exactly the type of pattern we would expect if the MKM model were to be transformed into the M&R version. Salience modulation enhances learning of novel stimuli, and so improves early acquisition of category discrimination, and it also reduces generalisation. Losing this type of modulation would lead to slower learning (at least initially) and greater generalisation. We speculated that anodal tDCS to Fp3 might have disrupted salience modulation by means of prediction error leading to Ambrus *et al's* result.

If we now consider how tDCS might influence the brain's computation and use of prediction error, Reinhart and Woodman (2014) in a recent paper have shown that anodal tDCS over frontal regions can change prediction error. They used anodal stimulation at FCz and were able to show that this produced enhanced learning and selectively enhanced neural correlates of prediction error. The most obvious conclusion to draw from this study is that 1.5 mA anodal stimulation applied with their electrode montage has the effect of amplifying prediction error, which will both speed learning and

lead to the neural signature they found. This is not the effect we postulated in response to Ambrus *et al's* data, but it does suggest that prediction error can be influenced by anodal tDCS, and of course the locus of stimulation is rather different in Reinhart *et al's* work.

Our approach in the studies reported in this paper is to take something from the approaches of Ambrus *et al's* (2011) - because they were able to influence categorisation quite directly - whilst holding that of Reinhart *et al* (2014) in mind - because they have good evidence for changing prediction error. Hence we employed a similar electrode montage to that use by Ambrus *et al* (2011) stimulating Fp3, and increased the current from the 1mA they used to 1.5 mA in the hope of maximising our chance of observing an effect on categorisation. If we were to observe such an effect, then we would consider the possibility that this effect would be due to our changing the contribution of prediction error in influencing learning and performance on the categorisation task. In this way we hoped to develop a procedure that would allow us to both influence categorisation and the perceptual learning that follows on from categorisation, which in turn would allow us to probe the mechanisms underlying both, using the MKM modification of M&R as our starting point for interpreting our results. Note that our procedure, which is akin to that used in earlier studies, employs electrodes (see later) that do not have a strongly focal effect, so that the stimulation we provide is perhaps best functionally described as Left DLPFC rather than trying to claim any greater specificity.

Experiment 1

Experiment 1 is a conceptual replication of Ambrus *et al* (20111), using a classic categorisation paradigm based on early work by Posner and Keele (1968) and Homa, Sterling and Treppel (1981) designed to reveal any prototype effect. We use three prototype-defined categories of chequerboards, with the exemplars in each category generated by adding noise (randomly changing a certain number of squares) of the prototype for that category. Participants are trained to classify exemplars into these three categories by trial and error, and then tested on exemplars and the prototypes (which are never shown in training) to allow us to determine if an exemplar effect has occurred. Three types of stimulation, Anodal, Cathodal and Sham are used, but all employ the same Fp3 electrode placement used by Ambrus et al.

Method

Participants: 50 University of Exeter students (17 male) with a mean age of 21.5 years (sd 2.93) participated in the study. Two were excluded before analysis due to procedural complications leaving 48.

Stimuli: These were 16x16 chequerboards containing approximately 50% black and 50% white squares. Four prototypes were created that were constrained to share 50% of their squares with one another, and also to consist of relatively clearly demarcated regions of black and white. This was achieved by making the colour of a given square depend on that of its near neighbours. Thus, if they were predominantly black then it was likely to be black, and vice-versa if the neighbours were predominantly white. Exemplars were generated by adding noise. A randomly chosen 96 squares would be set at random in a given prototype to generate an exemplar of that category, so that on average 48 squares are changed from the category prototype (see Figure 2). In this way as many exemplars as were desired could be created. We used a total of 128 chequerboard exemplars from each of the four categories in these experiments, though not all of these stimuli would be used for a given participant. The stimuli used in the experimental phases (categorisation and test) were counterbalanced across subjects.

Participants were required to separate these chequerboard stimuli into three categories (A, B and C) during the training and test phases (see Figure 2). In the training phase, 64 novel exemplars from each of the three categories were presented to participants in a randomized order. In the test phase, 10 of these previously seen exemplars from each category were presented to participants along with 10 novel chequerboard stimuli from each of the three categories and the three previously unseen category prototypes. The prototype stimuli were presented twice each during test. Participants made category responses to stimuli using the “C”, “V” and “B” keys on a keypad.

Figure 2 about here please

tDCS: This was delivered by a battery driven constant current stimulator (Neuroconn) using two electrodes covered by 5cm x 7cm pieces of pre-dampened synthetic sponge. One electrode montage was used: the first electrode (to which polarity refers) was placed over the left PFC (Fp3) and the reference electrode was placed on the forehead above the right eye. First electrode placement was determined by locating the Cz for each of the subjects (half the distance between the inion and nasion areas) and then moving 7 cm anterior relative to the Cz and 9cm to its left (see Figure 3).

Figure 3 about here please

Current was applied 1.5 min before the participants began the categorisation task (whilst listening to instructions) and from then on making 10 min stimulation in total. tDCS was delivered with an intensity of 1.5mA, and a fade-in and fade-out of 5 sec for the Anodal and Cathodal groups. Sham received the same 5 sec fade-in and fade-out, but only 30 sec stimulation between them, which terminated before categorisation commenced. A double blind procedure was used, by having two experimenters, one (primary) who actually ran the participant, and another (secondary) who set up the stimulation according to specifications provided by a third party. The connections to the stimulator were concealed by the secondary experimenter so that neither primary experimenter nor participant could determine the polarity of stimulation. In Experiment 1 we compared Anodal, Cathodal and Sham groups.

Design and procedure: In a between-subjects design the 48 participants were randomly assigned to one of 3 conditions: anodal stimulation, cathodal stimulation, and sham. Thus, all conditions contained 16 participants.

Once participants had been set up for tDCS stimulation they were informed that they would see different black and white chequerboard stimuli that they had to categorise into category A, B or C, and were shown the three buttons on the keyboard that they were to use ('C', 'V' and 'B' respectively). After the tDCS stimulator was switched on, the participant then read through three screens of more detailed instructions about the task, which lasted approximately 1.5 minutes. The training phase then began which contained 192 novel category stimuli presented in three blocks of 64 randomized trials with self-paced breaks separating each block. After a fixation cross, one stimulus was presented for 3 seconds during which the participant made their category response on the keyboard. The stimuli remained on the screen for the full 3 seconds. Feedback was presented after every trial.

After the participant finished the training phase, the primary experimenter switched off the tDCS stimulator and informed participants that no current was now going through the electrodes. They were then informed that there was a final block to the task, using the same categories as before, but this time with no feedback. This test phase had 66 trials of randomised exemplar and prototype stimuli.

Results

The crucial dependent variable was mean accuracy proportion (out of 1) of category responding during the test phase of the experiment (Figure 4).

Figure 4 about here please

To examine the prototype effect, the difference between accuracy in responding to exemplar and prototype stimuli during test was investigated. The average accuracy in responding to exemplars during test was calculated for each participant, and the mean accuracy of responding to these exemplars was then subtracted from the accuracy in responding to the prototypes during test (Figure 5). This difference was then added entered into a univariate analysis as a dependent variable with condition as the fixed factor. The main effect of condition on this measure of the prototype effect approached significance (p =. 081). There was no significant difference when comparing cathodal stimulation to Sham. However, when comparing the prototype effect in the anodal stimulation condition to the Sham control group there was a significant difference (p = .03) indicating that the prototype effect was smaller in the Anodal group. Comparing the prototype effect under anodal stimulation to the cathodal stimulation condition there was also a similar significant difference between conditions (p = .043), i.e. a greater prototype effect under cathodal stimulation compared with anodal stimulation. It is the lower accuracy on prototype trials in the anodal condition that seems to be driving these results.

Figure 5 about here please

Differences between exemplar and prototype response accuracy were also compared with the null hypothesis of a difference of zero between the two measures. There was no reliable difference found in the Anodal condition, however, Cathodal and Sham conditions both produced significant effects on this test (p < .05) indicating a significant prototype effect for these conditions (see Figure 5).

