

An examination of the neural correlates and behavioural phenomena of category learning

Submitted by Kathryn Louise Carpenter to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Psychology in January 2017

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgment

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:.....

Acknowledgments

Writing this thesis has undoubtedly been the greatest experience, and it would not have been possible without a huge network of people. Firstly, for unwavering support, encyclopaedic knowledge, and utter patience I thank Dr Fraser Milton. I would also like to thank Dr Abdelmalek Benattayallah, Dr Charlotte Edmunds, Prof. Ian McLaren, Rossy McLaren and Prof. Frederick Verbruggen, and Prof. Andy Wills, without whom the research presented in this work would have been much harder to complete. I am grateful to Ben Hurlock, Robyn Lee, Colomba Lyall, and Holly Wilson for their assistance with the tDCS experiments, and also to the students of the University of Exeter for participating in my studies.

I would like to acknowledge how lucky I was to share an office with such wonderful people; Lorna Hardy, Amy McCarthy, Lucy Porter, Fayme Travis, and especially Maisy Best and Will Bowditch provided reassurance, inspiration and humour during days of elusive participants and statistical mishaps.

I dedicate this thesis to my family. I am especially indebted to my Mum, Pam Carpenter, who always pushes me to be the very best version of myself, and my Dad, Mike Carpenter, who is the hardest working person I know and always believes in me – without their love and (colossal) support I would have never considered this path. I also thank my sister Emmily Carpenter for unconditional friendship, and my brother and sister in law, Justin Carpenter and Terri Anthony, for their support, and also ‘Game Nights’ – the best stress relief. My final PhD year was also brightened by the arrival of Spencer Michael, who has brought happiness to everyone. I am also grateful to Jillian Webber, who will always underestimate how important she is to her ‘old girls’. Finally, I owe my sanity to Shaun Edworthy, my rock, my world, my everything.

Abstract

This thesis investigates the neurobiological pathways that underpin learning of visual categories, and the behaviour associated with these neural systems. The work contains two strands. The first assesses the neural and behavioural predictions of the COmpetition between Verbal and Implicit Systems (COVIS) account of category learning. The second aims to examine the brain regions implicated in the prototype effect after transcranial Direct Current Stimulation (tDCS) to the left dorsolateral prefrontal cortex (DLPFC).

COVIS predicts there are separate explicit and implicit category learning systems. According to COVIS, the explicit system optimally learns rule-based (RB) categories and relies upon the frontal lobes for working memory (WM) and executive functioning processes, and the medial temporal lobes (MTL) to store decision boundaries. In contrast, the implicit system employs the basal ganglia to procedurally learn information-integration (II) categories through stimulus-response associations.

Experiment 1 found little evidence of separable implicit or explicit systems in an fMRI study that investigated category decision making processes during RB and II category learning using conditions matched in difficulty, category separation and number of relevant stimulus dimensions. Contrary to the predictions of COVIS, the MTL was more active during the II condition compared to the RB condition, an area that should be more engaged by the explicit system. There was also extensive neural activation overlap found between RB and II learning. Experiments 2 and 3 aimed to generalise these neural findings to activation during feedback processing in RB and II conditions. Experiment 2 was a behavioural study which showed that adding a feedback delay necessary for fMRI data analysis did not differentially impact RB or II learning.

Experiment 3, including this feedback delay, found the same neural pattern of results as Experiment 1 offering further support that the MTL is more engaged in II learning than RB learning. There was also again considerable overlap in the regions involved in the two tasks. Taken together, Experiments 1 to 3 found no evidence for the neurally dissociable category learning systems predicted by COVIS.

Experiments 4, 5 and 6 investigated the behavioural dissociation reported by Smith et al. (2014) that deferring feedback to the end of a six trial block selectively impairs II learning compared to a unidimensional RB condition. Experiment 4 replicated this result. However, when equating the number of dimensions relevant for RB and II learning in Experiment 5, both conditions were hindered by deferring feedback, with Experiment 6 confirming that conjunctive RB learning was impaired by deferred feedback compared to immediate feedback. I concluded that the dissociation reported by Smith et al. is attributed to the use of a unidimensional category as a comparison for II performance, and that when the number of relevant stimulus dimensions between conditions are controlled there is little evidence for the separable systems of COVIS. Experiment 7 used tDCS to investigate if RB or II learning was differentially affected by anodal stimulation to the left DLPFC. Although there was no significant difference in learning between category conditions, during anodal stimulation participants improved less across blocks than those receiving sham stimulation. While the results suggest that the effect of tDCS on RB and II learning may be more tangible during stimulation, the numerical pattern of the data warrants further research into the possibility that RB participants are more affected by tDCS than II participants after stimulation to the left DLPFC.

Strand 2 of this thesis aimed to further previous work that suggests anodal stimulation to the DLPFC during a prototype distortion task induced a prototype effect

(better responding to unseen prototype trials than other category exemplars derived from this prototype) that was not present in sham participants. Contrary to this past work, Experiments 8 and 9 found that anodal stimulation to the left DLPFC inhibited a prototype effect that was present in sham participants. Experiment 10 implemented a combined tDCS and fMRI task and found that anodal participants engaged the stimulated DLPFC and the MTL more than sham participants in measures of the prototype effect. Based on these findings, this thesis argues that anodal tDCS to the left DLPFC inhibits perceptual learning by disrupting error prediction processes. Anodal participants are also considered to use generalization more than sham participants when perceiving category exemplars, a process attributed to the MTL.

