Please cite as:

Civile, C., Verbruggen, F., McLaren, R., Zhao, D., Ku, Y. and **McLaren, I.P.L.** (2016). Switching off perceptual learning: tDCS to left DLPFC eliminates perceptual learning in humans. *Journal of Experimental Psychology: Animal Learning and Cognition*, 42, 290-296.

Switching off perceptual learning:

Anodal tDCS at Fp3 eliminates perceptual learning in humans

Ciro Civile[1], F. Verbruggen[1], R.McLaren[1], D. Zhao[2], Y. Ku[2] and I.P.L McLaren[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 article 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

Perceptual learning can be acquired as a result of experience with stimuli that would otherwise be difficult to tell apart, and is often explained in terms of the modulation of feature salience by an error signal based on how well that feature can be predicted by the others that make up the stimulus. In this paper we show that anodal tDCS at Fp3 directly influences this modulation process so as to eliminate and possibly reverse perceptual learning. In two experiments, anodal stimulation disrupted perceptual learning (indexed by an inversion effect) compared with sham (Experiment 1) or cathodal (Experiment 2) stimulation. Our findings can be interpreted as showing that anodal tDCS severely reduced or even abolished the modulation of salience based on error, greatly increasing generalisation between stimuli. This result supports accounts of perceptual learning based on variations in salience as a consequence of pre-exposure, and opens up the possibility of controlling this phenomenon.

Keywords: perceptual learning, tDCS, associative learning, prediction error, salience modulation

Perceptual learning is fundamental to human cognition. It manifests as an enhancement in the ability to distinguish between similar stimuli (that otherwise would be very hard to tell apart) as a consequence of experience with them, or with other similar stimuli. This enhancement can be apparent in both learning and memory tasks involving these stimuli (e.g. McLaren, 1997). In this paper we will be concerned with those examples of perceptual learning that constitute more than just enhanced sensitivity to a stimulus dimension, and employ relatively complex stimuli (Mackintosh, 2009). It is this form of perceptual learning that allows the wine connoisseur to distinguish between different wines (James, 1890; see Hall, 1980 for a review; and McLaren and Mackintosh, 2000 for an analogous experiment with vinegars), the bird watcher to tell one warbler from another, and, we would argue (McLaren, 1997; Civile, Zhao, Ku, Elchlepp, Lavric, & McLaren, 2014) the rest of us to individuate faces.

Civile *et al* (2014) have shown that one consequence of perceptual learning is an inversion effect that can be demonstrated when performing an old/new recognition experiment with stimuli taken from a familiar (pre-exposed) category. Perhaps the best known instance of this phenomenon is the face inversion effect first reported by Yin (1969), Civile *et al* (2014) have recently shown that a similar effect can be obtained with arbitrary, artificial stimuli (checkerboards), as long as they meet certain prescriptions. Participants in these experiments were first familiarised with checkerboards from two prototype-defined categories (by learning to categorize exemplars from both categories), and then given a study phase using new checkerboards taken from one of the familiar categories intermixed with other checkerboards taken from a novel category. Checkerboards from either category could be presented in either an upright or an inverted orientation. In the later recognition phase they had to identify which checkerboards they had seen during the study phase. If the checkerboards had been taken from the familiar category then recognition was much better for upright checkerboards than for inverted ones. Orientation made no difference for checkerboards taken from the novel category. This inversion effect, first discovered in checkerboards by McLaren (1997), parallels that found with Greebles (artificial stimuli that share a common spatial configuration) by Gauthier and Tarr (1997) and can be used as a marker for perceptual learning. It manifests reliably, and both McLaren (1997) and Civile *et al* (2014) have shown that it is contingent on experience with a prototype-defined category. It is for these reasons that we will use it as our main index of perceptual learning in this paper.

The aim of the research reported in this paper was to investigate the effect of tDCS (trans-cranial Direct Current Stimulation, see Utz, Dimova. Oppenlander, & Kerkhoff, 2010 for a review) to dorsolateral pre-frontal cortex (DLPFC)[1] on perceptual learning of the type identified by Civile *et al* (2014). The procedure used by Civile *et al* (2014) employs a categorization phase to pre-expose participants to our stimuli - so we adopted the montage (see Figure 1) used by Ambrus, Zimmer, Kincses, Harza, Kovacs, and Paulus (2011), which has been shown to affect (and possibly eliminate) the prototype effect otherwise produced by training people to categorize stimuli. As we use prototype-defined checkerboard categories in our experiments, and formation of a strong representation of the prototype is a pre-requisite for perceptual learning according to theories such as MKM (McLaren, Kaye and Mackintosh, 1989), their Fp3 montage seemed appropriate for our purposes. We speculated that anodal tDCS using this montage might disrupt perceptual learning, and if this was the case it could eventually enable us to influence the mechanisms controlling representation development. Such a result would also have important theoretical implications, and could help us discriminate between salience modulation and comparison-based models of perceptual learning.