Discussion

Our results are broadly in line with those of Ambrus et al (2011), in that we have also shown that anodal stimulation at Fp3 leads to a significant reduction in, perhaps even elimination of, the prototype effect. Whilst accuracy scores are significantly higher for the prototype than for exemplars under Cathodal and Sham stimulation, this difference disappears under anodal stimulation and the difference between these differences (i.e. prototype effect for Anodal vs. prototype effect for Cathodal or for Shams) is also significant. Ambrus *et al* (2011) also found that anodal stimulation to left DPLFC eliminated a prototype effect that was otherwise significant in Sham controls, though in their case this was accompanied by significantly lower performance to prototypes in the Anodal condition relative to Shams as well, a result that is not significant in our data though the numerical trend is the same. Our results do allow us to extend Ambrus *et al's* conclusions, however, as we have been able to show that

cathodal stimulation is not different to sham stimulation with our procedures (Ambrus *et al* did not run a left DPLFC cathodal group). Thus, our effect is a selective one, in that only anodal stimulation of left DPLFC eliminated the prototype effect in our experiment.

We will forgo further analysis of this result until we have reported the results of Experiment 2 which also investigates the effects of tDCS to left DLPFC, but this time using a version of our categorisation task that is identical to that used in our earlier perceptual learning experiments (Civile *et al*, 2014).

Experiment 2

Here we carry out two replications of an experiment that exactly duplicates the categorisation training procedure adopted by McLaren (1997) and also used by Civile *et al* (2014). This was done in order that our results could be extrapolated to these perceptual learning experiments, allowing us to predict the consequences tDCS for perceptual learning in future experiments. In this procedure only two chequerboards are used as base patterns or prototypes (i.e. there are only two categories in play), and exemplars are generated from them as before by adding noise, which simply involves changing a random selection of the squares in the prototype. Participants are then trained to distinguish between exemplars drawn from these two categories using a trial and error procedure with feedback before being tested for classification accuracy to both category exemplars and their prototypes (which, as in Experiment 1, are never seen in training). Experiment 2a uses this paradigm and contrasts anodal tDCS to Fp3 in the Experimental group with a Sham control. Experiment 2b uses a cathodal stimulation group as the comparison with the Experimental group receiving anodal stimulation. The cathodal control has the advantage that stimulation occurs in exactly the same way as for anodal stimulation (but with reversed polarity). We took this opportunity to see if it would produce similar results to sham stimulation.

Method

Stimuli: These were as before but only two prototype-defined categories were used (A and B in Figure 2).

Participants: Experiments 2a and 2b each had 16 undergraduate participants per group and were run in Shanghai, China, at East China Normal University.

tDCS: Stimulation was as in Experiment1. In Experiment 2a we compared Anodal and Sham groups. In Experiment 2b we compared Anodal and Cathodal groups.

Categorisation task: Participants were asked to categorise chequerboards into two different categories (in this case A and C, see Figure 2). Chequerboards were presented one at a time for classification. They were presented for 4 seconds. Participants had to press either the "x" or the "." key to categorise the stimulus. The experiment moved to the next stimulus only after the 4 seconds had passed. Participants received feedback as to whether their response was correct or not. 128 Exemplars were presented, 64 from category A and 64 from category C. In the test phase participants were asked to categorise chequerboards (self-paced) without feedback. They were given one presentation of eight old exemplars from each category (exemplars used in training), eight new exemplars from each category, and two presentations of both category prototypes.

Results - Experiment 2a

Figure 6 gives graphs of mean accuracy for Experiment 2a. A strong prototype effect was obtained under anodal tDCS, but was absent in the Sham group; $p<.05$ for comparisons between the prototype and mean performance on the exemplars in the anodal condition. The interaction for these effects with group (Anodal vs. Sham) did not, however, reach significance ($p=.15$). There is some evidence that the effect of anodal tDCS was to suppress performance to the exemplars, in that there was a significant difference between Anodal and Sham groups for the New exemplars, $p=.042$. Clearly, given our earlier results and those of Ambrus et al (2011), this set of data came as something of a surprise. Further interpretation of this result will be postponed, however, until we have considered the results of Experiment 2b.

Figure 6 about here please

Results - Experiment 2b

Figure 7 gives the graphs for Experiment 2b. Once again a prototype effect was obtained under anodal tDCS, $p=.005$, but not under cathodal tDCS, which gave results very similar to those obtained in the Sham group of Experiment 2a. There was some evidence that the Anodal group prototype effect was significantly stronger than that in the Cathodal group, $p=.078$ for the interaction using the average of the two types of exemplar to compare to the prototype. There is also evidence that anodal tDCS suppresses test performance to exemplars, as there is a significant Group difference, this time for Old exemplars, $p=.037$.

Figure 7 about here please

Discussion

Taken together, the results of Experiment 2 suggest that anodal tDCS reduces accuracy on test to exemplars in this type of categorisation task. It leaves performance to prototypes relatively unaffected, however, which leads to the emergence of a prototype effect when we compare performance on the prototype to that on other exemplars. Before accepting these conclusions, however, we acknowledge that there is an obvious issue with these results that makes their interpretation more difficult. Performance in the Sham or Cathodal groups is near ceiling, particularly for the prototypes. This makes it hard to tell whether the absence of any prototype effect in these groups is real - or is due to this ceiling effect. If it is the latter, then it may be that anodal tDCS simply reduces test accuracy below ceiling, allowing a prototype effect that was, in some sense, always there to emerge. Another possibility, however, is that anodal tDCS selectively enhances the prototype effect in these experiments, and that it's appearance is not a simple consequence of an overall reduction in performance allowing an effect that was present but masked to become visible. We will focus on this last possibility in what follows, as we have been unable to generate a plausible account of how anodal tDCS could reduce overall performance in the two category case, but selectively reduce performance to prototypes in the three category problem.

On the face of it, the results of Experiment 1 and Experiment 2 appear to be incompatible. In Experiment 2, as we have just seen, we have evidence for anodal tDCS using our electrode montage producing a stronger prototype effect than that shown in our control groups (using either Sham or Cathodal stimulation). In Experiment 1 we obtained the converse pattern of results, the prototype effect in the Anodal group was this time significantly weaker (and actually absent) than in either Sham or Cathodal groups. It is true that because of the nature of the problems there are some parametric differences in stimulation between the two experiments. tDCS stimulation will have been active for about half of the training phase in Experiment 1, but the full training phase in Experiment 2. But this, on it's own, would seem an unlikely candidate to explain the opposite effects of the two experiments, and in any case the effects of tDCS stimulation are thought to last well beyond the active stimulation period. So how are we to explain this pattern of results?

We believe that the key to understanding this pattern lies first of all with the prototype effect (or lack of it) demonstrated in the control conditions where tDCS can be assumed to not have any

significant influence. In the two category problem used for Experiment 2 there was no prototype effect in these control groups. In the three category problem used in Experiment 1 there was a significant prototype effect in both control groups. The stimuli and procedures in both experiments are the same, with the proviso that we used an extra category in Experiment 1, so this difference (no prototype effect vs. prototype effect) can most probably be ascribed to the use of three rather than two categories. This would have the effect of influencing performance levels not only because there are three possible choices instead of two, but also because the amount of generalisation between categories has increased (because now each test stimulus in the three category problem would be receiving generalisation from exemplars of two different categories in addition to members of its own category, rather than from just one).

This extra generalisation between categories would also, somewhat paradoxically, produce a stronger perceptual learning effect for the three category problem than would be the case in the two category problem. The extra generalisation makes the perceptual discrimination between categories more difficult, but the perceptual learning effect addresses this issue, by enabling the representations of the exemplars and prototypes from the three categories to become more distinct, and consequently there is more scope for this effect to manifest in these circumstances We will go into considerable detail on exactly how this might be achieved shortly, but our argument is that this stronger perceptual learning effect in the three category problem is particularly marked between categories, making them more easily distinguishable from one another and this enhances the prototype effect.

Our explanation of the results for the control conditions is thus based on a trade-off between generalisation between categories (which on its own reduces classification performance) and enhanced between-category perceptual learning, which we will argue assists classification of prototypes more than exemplars. In the two category problem the former effect dominates, and generalisation between categories is such that it counteracts any advantage that the prototype might have over other exemplars. In the three category problem the balance shifts, and now perceptual learning makes the categories more discriminable and the prototype effect emerges. We will show how this can happen shortly, but note that some explanation for this (reliable) difference between control conditions has to be given, and this is the most plausible account available to us.

Our explanation of the results in the anodal tDCS conditions is that this stimulation abolishes perceptual learning leaving enhanced generalisation, both between and within categories. The effect

of the enhanced generalisation within-category is to strengthen the prototype effect, but the effect of the between-category generalisation will be to reduce it. The first dominates in the two category problem, but the second is the more important factor in the three category problem because the amount of between category generalisation is doubled. Hence the prototype effect in the two category problem becomes detectable under anodal tDCS (and may be potentiated by a reduction in performance from ceiling - we cannot rule this out); but the prototype effect that was already detectable in the three category problem is reduced and becomes non-significant in the three category case.