Table of Contents

List of Figures	11
List of Tables	13
Declaration	15
Chapter 1. General introduction	16
Strand 1: The COmpetition between Verbal and Implicit Systems	17
Model	
The COVIS model	17
Behavioural dissociation research	20
Neuroimaging evidence investigating the COVIS model	24
Research aims of Strand 1	28
Strand 2: The prototype effect	30
Neuroimaging research	33
TDCS and the prototype distortion task	34
Research aims of Strand 2	37
Chapter 2. Investigating the neural correlates of rule-based and information-integration category learning with fMRI	39
Experiment 1	38
Method	42
Participants and design	42
Stimuli	43
fMRI imaging	44
Procedure	45
fMRI data analysis	46
Results	48
Behavioural analysis	48
Imaging analysis – ‘odd-or-even?’ baseline measure	50
Analysis of all runs	50
Analysis of runs 3 and 4 only	56
Modelling analysis	58
Imaging analysis – incorrect trials baseline measure	64
Discussion	67

Chapter 3. Investigating the feedback processing systems of rule-based and information-integration learning with fMRI	76
Introduction	76
Experiment 2	82
Method	82
Participants	82
Design	83
Stimuli	83
Procedure	84
Results	85
Discussion	87
Experiment 3	88
Method	89
Participants	89
Stimuli and procedure	89
FMRI imaging	89
FMRI data analysis	90
Results	91
Behavioural analysis	91
Modelling analysis	92
Imaging analysis of both learning sessions	93
Imaging analysis of session 1	95
Imaging analysis of session 2	97
Imaging analysis comparing session 1 and session 2	100
Analyses of correct and incorrect trials	101
Session 1	101
Session 2	105
Discussion	107
General discussion	108
Appendix 3.1: Imaging analysis of activation during decision making	113
Imaging analysis of both learning sessions	113
Imaging analysis of learning session 1	114
Imaging analysis of learning session 2	115

Chapter 4. The effect of deferring corrective feedback on rule-based and information-integration category	118
Introduction	118
Experiment 4	121
Method	121
Participants and design	121
Stimuli	122
Procedure	123
Modelling analysis	125
Results	125
Modelling analysis	127
Discussion	128
Experiment 5	129
Method	129
Participants and design	129
Stimuli	130
Procedure and analysis	130
Results	131
Modelling analysis	132
Discussion	132
Experiment 6	134
Method	135
Results	136
Modelling analysis	137
Discussion	137
General discussion	138
Appendix 4.1: Modelling analyses containing all blocks of the experiment	143
Chapter 5. Investigating the effect of anodal tDCS on rule-based and information-integration category learning	144
Experiment 7	144
Method	151
Participants	151
Design	152

Stimuli	152
TDCS	153
Procedure	154
Results	155
Discussion	159
Appendix 5.1: TDCS safety questionnaire	164
Chapter 6. Investigating the prototype effect with transcranial Direct Current Stimulation	167
Introduction	167
Experiment 8	169
Method	170
Participants	170
Stimuli	170
TDCS	172
Design and procedure	173
Results	174
Discussion	179
Experiment 9	181
Method	183
Participants	183
Stimuli	184
TDCS	184
Design and procedure	184
Results	185
Repeated trial and non-repeated trial accuracy	185
Analysis of training and test accuracy	185
Combined analysis of Experiments 8 and 9	188
Discussion	193
Experiment 10	194
Method	195
Participants	195
Stimuli	195
TDCS	196
fMRI data acquisition	196

Design and procedure	196
FMRI data analysis	197
Results	198
Behavioural analysis	198
Imaging analysis – the training phase	202
Imaging analysis – the prototype effect	203
Imaging analysis by trial type in the test phase	205
Discussion	210
General Discussion	211
Chapter 7. General discussion	220
Chapter 2 (Experiment 1)	221
Chapter 3 (Experiments 2 and 3)	223
Chapter 4 (Experiments 4, 5 and 6)	226
Chapter 5 (Experiment 7)	229
Chapter 6 (Experiments 8, 9 and 10)	230
Concluding remarks	233
References	235

List of Figures

Figure 1.1	Illustration of unidimensional, conjunctive and information-integration category structures (Zeithamova & Maddox, 2006)	18
Figure 1.2	The category structures of Maddox and Ing (2005)	21
Figure 1.3	The category structures of Nomura et al. (2007)	25
Figure 1.4	The results of McLaren et al. (2016; Experiment 2a)	36
Figure 2.1	The category structures of Filoteo et al. (2010)	40
Figure 2.2	The category structures used in Experiment 1	44
Figure 2.3	Mean accuracy of performance in Experiment 1	49
Figure 2.4	Neural areas active in all runs in II and RB learning	51
Figure 2.5	Neural areas more active in the incorrect trials in II compared with RB participants	53
Figure 2.6	Neural areas more active in the II condition than the RB condition in all runs	54
Figure 2.7	Neural areas active in runs 3 and 4 during II and RB learning	56
Figure 2.8	Neural areas more active in II compared with RB participants in runs 3 and 4	58
Figure 2.9	Neural areas active in all runs of the modelling analysis	61
Figure 2.10	Areas of the MTL more active in correct trials than incorrect trials in II compared with RB participants	67
Figure 3.1	The category structures used by Maddox and Ing (2005)	78
Figure 3.2	The category structures used in Chapter 3	83
Figure 3.3	Mean accuracy of performance in Experiment 2	86
Figure 3.4	Mean accuracy of performance in Experiment 3	91
Figure 3.5	Analysis of neural activation in all runs of Experiment 3	94
Figure 3.6	Analysis of brain activation in session 1 of Experiment 3	96
Figure 3.7	Analysis of brain activation in session 2 of Experiment 3	98
Figure 3.8	Analysis comparing activation in session 1 and session 2	101
Figure 3.9	Areas more active in correct trials than incorrect trials in session 1	102
Figure 3.10	Areas more active in correct trials than incorrect trials in session 2	105