---

[1] We do not mean to imply that we selectively stimulated DLPFC, simply that the electrode montage was such that the Anode was situated over this region (at Fp3). tDCS using 35 cm^2 electrodes does not have a particularly focal effect.

Method

Participants

The first experiment (N=16 per group) was run in Exeter with undergraduates as participants. The second experiment (N=36 per group) was run in two replications, 16 per group in Shanghai, China, and 20 per group in Exeter, UK, again with undergraduates as participants. The permissible age range was 18-35 in both experiments; and the mean age in Experiment 1 was 20 with 72% females, and in Experiment 2 the mean age was 22 with 75% females. Analyses with replication as a factor showed that it did not interact significantly with any others in this experiment (max. $F(1,68)=1.83$, $p=.18$ uncorrected for multiple comparisons), nor does its inclusion materially change our analysis and so we collapsed over it. All participants gave informed consent before taking part in these experiments, which were approved by the ethics committee at the University of Exeter.

Stimuli

These were 16x16 checkerboards 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 color of a given square depend on that of its neighbors, such that if they were predominantly black then it was likely to be black and vice-versa. Exemplars were generated from the prototypes by adding noise (examples can be seen in the lower panels of Fig. 1). A randomly chosen 96 squares would be set at random in a given prototype to generate an exemplar of that category. We used 128 checkerboard exemplars from each of the four categories in this experiment, though not all of these stimuli would be used for a given participant. The stimuli used in the experimental phases were counterbalanced across subjects, such that each stimulus was used equally often as one drawn from a familiar category and as one drawn from a novel category.

tDCS

All participants first completed a brain stimulation safety-screening questionnaire. Stimulation was delivered by a battery driven constant current stimulator (Neuroconn) using two electrodes in 5cm x 7cm pieces of water-soaked synthetic sponge. One electrode was placed over the left PFC (Fp3) and the reference electrode was placed on the forehead, just above the right eyebrow (see Figure 1). Current was applied 1.5 min before the participants began the categorization task (whilst listening to instructions) and throughout categorization, and lasted for about 8.5 min (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 categorization commenced. The electrodes were left on the participant throughout the experiment. We used a double-blind procedure. One experimenter ran the participant, and another set up the stimulation according to specifications provided by a third party. The connections to the stimulator were concealed so that the polarity of stimulation could not be determined during the session.

Behavioral Procedure

Participants were pre-exposed to two prototype-defined checkerboard categories during tDCS (categorization phase). This phase took approximately 11 minutes, and began 1.5 minutes after stimulation commenced. The next phase began immediately, in which they studied a number of exemplars drawn from one of the familiar categories and from another novel category, in either upright or inverted orientations (study phase). Clearly the use of "upright" and "inverted" to describe the stimuli drawn from the novel category requires further explanation. In the case of this novel category, they have meaning in the sense that they are defined by the orientations used for these stimuli when employed as exemplars of the familiar category for other participants as a consequence of our counterbalancing (see next section for some further discussion of this point). This phase took approximately 4.5 minutes. They then immediately moved on to the recognition task (test phase) in which they had to classify checkerboards from both categories as either “old” (studied) or “new” (not studied). This took approximately 11 minutes.

Pre-exposure phase. In the pre-exposure phase, participants had to categorize checkerboard stimuli into two classes A vs. B. They saw 64 exemplars of each category (128 stimuli in all) and had to classify these exemplars by trial and error. A fixation cross was first shown for 1 sec, followed by a checkerboard, after which participants then had 4 sec to make their classification. They were given feedback immediately afterwards. One of these categories later served as the familiar category in the experiment. None of the exemplars used in this phase were employed in the later phases of the experiment. The purpose of this phase was more to expose participants to the categories than to succeed in training them in this classification. Nevertheless, in Experiment 1 mean categorization accuracy was 80.0% (SE=4.8%) correct for Anodal participants and 77.1% (SE=5.1%) for Sham participants, both being well above chance and not significantly different from one another. Similarly mean categorization accuracy in Experiment 2 was 78.7% (SE=3.2%) correct for Anodal participants and 77.9% (SE=3.5%) for Cathodal participants, again both values are above chance and do not significantly differ from one another.