The analysis thus far may seem rather ad-hoc and designed to describe rather than explain our data. Note, however, that there has to be some explanation for the otherwise rather counter-intuitive pattern of results obtained across Experiments 1 and 2, and that our explanation of the effects in the control groups follows from an application of the McLaren, Kaye and Mackintosh (1989) model of perceptual learning and categorisation and it's recent variants (McLaren and Mackintosh, 2000; McLaren, Forrest and McLaren, 2012) discussed in our introduction. Our hypothesis is that the modulation of salience based on the error term that forms a vital part of MKM model is disrupted by anodal tDCS so that the model in essence reverts to McClelland and Rumelhart's (1985) model of categorisation inasmuch as perceptual learning or representation development is concerned. This hypothesis is explored in detail in the computational analysis that follows.

Perceptual Learning and Categorisation under tDCS

The top middle panel of Figure 8 shows how the salience (activation) of the elements (representations of sets of features) of each category prototype will be affected by experience of exemplars from categories A and C if we adopt the MKM approach to salience modulation via prediction error. Note that all the elements needed to represent all three categories (A, B and C) are shown for completeness, but that exposure is only given to two of them, A and C, for this example. Those elements that are more predictable and are more often encountered will be those with lower salience (darker shading). Thus, the elements shared by the A and C prototypes (abc and ac) are less salient than a or ab elements (only present in A). Given that exemplars from the B category are not pre-exposed in this example the b elements can only occur by virtue of the random noise added to construct exemplars from the prototypes. The right panel shows how the modulation of salience

across elements changes when all three categories are experienced. In particular, the shared prototypical elements, abc, become even less salient. The effect is that discrimination between the three category prototypes is actually better than when only two categories were trained because perceptual learning is more effective.

Figure 8 about here please

The bottom panels of Figure 8 show what happens when the salience modulation mechanism in MKM is removed. The salience (activation) of elements representing the stimulus features now reverts to that in McClelland and Rumelhart's (1985) model of categorisation, with units receiving more internal input having higher (rather than lower as in MKM) activations. In effect, this gives the common elements an advantage that can be seen in both lower panels. They become increasingly salient, and this leads to very strong between and within-category generalisation. Table 1 gives the relative proportions of the different elements making up each stimulus for the average A exemplar and C exemplar as well as the A and C prototypes using a simple model that, as a first approximation, equates each square in a chequerboard with a feature. By combining this information with the expected salience of these elements shown in Figure 8, it is possible to get a sense of how much one stimulus will generalise to another as a result of categorisation training.

Table 1 about here please

We can see immediately that the MKM model predicts that exemplars will contain novel (noise) elements that are of relatively high salience, and that the prototypical elements will be more numerous, but less salient. The prototypes are exclusively composed of relatively low salience prototypical elements and do not overlap as much as exemplars drawn from the two categories. The consequence of this is that generalisation from say the trained A exemplars to C exemplars will be somewhat greater than to the C prototype. This effect is symmetrical (the C exemplars generalise to the A exemplars to the same extent), and so the chance of mistakenly calling an A exemplar a member of the C category will be somewhat greater than that of calling the A prototype a member of the C category.

Table 2 gives the calculated expected generalisation (based on Figure 8 and Table 1) to/from trained A exemplars to each of the four stimulus types considered in our earlier table, and it confirms our analysis. If we begin by looking at the 2 Categories MKM column of the table, it shows (perhaps rather surprisingly) that the generalisation from one of the trained A exemplars to this typical A

exemplar (.609) will be greater than that stimulus' generalisation from (or to) the A prototype (.596), but the difference between these values is not large. We can estimate the generalisation that occurs on average from the C category exemplars and the C prototype to/from this A exemplar by looking at the C prototype and C exemplar rows of the table. These give generalisation from an A exemplar to these stimuli, but by symmetry they give us the values we will require for our calculations. Thus, the generalisation from C exemplars to an A exemplar (.430) will be considerably greater than the generalisation from the C prototype to A exemplars (.340), which is also the value for generalisation from C exemplars to the A prototype. The result is a larger difference in generalisation for the A prototype to the A category exemplars compared to the C category exemplars (.596 - .340 = .256) than for the A exemplars to the same stimuli (.609 - .430 = .179). In other words, it predicts a prototype advantage, but does it predict a detectable prototype effect?

Table 2 about here please

To answer this question we need to convert generalisation into choice. The models themselves do not stipulate the requisite decision mechanisms to function as stand-alone classifiers. Hence we used a minimalistic approach to converting generalisation into choice behaviour that was simply designed to demonstrate that the MKM model could produce the correct pattern for the two and three category problems in the control groups, and that this would then change appropriately when error modulation of salience was disrupted. We employed a standard form of Luce's choice rule, using the exponential of the generalisation coefficient as our measure of category membership.

1. $P(A) = \frac{e^{ka}}{e^{ka} + e^{kc}}$

Where P(A) is the probability of classifying a stimulus as a member of category A, a is the summed generalisation to that stimulus from trained A exemplars, c is the summed generalisation from trained C exemplars, and k is a constant that captures the weight given to generalisation in a given task. We then needed to find k for our model. For the 2 Categories MKM coefficients we simply chose k so that it gave a ballpark fit to the accuracy data for the exemplars in our experiments (and we used this procedure for the other data as well). We adopted a value of 11, which resulted in P(A) for the prototypes being 0.94, and P(A) for exemplars being 0.88. These are a reasonable fit to the actual values across the two experiments, which are 0.94 and 0.925 respectively, though clearly the model value for the exemplars is a little low.

One point to make here about this very simple model is that we are simply assuming that each square in a chequerboard is a feature. This may be a useful approximation to reality for our purposes, but it completely fails to capture the fact that the prototypes (which were constrained to have regions of nearly all black or nearly all white) looked distinctly different to the exemplars, even those from their own category, which were necessarily less "blocky" in appearance because of the random noise used to generate them (see Figure 2). This would act to reduce the magnitude of any prototype effect in these experiments, and so our model is necessarily overestimating the size of the prototype effect actually obtained. Even given this, however, we can see that a prototype effect might be hard to detect for the two category case under our control conditions.

We can now look at the expected generalisation for the three category problem. This is shown in the 3 Categories MKM column, and gives a difference of .222 (= .395 - .173) for the prototype and .104 (= .526 - .422) for the exemplar. Clearly this is a larger disparity between prototype and exemplar generalisation (.118) than we had for the two category case (.077 where the values for the prototype and exemplars were .256 and .179), and as such could lead to a stronger prototype effect. We took k for the three category task to not necessarily be the same as in the two category task, and arrived at a value of 10. Clearly training on three categories rather than two might, in itself, affect the weight placed on the measure provided by generalisation (not least because as the number of categories increases so does total generalisation between them), but note that using the same value for k as in the two category case (i.e.11) leads to essentially the same pattern of results with this simple model of choice. The choice equation now becomes:

2. $$P(A) = \frac{e^{ka}}{e^{ka} + e^{kb} + e^{kc}}$$

This resulted in P(A) for the prototypes being 0.82 (which is somewhat too high), and P(A) for exemplars being 0.59 (which is too low), but represents a reasonable fit to the data and clearly makes the point that the prototype effect for the three category problem is predicted to be much greater than for the two category problem (a difference of 0.82 - 0.59 = 0.23 compared to a difference of 0.06 in the two category problem). It is no surprise on this analysis, then, that the prototype effect might be detectable in our controls for the three category, but not the two category problem.