Figure S.3.1	Analysis of areas active during decision making in Experiment 3	114
Figure S.3.2	Analysis of areas active during decision making in session 2 of Experiment 3	116
Figure 4.1	Examples of stimuli used by Smith et al. (2014)	122
Figure 4.2	Example of the RB and II category structures of Experiment 4	123
Figure 4.3	Mean response accuracy in Experiment 4	126
Figure 4.4	The category structures of Filoteo et al. (2010)	130
Figure 4.5	Mean response accuracy in Experiment 5	131
Figure 4.6	Mean response accuracy in Experiment 6	136
Figure 5.1	The category structures of Perry and Lupyan (2014)	148
Figure 5.2	The category structures of Filoteo et al. (2010)	153
Figure 5.3	Mean response accuracy in Experiment 7	156
Figure 5.4	Mean response accuracy in the first half of Experiment 7	158
Figure 5.5	Mean response accuracy in the second half of Experiment 7	159
Figure 6.1	Prototypical and typical exemplar stimuli for Chapter 6	172
Figure 6.2	Accuracy of responding in Experiment 8	176
Figure 6.3	The prototype effect in Experiment 8	178
Figure 6.4	Accuracy of responding in Experiment 9	187
Figure 6.5	The prototype effect in Experiment 9	187
Figure 6.6	Accuracy of responding in Experiments 8 and 9 combined	190
Figure 6.7	Accuracy of responding in Experiment 8 and Experiment 9	191
Figure 6.8	The prototype effect in Experiments 8 and 9 combined	192
Figure 6.9	Accuracy of responding in Experiment 10	199
Figure 6.10	The prototype effect in Experiment 10	201
Figure 6.11	Analysis of brain areas activated during the training phase	203
Figure 6.12	Brain areas more active in prototype trials than all exemplar trials	205
Figure 6.13	Analysis of brain areas active during prototype trials	207
Figure 6.14	Analysis of brain areas active during old trials	208
Figure 6.15	Analysis of brain areas active during new trials	209
Figure 6.16	Model representing perceptual learning with and without prediction error	215

List of Tables

Table 2.1	Brain regions more active in the II condition than the RB condition in the last runs of the study compared to the first runs	52
Table 2.2	Brain regions more active in the II condition than the RB condition in all runs	54
Table 2.3	Brain regions more active in the II condition than the RB condition in runs 3 and 4	57
Table 2.4	Learning strategies of the participants	60
Table 2.5	Brain regions active in the II and RB conditions in runs 3 and 4 of the modelling analysis	63
Table 2.6	Brain regions more active in correct trials than incorrect trials in RB and II conditions in all runs	65
Table 2.7	Brain regions more active in correct trials than incorrect trials in RB and II conditions in runs 3 and 4	66
Table 3.1	Learning strategies of the participants	93
Table 3.2	Brain regions more active in the II condition than the RB condition in session 2	99
Table 3.3	Comparisons of brain regions active in the RB and the II condition in incorrect and correct trials	104
Table 3.4	Brain regions more active in the II condition than the RB condition in incorrect trials	106
Table 4.1	Distributional characteristics of the stimuli used by Smith et al. (2014)	124
Table 4.2	Learning strategies used by participants in Experiment 4	128
Table 4.3	Learning strategies used by participants in Experiment 5	133
Table 4.4	Learning strategies used by participants in Experiment 6	137
Table S.4.1	Learning strategies used by participants in all blocks of Experiment 4	144
Table S.4.2	Learning strategies used by participants in all blocks of Experiment 5	144
Table S.4.3	Learning strategies used by participants in all blocks of Experiment 6	144

Table 6.1	Brain regions associated with the prototype effect more active in anodal participants than sham participants	206
Table 6.2	Brain region more active in sham participants than anodal participants in old trials.	209

Declaration

The research reported in this thesis was carried out at the University of Exeter between September 2012 and August 2016 and was supervised by Doctor Fraser Milton. Data from 36 participants of Experiment 1 was collected during a Masters degree at the University of Exeter. Data from 9 participants was collected in September 2013, and the results reanalysed and modelled as a complete data set of 45.

This thesis has not been submitted, in whole or in part, for any other degree, diploma or qualification at any university.

Kathryn Carpenter

January 2017

Chapter 1. General introduction

“Because our minds need to reduce information, we are more likely to try to squeeze a phenomenon into the Procrustean bed of a crisp and known category (amputating the unknown), rather than suspend **categorization**, and make it tangible.”

(Nassin Nicholas Taleb, 2010, pp. 105)

Many daily tasks are second nature to us. For example, when driving to work innumerable abstract stimuli let us know when to stop and go, what the road ahead may look like and where to park. We can easily interpret the meaning of road signs; we have learnt what actions to (or not to) perform when presented with these stimuli. Such learning is synonymous with our prehistoric ancestors who learnt to distinguish objects such as berries by their shape, size and colour into ‘poisonous’ and ‘edible’ categories. The broad aim of this work is to investigate how humans learn to group novel stimuli into meaningful categories.

The particular focus of this thesis will be how visual categories are learnt. Here the visual features of a stimulus can distinguish it as a member of a certain category (Richler & Palmeri, 2014). In such research, participants must learn to separate novel and abstract stimuli (e.g., Gabor patches: Maddox, Ashby, & Bohil, 2003; Nomura et al., 2007 and line stimuli: Filoteo, Lauritzen & Maddox, 2010) into two or more categories. This general introduction is separated into two strands, as is the entirety of the thesis. The first strand will focus on the ‘COmpetition between Verbal and Implicit Systems’ model of category learning (COVIS; Ashby, Alfonso-Reese, Turken, & Waldron, 1998) which posits that we have both an explicit and implicit category learning system which compete for control over the decision. The aim of Strand one is to examine the predictions of COVIS from both a behavioural perspective and a neuroscience perspective using functional Magnetic Resonance Imaging (fMRI) and transcranial Direct Current Stimulation (tDCS) methods.