Study phase. In the study phase, 64 checkerboard stimuli were shown one at a time in a random order for 3 sec each. All the checkerboards in this phase were novel to the participant. 16 of them were exemplars drawn from the familiar category shown in an upright orientation, i.e the orientation defined by experience of that category, 16 were exemplars from that familiar category but shown in an inverted orientation (i.e. rotated by 180°), 16 were exemplars taken from a category that was novel to that participant and shown in an upright orientation, and 16 were inverted exemplars from that novel category.

Test phase. In the test phase 128 stimuli were shown in total. Half of these were the 64 checkerboards studied earlier ("old" items) presented in the same orientation as during study, and another 64 were new checkerboards not seen in the study phase (or earlier) chosen so as to match those in the "old" set for category and orientation. This "new" set thus had another 16 upright and 16 inverted exemplars drawn from the familiar category, and 16 upright and 16 inverted exemplars taken from the novel category. Stimuli were presented one at a time and participants had to respond "old" or "new" within 4 sec, indicated by pressing one of two computer keys, "." and "x" counterbalanced across participants. Their accuracy scores were then converted into *d'* measures[2].

We use a number of planned comparisons in the analyses that follow, because our design and counterbalancing of stimuli lends itself to this approach, and because we have a considerable amount of data obtained with these stimuli and procedures on which to base our predictions for the control conditions. Following Civile *et al* (2014), our primary measure is the inversion effect given by comparing performance on upright and inverted exemplars taken from a given category. This difference is predicted on the basis of the experiments reported in Civile *et al* (2014) to be positive for the familiar category and is assessed by a one-tailed test. Comparison of this difference for the familiar category (which should be large and positive as a result of pre-exposure) to the same difference in the control stimuli (which should not be, but, given that we use a restricted set of stimuli, will be influenced by any

---

[2] We also computed the complementary measure of bias, but none of the analyses using this measure produced significant results so they are not mentioned here.

inadvertent difference in discriminability for upright and inverted exemplars) enables us to further assess the evidence for perceptual learning in this experiment, and we also use a 1-tailed here in assessing this interaction between our Familiarity and Orientation factors. We then further decompose this effect by comparing performance on upright exemplars taken from familiar and novel categories. This comparison is appropriate because the same stimulus sets are rotated across participants as part of our counterbalancing; so that each upright stimulus in the familiar category for a given participant will equally often serve as an upright stimulus from the novel category for other participants. Finally, we also assess whether performance is above chance in our control conditions, and for this we use a one-tail test against a mean of zero both on performance for the upright stimuli taken from a familiar category, and the average of performance on the novel stimuli, as these are the stimuli that we can make unambiguous predictions for. Civile *et al* (2014) found that performance on the inverted stimuli from the familiar category was reliably depressed in their experiments. All other tests except where noted are two-tail.

Results

Experiment 1

Experiment 1 (N = 16 per group) compared recognition performance after anodal or sham tDCS. We report significant and near significant results along with measures of variability (Standard Error or Mean Square Error) and effect size (Cohen's d followed by the 95% confidence interval for d).

Figure 1 about here please

We expected sham stimulation to allow the usual inversion effect for familiar category exemplars to manifest and begin with an analysis of the Sham condition to establish our baseline. As predicted, in the Sham condition ANOVA using factors of Familiarity