If we now consider the effect of turning off error-based modulation of salience to give something like the representation development that would be seen using the M&R model, then a quite different pattern emerges. First of all, generalisation increases a great deal – as can be seen by

looking in the two M&R columns of Table 3. This is exactly as would be expected given that perceptual learning (which has effectively been switched off) has the opposite effect to generalisation. The increased generalisation for the two category problem gives difference scores of 1 - .654 = .346 for the prototype and .754 - .558 = .196 for exemplars. The difference score for the prototype has improved relative to the .256 difference obtained using MKM, whereas the score for the exemplars has stayed about the same (it was .179). The prediction, then, is that the prototype effect should be enhanced by anodal tDCS in the two category case, as the disparity between prototype and exemplar difference scores is now .150 instead of the original .077. Translating the generalisation scores into choice probabilities requires that we make a new estimate of k here, as clearly tDCS could quite possibly have affected the weight placed on our measure of category membership in ways not captured by our model. A value of k = 8 gives us P(A) for the prototype as 0.94 and P(A) for exemplars as 0.82 in the two category problem, which is a good fit to our data and suggests that the size of the predicted effect has doubled. If we instead consider what happens for the three category problem then a different effect emerges. The original disparity between the generalisation differences for MKM was .118 (.222 - .104), but once we turn off error-based modulation it becomes .085 (.328 - .243). Clearly both generalisation scores have increased, but the increase has been greater for the exemplars and so the difference is smaller, and smaller still relative to the scores contributing to that difference. Translating these scores into choice probabilities we used a value for k of 5 to try and fit our data as best we could, which results in choice probabilities of 0.72 (too high) for the prototype and 0.63 (too low) for the exemplars. Clearly, this simple model of choice had considerable difficulty in fitting our data. But this exercise makes the important point that once again the changes in generalisation - which are all we are confident of in this modelling exercise (and even here we have caveats about the similarity of our prototypes to the exemplars) - do translate into changes in choice probability which fit the interaction in our data. In this case the predicted prototype effect, which was 23%, has now decreased to 9% indicating that it should become considerably more difficult to detect.

One point that may strike the reader about our analysis is that this final prototype effect for the three category problem under tDCS (an effect of 9%) is not so different to the effect of 12% predicted for the two category problem under tDCS which we wish to claim is detectable. The important points to make here are that first, the two effects occur at different levels of choice probability. A 12% difference when choice is in the 80%-90% range will have a lower variability

associated with it than a difference of 9% when choice probabilities are 60%-70%. Thus one may be detectable where the other is not. Second, we have tried to emphasize that it is the change in effect from control stimulation to experimental (anodal) stimulation that is the real prediction of interest here. We cannot (and do not wish to) lay claim to possessing a model that fits (in the statistical sense) our data, but we can claim that the changes in generalisation that occur in our model as a result of shifting from two category to three category problems, and as a result of switching off perceptual learning, accurately capture the pattern in our data. And this suggests a particular interpretation for the effects of anodal tDCS stimulation of DLPFC.

The real test of our position, of course, would be to look directly at the effects of anodal tDCS stimulation on perceptual learning. We are now able to unequivocally predict that this stimulation should disrupt perceptual learning and possibly even reverse it. We will now briefly consider a set of experiments that addresses this issue, using the same set of chequerboard stimuli and the design employed by Civile et al (2014) to look at perceptual learning in the context of inversion effects.

Perceptual Learning

We have already noted that perceptual learning affects the way we see the world and the objects in it, and that pre-exposure to stimuli enhances our ability to discriminate among or between them or other similar stimuli. In the lab, one of the most striking consequences of perceptual learning is the face inversion effect: upright faces are better recognised than inverted faces. This inversion effect is at least partly due to our extensive experience with faces, as exposure to artificial stimulus sets that have a structure akin to that possessed by faces leads to phenomena similar to those observed in face recognition, including inversion effects (McLaren, 1997; Gauthier and Tarr, 1997). For example, exposure to a set of prototype-defined chequerboards results in an inversion effect for exemplars from a familiar category but not for exemplars from a novel (not pre-exposed) category (McLaren, 1997, Civile et al, 2014). As we have already argued, this advantage for upright exemplars can be explained by associative models of perceptual learning that rely on differential latent inhibition of common elements. Exposure to exemplars from the familiar category leads to latent inhibition of the prototypical elements for that category (Figure 1, top half). When an exemplar drawn from that category is encountered, the elements that it shares with the prototype will be latently inhibited (making them less salient), whereas the elements that are unique to that exemplar will not suffer

greatly from latent inhibition (making them more salient). This will enhance discrimination between exemplars drawn from the familiar category (i.e. perceptual learning). Our associative model can explain a range of perceptual learning phenomena, including the inversion effect, as the latent inhibition mechanism only applies to what has been experienced, and participants have not experienced inverted exemplars during the earlier familiarization phase. Figure 1 (bottom half) also shows that losing the modulatory component producing differential latent inhibition should result in a loss and perhaps even a reversal of within-category perceptual learning.

Our plan, then, was to run what was essentially a replication of Civile et al (2014), but to apply anodal tDCS using our current electrode montage during the first, categorisation phase. This should disrupt any perceptual learning and increase generalisation between exemplars. Because perceptual learning is responsible for the inversion effect for exemplars drawn from a familiar category (i.e. one that has been trained) that we reliably see with this procedure, anodal tDCS should reduce (perhaps even reverse) this effect. The detailed results of these experiments will be reported elsewhere (Civile, Verbruggen, McLaren, Zhao, Ku and McLaren, in preparation) so we will only summarise them here. We used the same 16x16 chequerboards used in the earlier experiments with the addition of one extra category, D. Our experimental groups used anodal stimulation. The control groups used sham or cathodal stimulation. Participants classified exemplars from two prototype-defined chequerboard categories during tDCS (categorisation stage). They then studied exemplars drawn both from one of the now familiar categories and from another novel category in either upright or inverted orientations (study stage). Finally, in the recognition task (test stage) they had to classify chequerboards as either “old” (seen in the study phase) or “new” (not seen). Their accuracy scores were then converted into d' measures for use in our analyses.

In Figure 9 we give the combined Anodal stimulation vs. Control results for recognition in this final phase. As predicted by extrapolation from the Civile et al (2014) experiments and the results considered earlier, in the Control conditions we observed an inversion effect for familiar-category exemplars (Upright better than Inverted, $p = .013$) but not for novel-category exemplars. The perceptual learning effect was also reflected in the performance on upright exemplars taken from the familiar category being better ($p=.050$) than that on the matched exemplars (matched across participants) taken from the novel category. But under anodal stimulation the pattern was quite different. There was no inversion effect, and the effect that was there was significantly different to that

in controls (p=.045). Now the performance on upright exemplars taken from the familiar category was significantly worse (p=.005) than that on the matched exemplars taken from the novel category. In fact, if we compare performance on the upright exemplars taken from the familiar category in both conditions, the difference is also highly significant (p=.005), and in favour of the controls. Clearly the effects of familiarisation with the category have radically altered under anodal stimulation.

Figure 9 about here please

The most reasonable interpretation of these results is that Anodal tDCS has a selective effect on performance to upright exemplars drawn from the familiar category. We would argue that our data are consistent with anodal tDCS stimulation eliminating a modulatory input based on prediction error, leading to a loss of perceptual learning. The resultant system is then adequately described by simple delta rule algorithms of the type found in the M&R model of categorisation, and as such has a particular problem in dealing with familiar prototype-defined categories as a consequence of the increased generalisation between their exemplars. This leads to the poor performance on the upright exemplars drawn from the familiar category under anodal tDCS, compared to the otherwise superior performance exhibited to these exemplars under control conditions as predicted by MKM.

Conclusions

In Experiment 1 we were able to demonstrate that anodal tDCS to left DLPFC does indeed reduce the prototype effect that might otherwise be obtained after learning to categorise. This confirms the result of Ambrus *et al* (2011), and suggests that their result was not simply a matter of anodal tDCS reducing learning *per se*. We carried out Experiment 2 in order to set the stage for our subsequent investigation of perceptual learning and to confirm the results of Experiment 1. Our results were, on the face of it, anything but confirmation of Experiment1, in that far from reducing the prototype effect, anodal tDCS enhanced it. In fact, it produced a significant effect where under control conditions none had been detectable.

This initially surprising and contradictory result proved to be susceptible to a detailed analysis in terms of changes in generalisation brought about by 1) changing the number of categories from three to two and 2) using either anodal or control stimulation. The analysis relied on the assumption that the effect of anodal tDCS to left DLPFC was to disrupt modulation of the salience of stimulus representations based on error such as to transform a system for categorisation that under control conditions could be described by MKM, to one better thought of in terms of M&R. This effect

interacted with the increased generalisation between categories that occurred in the three category problem relative to the two category version, and so explained the different effects of anodal tDCS on the prototype effect in the different experiments. Our analysis is model-driven, and admittedly post-hoc, but it did make the prediction that perceptual learning due to pre-exposure during categorisation training should be eliminated, or even reversed, by anodal stimulation.

This prediction was fully borne out by the results of the final set of experiments reported here. Our control conditions showed our usual inversion effect in this analogue of the face recognition paradigm using chequerboards, but the inversion effect was not present under anodal tDCS. More importantly, whilst familiarisation with a category improved performance on upright exemplars drawn from that category, in that they were better discriminated in the old/new test than those drawn from a novel category, this effect was reversed under anodal stimulation. Finally, performance on the upright exemplars from the familiar category was significantly and selectively worse under anodal tDCS than in the control conditions, an effect entirely consistent with our hypothesis that anodal stimulation "turns off" perceptual learning and leaves participants with greatly increased generalisation.