Strand 2 of this thesis focuses on the prototype effect found in the category learning literature. Here, participants who have been trained on prototype derived categories respond more accurately to the previously unseen prototype than to previously seen or unseen exemplars of this category (e.g., Posner & Keele, 1968). Strand 2 will examine the prototype effect using tDCS and fMRI.

Strand 1: The Competition between Verbal and Implicit Systems model

Poldrack and Foerde (2008) consider the development of multiple memory systems theories as one of the most important contributions to neuroscience in the past quarter century, with one increasingly prominent line of research within this field focusing on the possibility of multiple systems of category learning (e.g., Ashby et al., 1998; Erickson & Kruschke, 1998; Newell, Dunn, & Kalish, 2011; Nosofsky & Johansen, 2000; see also Richler & Palmeri, 2014, for a review). Perhaps the most prominent multiple systems model of category learning is the COVIS account (Ashby et al., 1998). This dual system model is an excellent basis for category learning research as it makes clear behavioural and neurobiological predictions (Ashby et al., 1998).

The COVIS model theorises that there are two functionally and neurally separable systems responsible for category learning (Ashby et al., 1998). The explicit learning system is hypothesised to use working memory (WM) and executive functioning to generate and test verbal categorization rules which can be applied to stimuli in order to correctly classify them. The concept of executive functioning is likely to contain a broad set of processes necessary to exhibit task control such as task switching, updating WM and inhibiting irrelevant information (Miyake et al., 2000; Monsell & Driver, 2000). However, COVIS does not specify how such processes are implicated in explicit category learning, and so I will refer to executive functioning in rather broad terms throughout this thesis. The explicit system is optimally used when

learning to categorize rule-based (RB) category structures. In the COVIS literature, differing category structures are created by manipulating decision boundaries (see Figure 1.1). The decision boundary is a line drawn in perceptual space that separates stimuli that are plotted by values on category relevant dimensions, into perceptually different categories (Ashby & Gott, 1988). In RB categories these decision boundaries are easily verbalizable. Such RB categories can be unidimensional, where the defining rule is based on one stimulus dimension (e.g., Figure 1.1.a where the rule for category learning is “short lines in category A and long lines category B”). RB structures can also be conjunctive, which have a verbalizable decision boundary where two or more stimulus dimensions must be considered to make a category decision (Ashby & Gott, 1988; e.g., Figure 1.1.b where the rule is “short, upright, lines belong in category A; anything else belongs in category B”). The explicit system is proposed by COVIS to be the route of learning applied by default at the start of category acquisition (Ashby et al., 1998). However, if during category learning the explicit system is proven unsuccessful by corrective feedback, then a switch is made to use the implicit category learning system (Ashby et al., 1998; Filoteo et al., 2010).

This image has been removed by the author of this thesis for copyright reasons.

Figure 1.1. Examples of unidimensional, conjunctive, and information-integration category structures. Each open circle represents one member of category A; each filled square represents one member of category B. Figure adapted from Zeithamova and Maddox (2006).

The implicit system of COVIS is procedural based and predecisionally combines information from two or more unrelated stimulus dimensions (e.g., the angle and length of a line). Immediate corrective feedback regarding the category response made to a stimulus is essential for the implicit system to function adequately as it generates stimulus-response associations without reliance on WM and executive functioning (Ashby et al., 1998). The implicit category learning system is optimally applied when presented with information-integration (II) category structures (e.g., Figure 1.1.c) which are assumed to have difficult or impossible to verbalize decision boundaries (Ashby & Gott, 1988). When faced with II structures the explicit systems performance is suboptimal, so a switch is soon made to the implicit system (Ashby et al., 1998).

One notable aspect of COVIS that distinguishes it from other multiple systems accounts of category learning (e.g., ATRIUM, Erickson & Kruschke, 1998) is the detailed neurobiological predictions that it makes regarding the brain regions that underlie the different learning systems. In the explicit system, rule generation and hypothesis testing requires WM and executive functioning which takes place predominately in the prefrontal cortex (Ashby & Valentin, 2005). The particular rule to use is selected via the anterior cingulate (Maddox & Ashby, 2004), while the head of the caudate nucleus is responsible for mediating the switch to a different rule. Successful rules are stored in the medial temporal lobes (MTL) for future use (Ashby & Valentin, 2005). The MTL is also hypothesised to store representations of the decision boundaries used to separate the stimuli into categories (Nomura & Reber, 2008).

In contrast, the implicit system procedurally acquires the stimulus-response associations necessary for learning II categories (Ashby et al., 1998). The body and tail of the caudate nucleus receive representations of the visual stimulus perceived (Ashby & Valentin, 2005) and these cells project to the supplementary motor area via the globus

pallidus and the thalamus (Maddox & Ashby, 2004). When feedback indicates a correct response has been made, the substantia nigra releases dopamine which strengthens the association of the stimulus to the correct response (Ashby & Valentin, 2005). The putamen has also recently been proposed by Waldschmidt and Ashby (2011) to play a key role in the implicit system, as it is assumed to provide information to the motor regions (but see Ell, Marchant, & Ivry, 2006, who found that focal putamen lesions impaired RB but not II learning).