(Familiar vs. Novel category) and Orientation (Upright or Inverted) revealed that the interaction between Familiarity and Orientation was significant, F(1,15)=12.59, MSE=0.08, p=.001 1-tail, d=1.02, CI=1.36, 0.39. We also observed an inversion effect for familiar-category exemplars as assessed by a planned comparison, Upright better than Inverted, t(15)=1.86, SE=0.15, p = .041 1-tail, d=0.52, CI=0.89, 0.02, but not for novel-category exemplars under Sham stimulation (cf. Civile et al., 2014). Thus, our standard perceptual learning effect is present in the Sham participants. Participants do find this task difficult with these stimuli, as witness the generally low d' scores in Figure 1. Baseline performance on these stimuli is set by the novel stimuli for Sham participants, but note that the impression of higher performance on the control stimuli in the Sham condition given by Figure 1 is due to the superior performance on the inverted stimuli drawn from the novel category compared to those from a familiar category. This reflects the depressed performance expected on the inverted stimuli from the familiar category established in Civile *et al* (2014). Analysis of the upright stimuli from the familiar category, t(15)=2.05, SE=0.096, p=.029 1-tail, d=0.51, CI=0.94, 0.07 and of the averaged control stimuli from the novel category, t(15)=1.92, SE=0.11, p=.037 1-tail, d=0.48, CI=0.91, 0.04, confirms that performance was above chance in the Sham condition. It is not appropriate, given the existence of the inversion effect, to require that performance on the inverted exemplars from the familiar category be above chance, but it did not differ significantly from chance, t<1.

In the Anodal group, however, the Familiarity by Orientation interaction was not significant and there was no inversion effect for familiar-category exemplars (Upright was not better than Inverted). This difference between conditions resulted in a near significant Familiarity by Orientation by Stimulation (Anodal vs. Sham) interaction, F(1,30)=4.07, MSE=0.24, p = .053, d=0.71, CI=1.02, -0.02. In fact, performance in the Anodal condition on the novel-category upright exemplars was numerically superior to that on familiar-category upright exemplars, an effect that also approached significance, t(15)=2.08, SE=0.22, p = .055, d=0.78, CI=1.04, -0.01, and is the opposite of what might be expected as a result of perceptual learning on the basis of our previous studies. Finally, a direct comparison of performance on familiar upright exemplars also reveals that it was somewhat higher in the Sham group than in the Anodal group t(30)=2.02, SE=0.18, p=.053, d=0.71, CI=1.02, -0.02. Thus, anodal tDCS appears to have abolished the perceptual learning that would otherwise occur as a result of pre-exposure. Absolute performance on either the upright or inverted stimuli from the familiar category does not differ from chance, but performance for the novel category (assessed as the average of the two stimulus orientations) was actually numerically better than that in the Sham condition and differed significantly from chance, t(15)=2.64, SE=0.10, p=.009 1-tail, d=0.66, CI=1.11, 0.19. Therefore, the absence of an inversion effect in the Anodal group cannot be attributed to overall poor performance or a general failure to learn and remember. Instead, it appears to be specific to the exemplars shown during pre-exposure.

Experiment 2

In Experiment 2 (N = 36 per group, mean age 21.5 yrs., 75% female), we used cathodal stimulation as a control for the sensation of experiencing a 1.5mA current throughout exposure to the stimuli (Sham stimulation in Experiment 1 was only given for 30 sec). Our predictions for the Cathodal group were that the inversion effect should be the same as or stronger than that in the Sham group in Experiment 1 (given that this is the reverse polarity to Anodal stimulation), and our predictions for the Anodal group were that this inversion effect would be eliminated and possibly reversed based on the results of Experiment 1. This general pattern was present in our data, this time giving rise to a significant overall three-way Familiarity by Orientation by Stimulation interaction, F(1,70)=4.28, MSE=0.195, p = .042, d=0.48, CI=0.68, 0.01. The Familiarity by Orientation interaction in our control condition was not significant this time, F(1,35)=1.90, MSE=0.183, p = .085 one-tailed, d=.31, CI=0.51, 0.10, but our standard perceptual learning effect was present in the Cathodal condition, with the familiar-category exemplars producing an inversion effect, Upright better than Inverted, t(35)=1.84, SE=0.113, p = .037 one-tailed, d=0.38, CI=0.59, 0.02 that was absent in the novel-category exemplars. Performance on upright exemplars from the familiar category was significantly better than chance in this condition, t(35)=2.89, SE=0.07, p=.003 1-tail, d=.48, CI=0.77, 0.19, but performance on inverted exemplars from this category was not, nor was the averaged performance on exemplars from the novel category, both ts<1.