Final Thoughts

The McLaren, Kaye and Mackintosh theory of latent inhibition and perceptual learning could, up until now, be seen as an abstract connectionist model of representation development that provided a good account of a fairly limited domain of animal and human behaviour. But the intention on the original author's part was always to apply it more widely, and that is a challenge that we have taken up with our recent research into categorisation and perceptual learning. We now have some hints about the neural mechanisms underlying perceptual learning, and they fit very well within the framework provided by that theory. This serves to remind us that Nick Mackintosh's vision in extrapolating from sophisticated behavioural experiments to detailed theoretical mechanisms was quite extraordinary, and his theoretical insights into perceptual learning in humans are as relevant today as they were over twenty-five years ago.

Acknowledgements

IPL McLaren and Frederick Verbruggen were supported by a grant from the ESRC (ES/J00815X/1), and Frederick Verbruggen is supported by a starting grant from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ ERC Grant Agreement No. 312445. Kathryn Carpenter was supported by an ESRC studentship (ES/J50015X/1), Ciro Civile was supported by an Overseas Scholarship from the International Office at the University of Exeter, and Yixuan Ku by the National Key Fundamental Research (973) Program (2013CB329501) of China.

References

Ambrus G. G., Zimmer M., Kincses Z. T., Harza I., Kovacs G., Paulus W., and Antal, A. (2011). The enhancement of cortical excitability over the DLPFC before and during training impairs categorisation in the prototype distortion task. *Neuropsychologia 49*, 1974–1980.

Civile, C., Zhao, D., Ku, Y., Elchlepp, H., Lavric, A., and McLaren, I.P.L. (2014). Perceptual learning and inversion effects: Recognition of prototype-defined familiar chequerboards. *Journal of Experimental Psychology: Animal Behavior Processes*, 40,144-61.

Civile, C., Verbruggen, F., McLaren, R., Zhao, D., Ku, Y. and McLaren, I.P.L. (in preparation). Switching off perceptual learning: tDCS to left DLPFC eliminates perceptual learning in humans.

Fregni, F., Boggio, P. S., Nitsche, M. A., Bermpohl, F., Antal, A., Feredoes, E. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research, 166*, 23–30.

Gauthier I. & Tarr M.G. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research, 12, 1673-1682.*

Homa, D., Sterling, S. and Trepel, L. (1981). Limitations of Exemplar-Based Generalisation and the Abstraction of Categorical Information. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 418-39.

Kincses T. Z., Antal A., Nitsche M. A., Bártfai O., Paulus W. (2003). Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the human. *Neuropsychologia*, 42, 113–117.

Lupyan G., Mirman D., Hamilton R., Thompson-Schill, S.L. (2012) Categorisation is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition 124*, 36–49.

McClelland, J.L. and Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159-197.

McLaren, I.P.L. (1997). Categorisation and perceptual learning: An analogue of the face inversion effect. *The Quarterly Journal of Experimental Psychology* 50A, 257-273.

McLaren, I.P.L., Leevers, H.L. & Mackintosh, N.J. (1994). Recognition, categorisation and perceptual learning. In C. Umilta & M. Moscovitch (Eds.) *Attention & Performance XV*.

McLaren, I.P.L. and Mackintosh, N.J. (2000). An elemental model of associative learning: Latent inhibition and perceptual learning. *Animal Learning and Behavior*, 38, 211-246.

McLaren, I.P.L., Forrest, C.L., McLaren, R.P. (2012). Elemental representation and configural mappings: combining elemental and configural theories of associative learning. *Learning and Behavior*, *40*, 320-333.

McLaren, I.P.L., Kaye, H. & Mackintosh, N.J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. Oxford university press.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353-363.

Reinhart, R.M.G. and Woodman, G.F. (2014). Causal Control of Medial-Frontal Cortex Governs Electrophysiological and Behavioral Indices of Performance Monitoring and Learning. *Journal of Neuroscience*, 34, 4214-27.

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: Appleton-Century-Crofts.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141-145.

Figure and Table Legends:

Table 1: This shows the percentage of features of different types (elements) present in each of four different stimuli drawn from two different categories. Each element label refers to Figure 7 (middle and rightmost panels) and denotes features that are present in one or more of the three possible categories. The table also makes clear the extent of feature overlap between any two stimuli once it is born in mind that exemplars are generated from the prototypes by randomly changing the elements of the prototype, and that 25% is the maximum allocation of elements of any one type.

Table 2: This shows the expected generalisation (minimum=0, maximum=1) between each of the four stimuli in the table and a typical trained exemplar drawn from Category A. Generalisation is calculated using either the MKM or M&R salience for the elements comprising the stimulus. Note that the generalisation from A exemplars to the C prototype will be the same as that expected for the C exemplars to the A prototype, and so can be used in conjunction with the figures for the A prototype to calculate the probability of labelling prototype A as a member of the A category.

Figure 1: Top half. This illustrates how modulation driven by prediction error (as in MKM) can be used to influence feature salience for a single category. The result, shown in the temperature diagram, is that stimulus features that are more predictable become less active (darker shading) leading to latent inhibition (slower learning as a consequence of pre-exposure). This improves discrimination between members of a prototype-defined category as it relies upon the less predictable features unique to each stimulus. The bottom half of the figure shows how disrupting this modulatory input (e.g. using tDCS) reverses this effect (as in M&R), making the common, prototypical features of the stimuli the most salient (lighter shading).

Figure 2: Examples of the prototypes (top row) and exemplars (bottom row) from the categories used in the experiments reported in this paper. Please see the text and Civile, Zhao, Ku, Elchlepp, Lavric and McLaren (2014) for more details about the characteristics of our prototype-defined categories of chequerboards.

Figure 3: The figure illustrates the electrode configuration and the tDCS apparatus used in these experiments.

Figure 4: Average accuracy for each group (Anodal, Cathodal, Sham) broken down to show overall performance during training and then test performance to exemplars and prototypes. Error bars are SE of the mean.

Figure 5. The difference in response accuracy between prototypes and the average of responding to exemplars in the test phase of the experiment. Error bars are SE of the mean.

Figure 6: The graph shows mean accuracy during test for old and new exemplars drawn from the trained categories as well as performance on the prototypes for those categories. The chequerboards shown are typical exemplars / the prototype for the A category, but the average is for both categories. Error bars show SE of the mean.

Figure 7: The graph gives mean accuracy for old and new exemplars as well as the prototypes during test based on performance on both categories. This time the figure displays typical exemplars / the prototype for the C category.

Figure 8: Top Panels: This illustrates how modulation driven by prediction error (as in MKM) can be used to influence feature salience. The result, shown in the temperature diagram, is that stimulus features that are more predictable become less active (darker shading) leading to latent inhibition (slower learning as a consequence of pre-exposure). This improves discrimination between members of a prototype-defined category as it relies upon the less predictable features unique to each stimulus. The bottom panels show how removing this modulatory input (as in M&R) reverses this effect, making the prototypical features of the stimuli the most salient (lighter shading). The two category case (centre) where exposure is only to A and C categories, and the three category case (right) are illustrated in terms of the prototypes for each category, and are labelled to show the differential effect on the elements that make up each category prototype.

Figure 9: Combined results of perceptual learning experiments. Lighter bars are for Anodal stimulation and darker bars for control stimulation. The y-axis gives d' scores for the old/new recognition task (higher=better, 0=chance), and the four different stimulus conditions are shown on the x-axis.