Behavioural dissociation research

Dissociations between II and RB learning have been found using many behavioural manipulations. II learning has been found to be disrupted by changing the appropriate response buttons whilst RB learning is not, supporting the prediction that procedural stimulus-response associations are necessary in II learning (Ashby, Ell, & Waldron, 2003). Additionally, deferring feedback to the end of a six trial block (Smith et al., 2014), and providing the category label prior to participants making the response (Ashby, Maddox, & Bohil, 2002), have all led to II learning impairment but intact RB performance. Similarly, some studies have suggested that increasing the number of categories to learn (Maddox, Filoteo, Hejl, & Ing, 2004) or reducing feedback processing time (Maddox, Ashby, Ing, & Pickering, 2004) disrupts RB but not II learning. However, there are a growing number of studies that question the dual system interpretation of the dissociations by suggesting that these dissociations can be explained by methodological artifacts and a single explicit system of category learning (e.g., Newell, Dunn, & Kalish, 2010, 2011; Newell, Moore, Wills, & Milton, 2013; Nosofsky & Kruschke, 2002; Stanton & Nosofsky, 2007, 2013)¹. This section will now provide several examples to give a flavour of the competing viewpoints.

¹ Confounding variables in the studies reviewed in Chapter 1 will be discussed in more depth in the proceeding chapters.

COVIS predicts that for successful II learning corrective feedback must be immediately administered. This is so that dopamine can be released and strengthen the synaptic pathways between representations of the stimulus seen and the response made (Ashby et al., 1998; Ashby & Valentin, 2005). However, if feedback is delayed the transient synaptic activity between representations is depleted which disrupts the stimulus-response associations necessary for II learning (Ashby & Valentin, 2005).

Supporting these predictions, delaying feedback has been found to impair performance in II learning conditions (e.g., Maddox, Ashby & Bohil, 2003). Maddox and Ing (2005) delayed giving participants corrective feedback for 5 seconds when learning Gabor patch stimuli in II or RB conditions. RB learning was unaffected by delaying feedback. Yet, delaying feedback interrupted II performance when compared with performance in immediate feedback II conditions due to the reliance of the implicit system on procedural processes.

However, the stimuli in the RB condition of Maddox and Ing (2005) were further away from the decision boundary than II stimuli (see Figure 1.2) introducing a difference of ‘category separation’ between RB and II conditions. In the RB condition,

This image has been removed by the author of this thesis for copyright reasons.

Figure 1.2. The category structures used by Maddox and Ing (2005); (a) The RB structure; (b) The II structure. The lines represent the decision boundary separating the stimuli into four categories. Figure adapted from Maddox and Ing (2005).

the mean distance of stimuli from the decision boundary as plotted in stimulus space was greater relative to the II condition. Therefore, on average RB stimuli were further away from the category decision boundary, making the RB categories more separated. Consequently, the II categories were perceptually more similar to each other than RB categories perhaps resulting in the generation of weaker category representations in the II condition. This may result in participants in the II condition becoming more susceptible to less optimal feedback manipulations (such as a delay), as learning in this condition is more difficult (Newell et al., 2011; Stanton & Nosofsky, 2007).

It has also been questioned whether Maddox and Ing's use of a random Gabor patch as a mask presented to participants on the screen between making the response and receiving feedback may have affected learning (Newell et al., 2011). In deferred feedback conditions the mask was presented for a total of 5 seconds per trial, whereas it was only presented for 750ms in immediate trials. The presentation of the Gabor patch mask for 5 seconds in delayed conditions could result in the mask being incorrectly represented as a category stimulus (Newell et al., 2011). This could selectively impede II learning which is predicted to rely upon the formation of stimulus-response associations. Conversely RB learning would be unaffected as it is not reliant on procedural processes; instead the original category stimulus can be held in mind by WM before feedback is presented (Ashby et al, 1998). Indeed, contrary to the predictions of COVIS, Dunn, Newell, and Kalish (2012) found that when using a mask dissimilar to the category stimuli in the delayed feedback paradigm presented above, a dissociation is no longer apparent between RB and II learning as neither RB or II performance is affected by delayed feedback presentation.

A dissociation between performance in RB and II tasks has also been reported by Filoteo et al. (2010). Participants learnt either three-dimensional conjunctive RB or

II categories either with, or without, a concurrent WM load. In the load condition participants performed a digit probe task where they were shown an array of numbers and, after a delay of 100ms, had to indicate if a probe was included in the original display. While the performance of RB participants was numerically better in the no-load condition than the load condition, participants learning the II condition with a concurrent load performed significantly better than II participants who did not complete a concurrent WM task. Filoteo et al. (2010) concluded that this was strong evidence for the COVIS model (Ashby et al., 1998). The taxation of WM would be predicted to improve II learning as it would enable the quicker abandonment of the default, suboptimal explicit system in the II condition and facilitate a switch to the more optimal implicit system that does not rely on WM (Ashby et al., 1998).

However, it has been highlighted that in Filoteo et al. (2010) the inter-trial interval (ITI) was 2500ms longer in concurrent load conditions than in the non-concurrent load conditions (Newell et al., 2013). Therefore, participants under concurrent load had more time to explicitly consider the stimulus they had been presented with and the feedback received than participants in the no-load conditions. Newell et al. (2013) replicated the II condition of Filoteo et al. (2010). The WM task in the load conditions was also manipulated to occur either before or after the categorization decision with long or short ITIs between trials. Participants who performed the task with a longer ITI, and therefore had more time to process feedback, were more accurate compared with participants in short ITI conditions, regardless of whether the WM task was completed before or after the categorization trial. This suggests that II participants in the concurrent load condition in Filoteo et al. may demonstrate better learning than II participants who did not perform the concurrent task because they had more time to explicitly process the feedback.