Consistent with Experiment 1, the inversion effect for exemplars drawn from a familiar-category was absent in the Anodal group. Again, there was evidence that novel-category upright exemplars were better recognised than familiar-category exemplars in this group, t(35)=2.03, SE=0.104, p = .05, d=0.42, CI=0.67, -0.01, which is opposite to what would be conventionally expected and opposite to the numerical effect in the Cathodal condition. As a consequence, performance on familiar category upright exemplars in the Cathodal group was significantly superior to performance on the same exemplars in the Anodal group, t(70)=2.97, SE=0.10, p = .004, d=0.70, CI=0.839, 0.146. Overall performance in the Anodal condition was poor, with none of the stimuli producing learning that differed significantly from chance, max t(35)=1.3. Nevertheless, when combined with the results of Experiment 1, the most comprehensive interpretation of these results is that anodal tDCS has had a selective effect on performance to upright exemplars drawn from the familiar category. There is also some evidence for a reversal of perceptual learning (i.e. enhanced generalisation) in the Anodal condition.

Discussion

Over two studies we have demonstrated that the inversion effect, our index of perceptual learning, that otherwise occurs in our control conditions is abolished, and even reversed by anodal tDCS to Fp3. We will now argue that this has important ramifications both for our understanding of what tDCS does, and for theories of perceptual learning. It also opens the door to possible applications of this technique in controlling perceptual learning in, for example, face processing.

We begin with the implications for our understanding of the effects of tDCS and it's impact on the inversion effect. The conventional result is for anodal tDCS to enhance learning by virtue of increasing cortical excitability (see e.g. Kincses, Antal, Nitsche, Bartfai and Paulus, 2003; Fregni, Boggio, Nitsche, Bermpohl, Antal, Feredoes *et al*, 2005 for examples; and Fertonani, Pirulli and Miniussi, 2011 for an example of an improvement in a different type of perceptual learning). Here, we have demonstrated something different; that anodal tDCS to Fp3 abolishes the inversion effect that would otherwise occur in our task. Instead of performance to upright exemplars drawn from a familiar category being superior to that for inverted exemplars drawn from the same category, as was the case for our control conditions, performance was numerically worse. In order to establish the lack of any inversion effect (defined as the difference between performance on the upright stimuli and inverted stimuli taken from the familiar category) in our data we computed Bayes factors using a half normal distribution starting at 0 (see Dienes, 2011) for this effect under anodal stimulation in each of our experiments using the control condition in each experiment to provide the necessary priors (the SD). The combined Bayes factor for the two experiments (obtained by multiplying the separate Bayes factors) was B=0.16, good evidence for the hypothesis that the effect in this condition was drawn from the null distribution (mean of zero) rather than that of the controls showing the inversion effect. This analysis confirms that anodal tDCS using our electrode montage has eliminated perceptual learning in this paradigm. In the absence of anodal stimulation, perceptual learning would confer an advantage that leads to the inversion effect found by Civile *et al* (2014) that is replicated here in our controls. This is, as far as we are aware, the first such demonstration of its kind. It is a quite striking empirical result, as it suggests that it may be possible to “turn off” perceptual learning in humans by means of anodal tDCS stimulation using our current electrode montage, in effect controlling perceptual learning in our participants.

Furthermore, performance on upright exemplars taken from the familiar category was worse after anodal stimulation than after sham or cathodal stimulation (although the difference with sham just failed to reach significance). In both experiments, we also found that, under Anodal stimulation, performance on upright exemplars taken from the familiar category was worse than performance on novel exemplars. Combined, these findings suggest that anodal stimulation not only abolished the advantage in discrimination that perceptual learning confers, but actually increased generalisation between stimuli that were taken from the familiar category. In a moment we will explain how this might fit with salience modulation theories of perceptual learning, but first we will consider how this result fits into an account based on comparison between stimuli over the course of experience with them (see for example Mundy, Dwyer and Honey, 2006; Mundy, Honey and Dwyer 2007, Wang, Lavis, Hall and Mitchell, 2012; Jones and Dwyer, 2013). The general idea here is that comparison between stimuli helps participants discover the important differences between them, and that this later assists in discrimination. We have at least two problems with such an account of our results. The first is that it does not easily explain the basic effect in our controls (as we argue in Civile *et al*, 2014). To succeed in the recognition task, discrimination both between and within categories is necessary. It is hard to see how within-category discrimination would be improved by comparison of a set of stimuli during pre-exposure that are not used for the recognition task, given that the changes made to the prototype in order to generate exemplars are random. And there is no ready explanation for why the effect is only obtained with prototype-defined categories, not others of equal difficulty and learnability. But the second reason concerns the novel result reported here. It could be argued that anodal tDCS had in some way disrupted or prevented comparison from taking place or proving beneficial - thus leading to a loss of perceptual learning. But it is hard to see how this could lead to what is effectively a reversal of the effect, enhanced generalization, as this would require something less than "zero" comparison, and it is not at all clear what this could be.