Table 1

Stimulus	Elements	a	ab	ac	abc	bc	b	c	n
A prototype	%	25	25	25	25	0	0	0	0
A exemplar	%	20	20	20	20	5	5	5	5
C prototype	%	0	0	25	25	25	0	25	0
C exemplar	%	5	5	20	20	20	5	20	5

Table 2

Stimulus	2 Categories MKM	2 Categories M&R	3 Categories MKM	3 Categories M&R
A prototype	.596	1.00	.395	1.00
A exemplar	.609	.754	.526	.761
C prototype	.340	.654	.173	.672
C exemplar	.430	.558	.422	.518

Figure 1

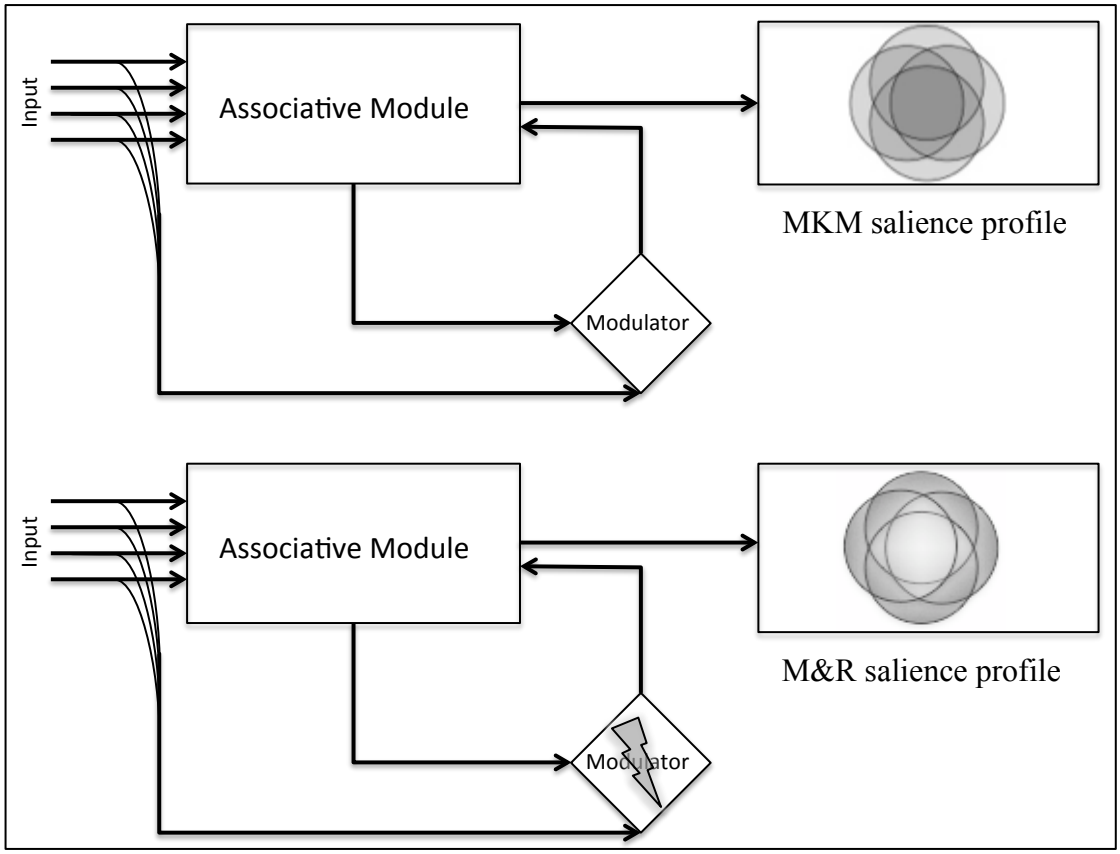
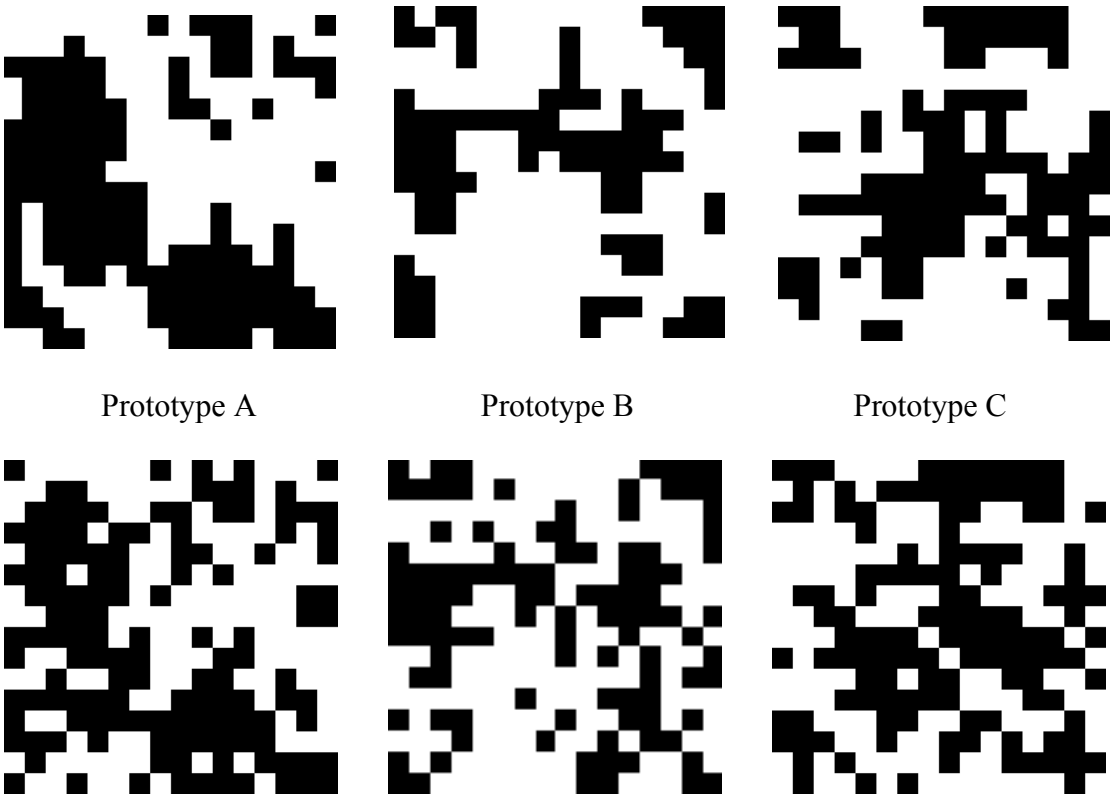

Input
Associative Module
MKM salience profile
Modulator
Input
Associative Module
M&R salience profile
Modulator

Figure 2



Prototype A | Prototype B | Prototype C

Typical A Exemplar | Typical B Exemplar | Typical C Exemplar

Figure 3

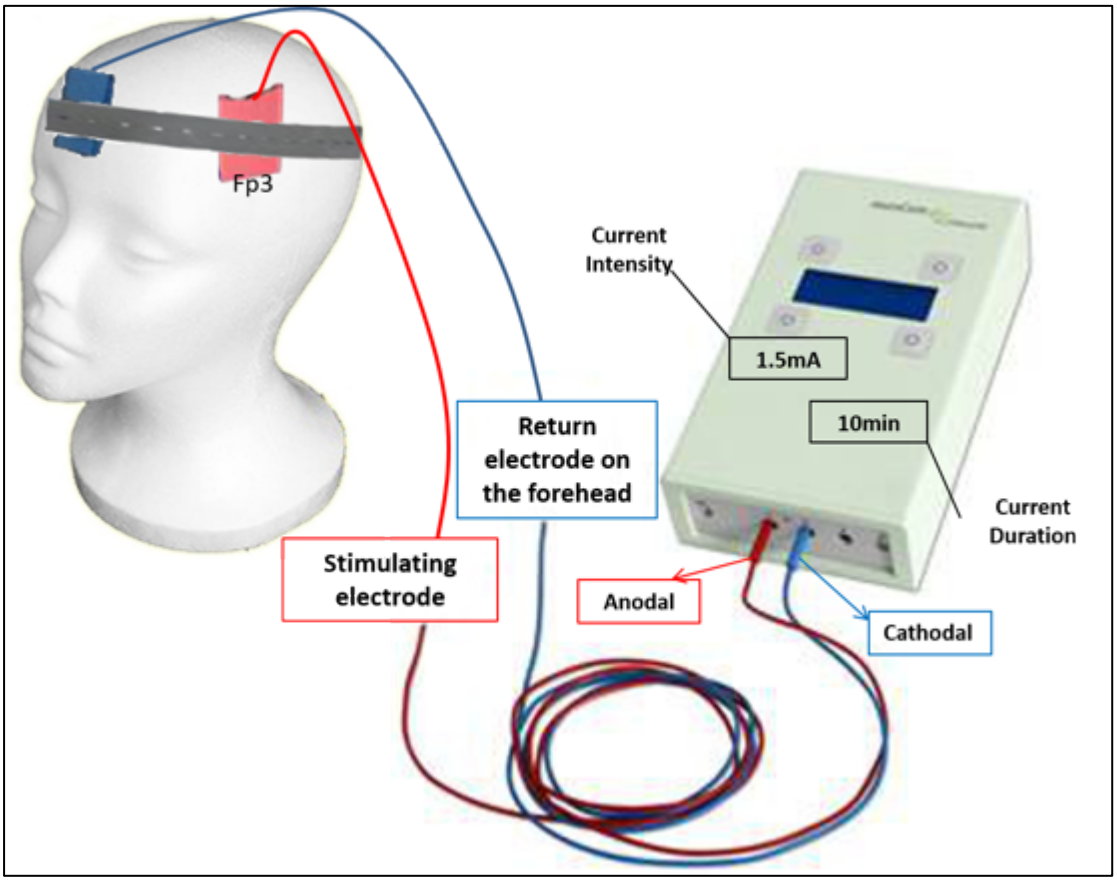

Fp3
Current
Intensity
1.5mA
10min
Current
Duration
Return
electrode on
the forehead
Stimulating
electrode
Anodal
Cathodal