What is apparent by the review presented is that many behavioural dissociations seemingly supporting COVIS predictions can be accounted for by methodological artefacts (for example the use of a longer ITI in concurrent load conditions in Filoteo et al., 2010). When accounting for such methodological issues, differences in performance between RB and II conditions are not replicated (e.g., Newell et al., 2012; Newell et al., 2013) which prevents a strong argument being made for separable category learning systems. In the current research field, the results of behavioural dissociation work are equivocal.

Neuroimaging evidence investigating the COVIS model

The unclear conclusions drawn from behavioural category learning research warrant the application of different research methods to explore the processes behind RB and II category learning. As suggested by Worthy, Markman, and Maddox (2013) an alternative approach to resolving this dispute is to focus directly on the neurobiological predictions of COVIS, where it has been argued that single system accounts cannot explain the evidence that separable neural systems are engaged during different types of category learning.

The neurobiological evidence for COVIS is currently surprisingly limited, there has been a paucity of studies directly comparing the brain systems involved in RB and II category learning (although cf., Aizenstein et al., 2000; Seger, Dennison, Lopez-Paniagua, Peterson, & Roark, 2011; Seger et al., 2000, of imaging work of category learning in a different context). Perhaps the most prominent study to examine this, though, was by Nomura et al. (2007). Participants completed either an RB or an II category learning task inside an MRI scanner. The RB category structure had an easy to verbalize unidimensional rule (e.g., Figure 1.3.a), while the II structure was based on that shown in Figure 1.3.b. Nomura et al. considered their results to be in line with

This image has been removed by the author of this thesis for copyright reasons.

Figure 1.3. The category structures used by Nomura et al. (2007); (a) The RB structure; (b) the II structure. The lines represent the decision boundary separating the Gabor patch stimuli into two categories. Figure adapted from Nomura et al. (2007).

COVIS - dissociable neural activation was found with the MTL more activated in RB compared with II learning, and the caudate body more activated in II than RB learning. Further evidence of separable systems was found in a reanalysis of Nomura et al.'s data which modelled participants' decision strategies (Nomura & Reber, 2008). Participants using RB learning strategies showed greater right prefrontal cortex activity than those using II strategies, and those utilizing II strategies had greater right occipital activation.

More recently, Soto, Waldschmidt, Helie, and Ashby (2013; see also Helie, Waldschmidt, & Ashby, 2010; Waldschmidt & Ashby, 2011) directly contrasted RB and II learning in a multi-voxel pattern analysis. While the study had multiple training sessions, the first scanning session (Training session 1 for the RB task and Training session 2, following 600 training trials in Session 1, for the II task) prior to the development of automaticity, is most pertinent for the current issue. While there was common activation between RB and II learning (for example in the globus pallidus and the extrastriate visual cortex), there were some differences in activation. For instance,

consistent with Nomura et al. (2007), the head of the caudate was activated more in RB learning, while activation in the caudate body/tail also differed between the RB and II tasks. However, it is difficult to know whether these neural differences were due to the engagement of separate systems or whether they were due to participants in the II condition having already received 600 training trials previous to the scanning session while participants in the RB condition had no prior training (brain activation alters over a relatively limited number of trials, e.g., Koenig et al., 2005; Milton & Pothos, 2011). While this issue was not the sole focus of Soto et al.'s study it does, nevertheless, compromise any direct comparisons in brain activation between II and RB learning prior to automaticity developing.

Milton and Pothos (2011) found a different pattern of results to Nomura et al. (2007), observing extensive overlap in activation between a unidimensional RB structure and a complex category structure assumed to have many of the properties of II categories (e.g., optimal decision bounds that were difficult to verbalize). In contrast to Nomura et al., neural differences between the II and RB conditions were minimal and restricted to greater activation in a small region of the left superior frontal lobe in the complex condition relative to the RB condition. While intriguing, one should not draw too strong an inference about these findings with regard to COVIS due to the differences in the stimuli that Milton and Pothos used compared to those traditionally administered in COVIS research. For instance, there were only 18 unique stimuli, with dimensions that were commensurable (rectangle height and ellipse width) and a decision boundary that was arguably easier to verbalize than the II structures typically employed (e.g., Figure 1.1.c). These differences between the decision boundaries employed in Nomura et al. (2007) and Milton and Pothos (2011) could account for the discrepant results found. Nevertheless, these findings indicate that further direct comparison of the neural correlates of RB and II category learning is needed.

The neural dissociations found by Nomura et al. (2007) can be questioned based on methodological grounds. The critical difference between the RB and II category structures is often assumed to be that the RB structure is easily verbalizable but the II structure is not. While the RB and II category structures used by Nomura et al. (see Figures 1.3.a and 1.3.b) differ convincingly on this factor, there are also non-essential differences between them that may potentially be driving the discrepancies in activation. For instance, the RB structure has only one relevant dimension while the II structure has two relevant dimensions, which means that selective attention is required for the RB condition but not for the II condition (Edmunds, Milton, & Wills, 2015; Nosofsky & Kruschke, 2002). When accounting for this non-essential difference between RB and II conditions behavioural dissociations supporting COVIS have been found to vanish (e.g., Edmunds et al., 2015). This is a concern that has been acknowledged by some COVIS theorists (e.g., Nomura & Reber, 2008; Xie, Maddox, McGeary, & Chandrasekaran, 2015; Zeithamova & Maddox, 2006). On a different note, multidimensional categorizations are typically more complex and require greater cognitive resources than unidimensional categorizations (e.g., Milton, Longmore, & Wills, 2008; Wills, Inkster, & Milton, 2015) and may be less intuitive, and therefore more difficult to learn (e.g., Pothos, & Bailey, 2009; Pothos & Close, 2008)². This could potentially be driving the more pronounced caudate body activation in the II condition than the RB condition, particularly given that the involvement of the basal ganglia is thought to be greater for more complex structures (e.g., Ell, Weinstein, & Ivry, 2010; Filoteo, Maddox, Salmon & Song, 2005). As the II structure is often more difficult to learn than the RB structure (e.g., Ashby, Maddox, & Bohil, 2002, Maddox, Ashby, & Bohil, 2003), Nomura et al. reduced