Turning now to other accounts of perceptual learning, given that we already have some evidence that points to the modulation of feature salience on the basis of error as one mechanism for perceptual learning (the dependence on category familiarity and the category being prototype-defined in McLaren, 1997 and Civile et al, 2014; the midpoint result in Experiment 2 of Mundy, Honey and Dwyer, 2007; see Mitchell and Hall, 2009 for a discussion), our departure point is the MKM theory of perceptual learning which attributes the effect to the differential latent inhibition of the common elements representing the stimuli (see McLaren, Kaye, & Mackintosh, 1989; McLaren & Mackintosh, 2000; and McLaren, Forrest, & McLaren, 2012). This model represents stimuli as sets of features represented by units in an error-correcting associative network (see Fig. 2). The error term controls learning, but is also used to modulate unit salience. Thus, the activity (salience) of a unit depends on how predicted it is. If it is well predicted by the other active units corresponding to features of the stimulus, then its error term will be low and its salience is low as well (latent inhibition). But, if its error term is high, then its activity will be high and it will easily enter into new associations. The advantage in recognition for upright exemplars taken from a familiar prototype-defined category is explained as being due to the relatively high salience of units that correspond to features that are distinctive for that exemplar, thus enhancing that exemplar's discriminability. This comes about because the units corresponding to prototypical features shared by most exemplars will predict one another well, leading to low error terms and low salience. The features that are unique to a given exemplar (or shared by relatively few exemplars) will be represented by units that are higher in salience, and so dominate learning helping to prevent unwanted generalisation from one exemplar (that has been seen in the study phase) to another (that has not). But this effect is contingent on experience with the category in an upright orientation, the advantage will be lost if the stimuli are upside down, thus leading to the inversion effect.

Figure 2 about here please

Figure 2 illustrates the operation of this model (see McLaren, 1997 for a more detailed account), and also indicates that if this modulation of salience in terms of error were to be disabled then it would produce a quite different effect. The model would revert to something similar to that of McClelland and Rumelhart's (1985) model of categorization (henceforth M&R), and would then predict a loss, and perhaps even a reversal of perceptual learning. This is because the common, prototypical elements would now be the most salient, and hence capture learning, leading to greater generalisation between stimuli. Thus, our data are consistent with the idea that 1.5mA anodal tDCS stimulation at Fp3 eliminates this modulatory input based on prediction error, leading to a loss of perceptual learning that manifests as increased generalisation. The behavioral consequences of this stimulation can be modelled by simple delta rule algorithms (M&R), which exhibit the increased generalisation between exemplars taken from familiar prototype-defined categories that we see in our data. In essence, we are arguing that anodal tDCS to Fp3 can change the system responsible for learning from that described by MKM to the original M&R model.

Another possible interpretation of our results is that rather than having a selective effect on perceptual learning and the inversion effect that we have used to measure this, anodal stimulation has instead simply reduced overall learning and hence performance in participants in that condition. This explanation is supported by the generally lower d' scores for the exemplars drawn from the familiar category under anodal stimulation, but is contradicted by the results for the exemplars taken from the novel category. In Experiment 1 these are higher than for the Sham condition, and in Experiment 2 they do not differ significantly from the Cathodal control. For these reasons we consider the explanation in terms of a selective effect on the exemplars from the familiar category, i.e an effect on perceptual learning rather than learning in general, is to be preferred.

An implication of our results is that our procedures might be used to effectively abolish perceptual learning in face recognition, if we follow the logic dictated by the experiments of McLaren (1997), Gauthier and Tarr (1997), Tarr and Gauthier (2000), Rossion, Gauthier, Goffaux, Tarr, and Crommelinck (2002), and Civile *et al* (2014). These studies point towards a role for perceptual learning in producing the inversion effect (at least in part) in faces. Given that we have abolished the inversion effect produced by perceptual learning in the experiments reported in this paper; then by extrapolation anodal tDCS to Fp3 might be expected to disrupt the face inversion effect as well. We must be cautious at this point, however, as we are not yet in a position to know for certain whether our procedures affect pre-exposure effects - and leave already established perceptual learning unaffected - or also extend to performance involving stimulus sets that are already familiar. Further investigation will be needed to clarify this intriguing possibility.

Acknowledgements

IPLM and FV are supported by a grant from the ESRC (ES/J00815X/1), and FV 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. CC 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.