Figure 4

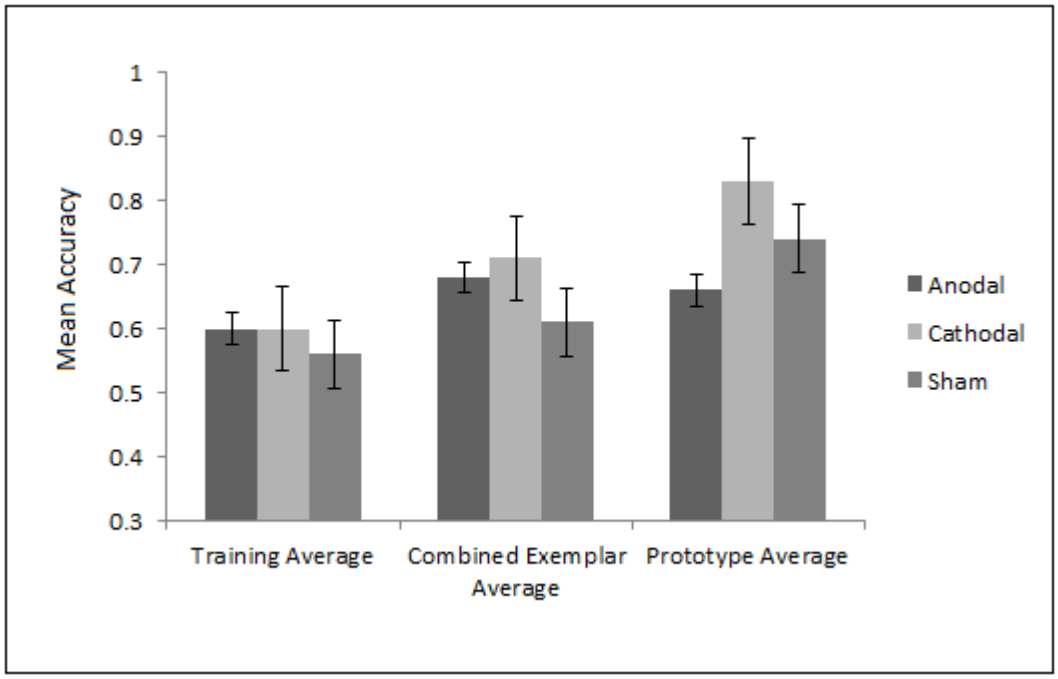

Mean Accuracy
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
Training Average
Combined Exemplar Average
Prototype Average
Anodal
Cathodal
Sham

Figure 5

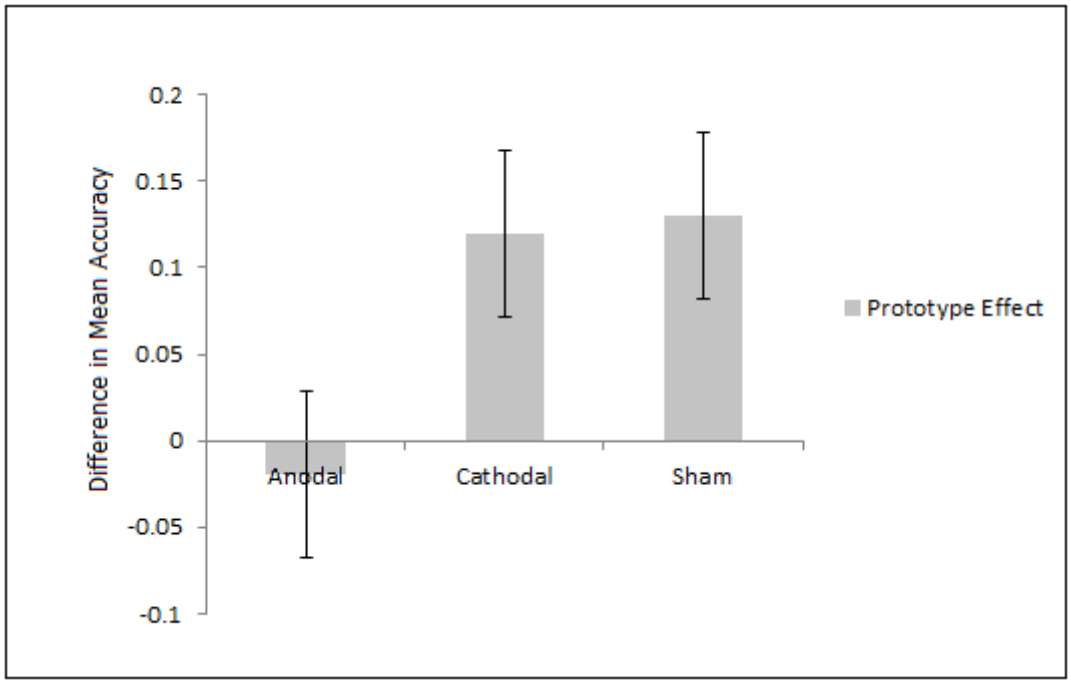

0.2
0.15
0.1
0.05
0
-0.05
-0.1
Difference in Mean Accuracy
Anodal
Cathodal
Sham
Prototype Effect

Figure 6

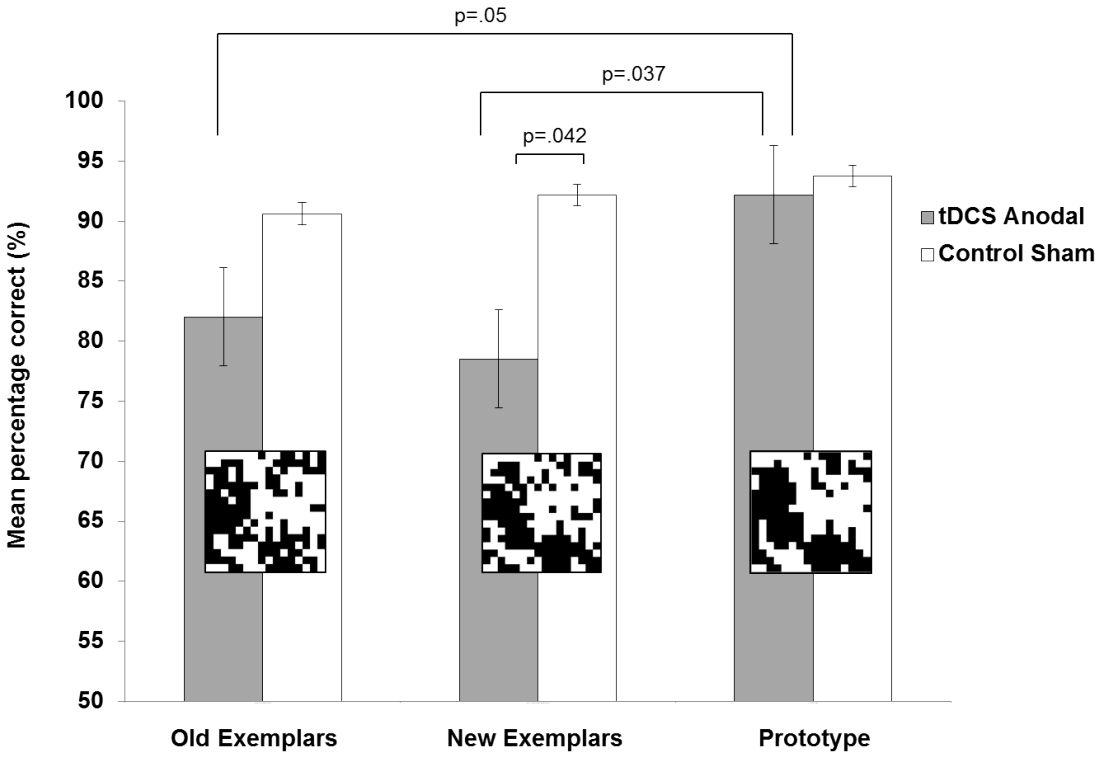

p=.05
p=.037
p=.042
100
95
90
85
80
75
70
65
60
55
50
Mean percentage correct (%)
tDCS Anodal
Control Sham
Old Exemplars
New Exemplars
Prototype