² While this thesis operationalises rule difficulty by the number of stimulus dimensions relevant to learning and the perceptual discriminability of the categories, there are other factors that contribute to rule complexity. For example, how intuitive the category is (i.e., Pothos, Chater, & Hines, 2011), the number of categories to be learned (e.g., Ashby & Maddox, 1992), and intrinsic variations in the difficulty of particular multidimensional rules such as conjunctive and disjunctive structures (e.g., Bourne, 1970).

the category separation in the RB condition relative to the II condition to minimize any performance differences between conditions. While this successfully matched learning rates, it effectively replaces one confound with another because the optimal decision boundary is more difficult to perceptually discriminate in the RB condition than the II condition (Stanton & Nosofsky, 2007). This confound is potentially critical given that COVIS assumes that the MTL is responsible for storing the precise placement of the decision bound (Nomura & Reber, 2008). The greater activation in the MTL for the RB condition compared to the II condition could, therefore, be due to this difference in category separation.

Research aims of Strand 1

In considering the neuroimaging and behavioural literature, it is clear that the debate in the research field as to whether single or multiple systems are responsible for category learning is far from settled. What is apparent is that in order to draw strong comparisons about brain activation in RB and II category learning, it is necessary to control for non-essential differences between the category structures. This has been achieved in previous COVIS related research (e.g., Filoteo et al., 2010; Zeithamova & Maddox, 2006) - but in no previous imaging study - by comparing the II category structure to a conjunctive, RB category structure (see Figure 1.1.b). The II and the conjunctive category structures both possess two relevant dimensions, have a similar error rate (Filoteo et al., 2010) and are closely matched for category separation. The neuroimaging study (Experiment 1) presented in Chapter 2 aims to implement these category structures to investigate whether separable systems present in RB and II learning can indeed be inferred when using conjunctive RB categories, or whether the previous neural dissociations found in previous work (e.g., Nomura et al., 2007) can be explained by extraneous differences between conditions.

Chapter 3 builds upon the research of Chapter 2, using fMRI to investigate whether separable neural systems of category learning can be found during feedback processing using the same controlled conditions as Chapter 2. A behavioural study was completed in order to investigate whether delaying feedback presentation to decorrelate neural activation during the feedback processing stage from the category decision phase would differentially hinder RB or II learning. By delaying feedback in category learning trials II learning may be impaired as stimulus-response associations could be disrupted (e.g., Maddox et al., 2003; Maddox & Ing, 2005). Having established that delaying feedback did not differentially affect RB and II performance, an fMRI study was then conducted aiming to examine the neural systems activated during RB and II feedback processing.

Chapter 4 critically evaluates the claim of Smith et al. (2014) that deferring feedback provides some of the strongest evidence for dissociable learning systems. However, like the neuroimaging work that motivated the studies in Chapters 2 and 3, there are non-essential differences between the RB and II conditions used by Smith et al. (2014) which complicate the interpretation of their results. As in Nomura et al., (2007), the RB category used was a unidimensional category whereas the II condition was multidimensional. To date, the conclusions drawn from Smith et al. (2014) have not yet been critically considered based on their inclusion of an extraneous variable between the category structures. Therefore, I conducted a series of studies comparing performance in conjunctive RB and II tasks under deferred feedback to assess whether the dissociation found by Smith et al. (2014) can still emerge when the non-essential difference has been controlled for.

Finally, Chapter 5 uses tDCS to investigate the effect of anodal stimulation on RB and II performance. It has been claimed that by using anodal stimulation

participants' performance on cognitive tasks can be enhanced (e.g., Coffman, Trumbo, & Clark, 2012; Fregni et al., 2005). However, to my knowledge, there are currently no tDCS studies that investigate the effect of brain stimulation on a participants' ability to learn RB or II category structures. Chapter 5 aimed to see if RB or II learning would be differentially affected by anodal tDCS to the left dorsolateral prefrontal cortex (DLPFC) and more generally whether there would be any effect of tDCS on category learning.

Strand 2: The prototype effect

Strand 2, like Strand 1, focuses on better understanding the neural underpinnings of category learning. Strand 1 considers category learning using a very specific procedure; however, it is likely that categories are acquired in many different ways (Love, 2002) and that the type of material to be learnt may have an important influence on categorization behaviour. Strand 2, whilst keeping with the general theme of examining trial-by-trial supervised learning, aimed to explore the neural processes of category learning from a different angle to build a broader, more nuanced picture of category learning systems. To do this, I looked at the classic prototype effect (e.g., Posner & Keele, 1968) from a neuroscientific perspective. In this behavioural effect, novel prototype stimuli are more accurately responded to compared with previously seen and novel exemplars derived from that prototype (e.g., Minda & Smith, 2001; 2002; Posner & Keele, 1968; Smith, Redford & Hass, 2008).