We would like to thank our research assistant R. Lee for her help with some of the tDCS setup procedures.

References

Ambrus G. G., Zimmer M., Kincses Z. T., Harza I., Kovacs G., Paulus W., et al. (2011). The enhancement of cortical excitability over the DLPFC before and during training impairs categorization 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 checkerboards. *Journal of Experimental Psychology: Animal Behavior Processes*, 40, 144-61.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology*: *General*, 115, 107-117.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6, 274–290.

Fertonani, A., Pirulli, C., & Miniussi, C. (2011). Random noise stimulation improves neuroplasticity in perceptual learning. *Journal of Neuroscience, 31*,15416–15423.

Fregni, F., Boggio, P. S., Nitsche, M. A., Bermpohl, F., Antal, A., Feredoes, E., et al. (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.

Hall, G. (1980). Exposure learning in animals. *Psychological Bulletin*, 88, 535-550.

James, W. (1890). *Principles of psychology*. New York: Holt.

Jones, S.P. and Dwyer, D.M. (2013). Perceptual learning with complex visual stimuli is based on location, rather than content, of discriminating features. *Journal of Experimental*

Psychology: Animal Behavior Processes, 39, 152-65.

Kincses T. Z., Antal A., Nitsche M. A., Bártfai O., and 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 SL (2012) Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition 124*, 36–49.

Mackintosh, N.J. (2009). Varieties of perceptual learning. *Learning & Behavior*, 37, 119–125.

McClelland, J.L. & 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). Categorization and perceptual learning: An analogue of the face inversion effect. *The Quarterly Journal of Experimental Psychology* 50A, 257-273.

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. In R.G.M. Morris (Ed.) *Parallel Distributed Processing - Implications for Psychology and Neurobiology*. Oxford, Oxford University Press.

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.

Mitchell, C. and Hall, G. (2014). Can Theories of Animal Discrimination Explain Perceptual

Learning in Humans? *Psychological Bulletin*, 140, 283-307.

Mundy, M. E., Dwyer, D. M. and Honey, R. C. (2006). Inhibitory associations contribute to perceptual learning in humans. *Journal of Experimental Psychology: Animal Behaviour Processes*, *32*, 178-184.

Mundy, M.E., Honey, R.C. and Dwyer, D.M. (2007). Simultaneous Presentation of Similar Stimuli Produces Perceptual Learning in Human Picture Processing. *Journal of Experimental Psychology: Animal Behavior Processes*, 33, 124-38.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M.-J., Crommelinck, M. (2002). Expertise training with novel objects leads to face-like electrophysiological responses. *Psychological Science*, *13*, 250-257.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3, 764-769.

Utz, KS, Dimova, V, Oppenländer, K, Kerkhoff, G (2010) Electrified minds: transcranial direct current stimulation (tDCS) and galvanic vestibular stimulation (GVS) as methods of non-invasive brain stimulation in neuropsychology—a review of current data and future implications. *Neuropsychologia 48*, 2789–2810.

Wang, T., Lavis, Y., Hall, G. and Mitchell, C.J. (2012). Location and salience of unique features in human perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*, 407-418.