Figure 7

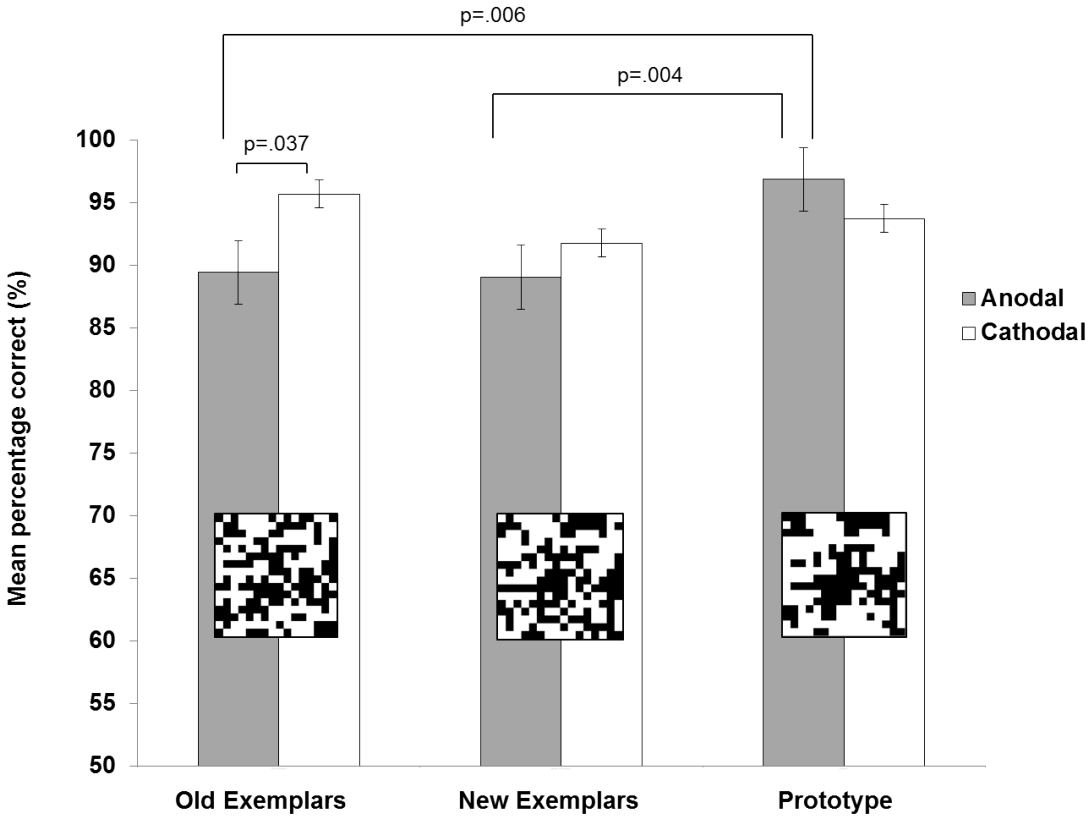

p=.006
p=.004
p=.037
100
95
90
85
80
75
70
65
60
55
50
Mean percentage correct (%)
Anodal
Cathodal
Old Exemplars
New Exemplars
Prototype

Figure 8

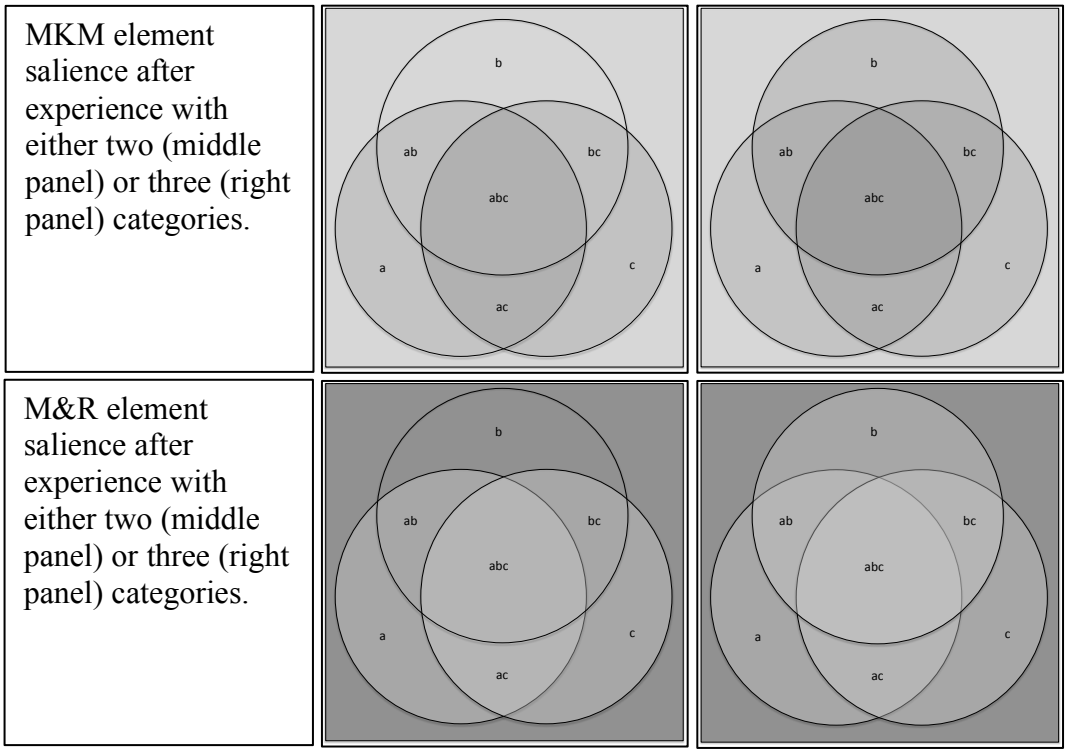
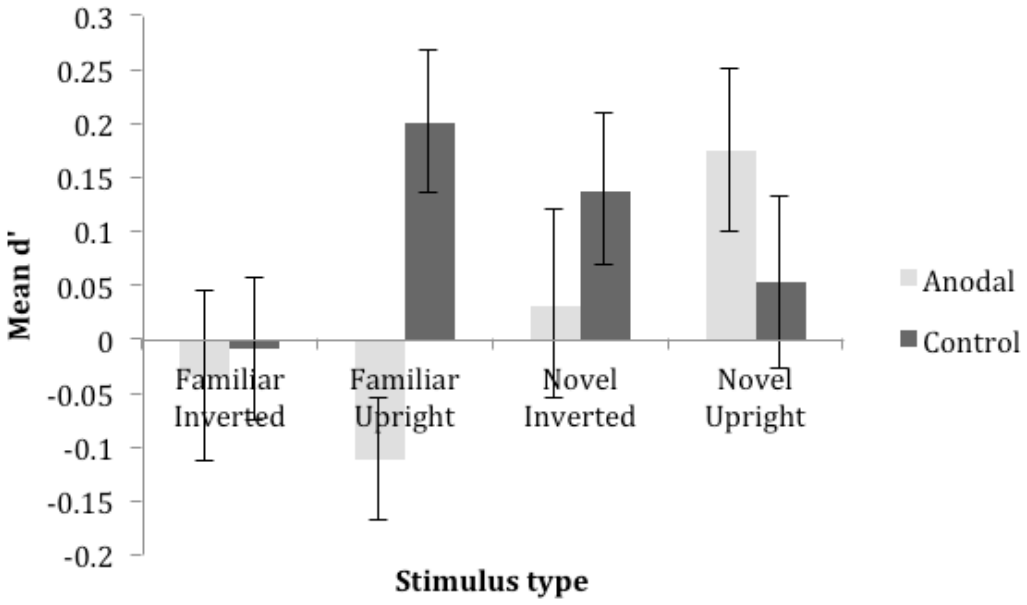

MKM element salience after experience with either two (middle panel) or three (right panel) categories.
b
ab
bc
abc
a
c
ac
b
ab
bc
abc
a
c
ac
M&R element salience after experience with either two (middle panel) or three (right panel) categories.
b
ab
bc
abc
a
c
ac
b
ab
bc
abc
a
c
ac

Figure 9


0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
Mean d'
Familiar Inverted
Familiar Upright
Novel Inverted
Novel Upright
Stimulus type
Anodal
Control