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

Figure 1

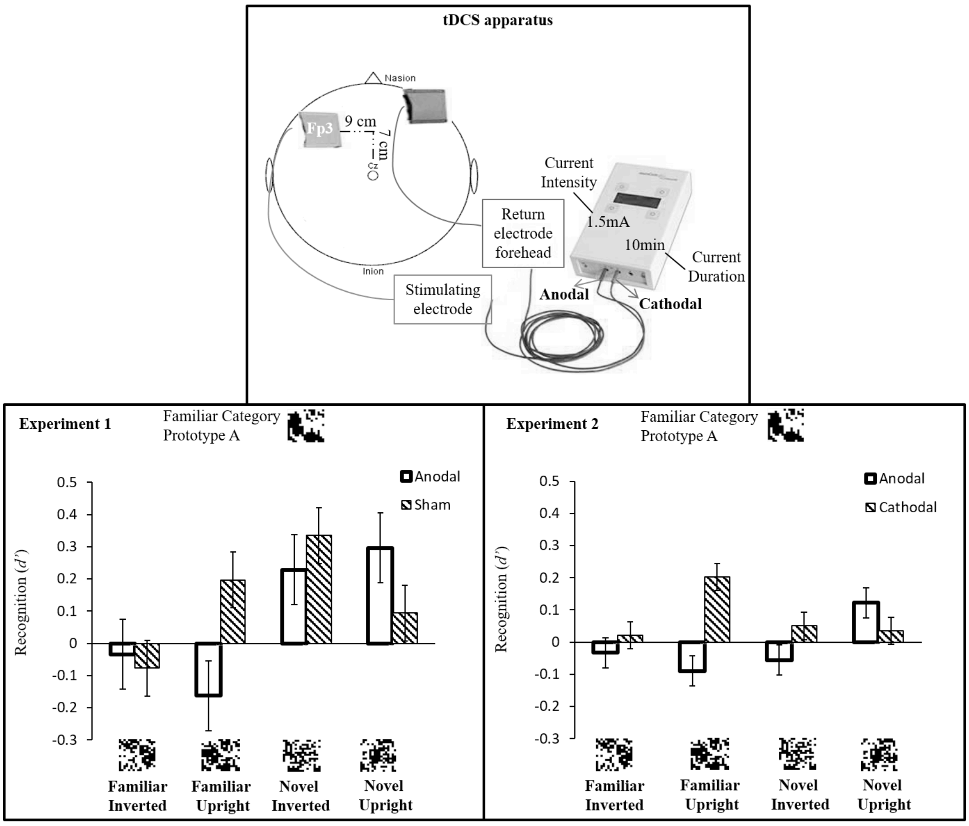

Figure 1: The top panel of the figure illustrates the electrode configuration and the tDCS apparatus used in both experiments. The results of Experiment 1 (bottom left panel) and Experiment 2 (bottom right panel) are shown below. Open bars are for anodal stimulation and hatched bars for control stimulation. 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 with illustrative exemplars derived from the relevant prototype (example given at top).

Figure 2

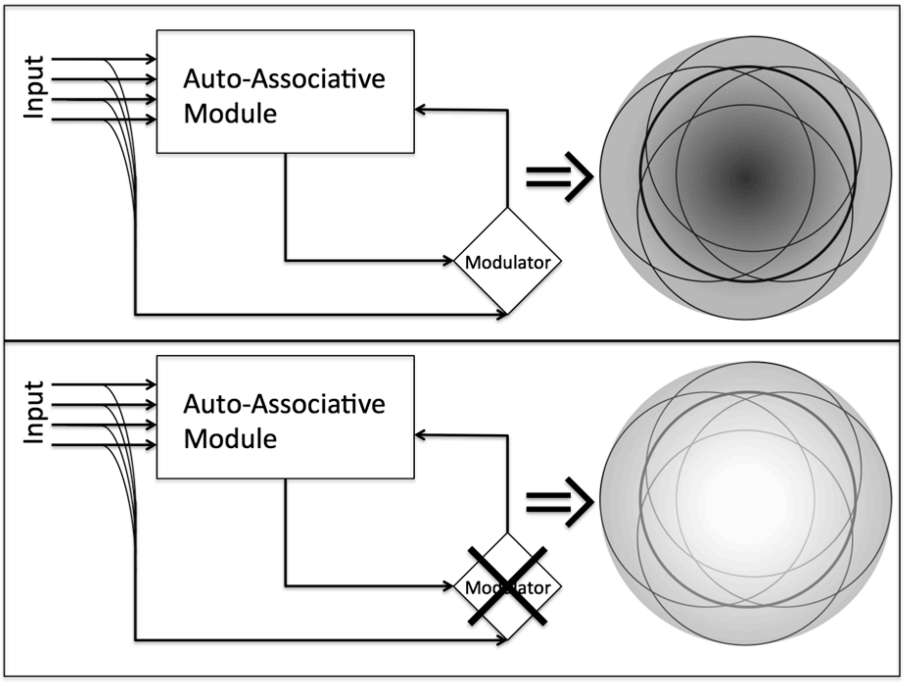

Figure 2: The top panel illustrates how modulation driven by prediction error can be used to influence feature salience. Stimuli are represented as sets of elements (each stimulus set contained within one circle) that can associate with one another in the Auto-Associative Module. The modulator determines element salience based on that element's prediction error. The result, shown in the temperature diagram right, is that stimulus features that are more predictable become less active (darker shading) leading to reduced salience and hence 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 panel shows how removing this modulatory input reverses this effect, making the common, prototypical features of the stimuli the most salient (lighter shading).