There are at least two types of theories that have been argued to account for behavioural prototype effects. Both suggest that items to be classified are compared to stored representations of a category. If perceived items reach a similarity threshold to that stored representation, then the item is treated as a member of that category (Minda & Smith, 2011). However, the theories differ on the form of the representation to which a novel stimulus item is compared. Exemplar theories, such as the Generalized Context

Model (Nosofsky, 1985; 1986), suggest that all experienced stimuli of a category are stored in memory, and the weighted similarity between the novel exemplar and a group of stored representations is computed in order to make the categorization decision. Exemplar theories such as this would account for prototype effects by arguing that prototype stimuli are more similar to stored exemplar representations than a novel exemplar stimulus, which would be more likely to depart from the central tendency. On the other hand, prototype models state that when perceiving a category of exemplars, the most common features of the stimuli are abstracted to form a prototypical representation of the category. Novel prototype stimuli are more accurately responded to compared with exemplars as they more closely match the single category representation stored in memory (e.g., Minda & Smith, 2001; 2002; Smith, et al., 2008).

The behavioural prototype effect can be investigated using an (A, B) prototype distortion task. In such tasks a prototype is constructed for each category, for example a random nine dot pattern (e.g., Posner & Keele, 1968). Exemplars of each category are then generated by distorting the prototype. In (A, B) tasks two or more categories are generated that are derived from different prototypes. The prototype derived exemplars are presented to participants in a training phase. Through trial and error participants are instructed to learn which stimulus belongs to which category. After participants have been trained on the categories they complete a test phase. They are shown exemplars from each category along with the prototypes which they must categorize based on what they learnt in training. The participant's accuracy in responding to this previously unseen prototype is compared to responding to exemplars that are distortions of the category prototype (e.g., Heindel, Festa, Ott, Landy, & Salmon, 2013; Posner & Keele, 1968).

In the literature, participants demonstrating a prototype effect more accurately categorize the previously unseen category prototype of a trained category than previously seen or unseen exemplars of the category (e.g., Metcalfe & Fisher, 1986; Zaki, Nosofsky, Jessup, & Unverzagt, 2003), or are more likely to falsely recognize a category prototype as a previously seen exemplar (e.g., Posner & Keele, 1968). The prototype effect has been found using a variety of stimuli, for example dot patterns (Posner & Keele, 1968), and faces (Cabeza, Bruce, Kato, & Oda, 1999). When using abstract shapes it has also been found that the prototype effect is maintained over four days, whereas performance on trained exemplars deteriorates over this period (Homa, Cross, Cornell, Goldman, & Schwartz, 1973).

Performance on (A, B) prototype distortion tasks has been found to be affected in patients. For example, Alzheimer patients were impaired on exemplar classification at test compared with healthy controls; however, they still demonstrated a prototype effect in accuracy scores (Heindel et al., 2013). Yet, some have argued that patients can retain the ability to respond accurately to novel exemplars as well as the category prototype when compared with healthy controls after training on a prototype distortion task (e.g., in amnesic patients, Knowlton & Squire, 1993; and in Parkinson patients, Reber & Squire, 1999).

Another prototype distortion task used in the literature is the (A, not A) task. Here, only one category is learnt during training, and at test participants then categorize previously seen exemplars from training, previously unseen (novel) exemplars, the previously unseen prototype, as well as randomly generated exemplars as being members of the trained category or not (Ashby & O'Brien, 2005). However, there are concerns raised that the (A, not A) task is too easy. For example, it has been found that participants can demonstrate category learning in an (A, not A) test phase without

having completed training (Palmeri & Flanery, 1999). Therefore (A, B) tasks will be the focus of the research in Strand 2. This introduction will now review the existing neuroimaging and tDCS evidence which was the basis of the experiments I conducted which used a prototype distortion task.

Neuroimaging research

Several studies have examined the neural substrates of learning in the prototype distortion task. However, only some of these use the (A, B) task. For example, Seger et al. (2000) examined the neural areas implicated by (A, B) prototype distortion tasks by presenting participants with two prototype derived visual grid pattern categories to learn through trial and error in an MRI scanner. Participants were only presented with exemplar stimuli and never shown the prototype of either category. Seger et al. found that the occipital lobe, the right DLPFC and inferior parietal regions to be active when learning to classify exemplars compared with a baseline task (random button press to a random grid pattern). The authors proposed that this network constitutes a visual reasoning function operational in the prototype distortion task. However, during later learning trials participants who classified the exemplars more accurately demonstrated bilateral DLPFC activation. The authors suggest that the activation of the left DLPFC suggests that analytical problem solving with verbal elements are implemented more in participants who perform more accurately at test. Similar to this work, (A, not A) tasks have also been found to implicate the occipital cortex (e.g., Aizenstein et al., 2000; Reber, Stark, & Squire, 1998). However, the more frontal and parietal systems found active in studies such as Seger et al. (2000) seem to be more specific to (A, B) task performance.

More recently, Zeithamova, Maddox, and Schnyer (2008) measured neural activation during an (A, B) task. The parahippocampal gyrus, the inferior parietal cortex, and the orbitofrontal cortex were found to be activated during the test phase of this task. Greater activation in the MTL and the orbitofrontal cortices was also found to be correlated with more accurate responding in the (A, B) test phase. These findings led to the conclusions that an episodic and MTL based mechanism is functioning during (A, B) learning and that this may represent associative or declarative memory processes. Although Seger et al. (2000) and Zeithamova et al. (2008) both used the (A, B) prototype distortion task, the prototype effect was not measured in either study and the brain activation during prototype presentation was not analysed. To my knowledge the neural substrates correlated with prototype presentation and with the prototype effect have not yet been investigated.

TDCS and the prototype distortion task

TDCS is a non-invasive brain stimulation technique that allows researchers to use small electric currents to modulate the intracerebral current flow under electrode sites to modify behaviour (Nitsche & Paulus, 2011). A common finding is that anodal tDCS can affect performance on cognitive tasks by modulating cortical excitability (e.g. Kuo et al., 2008). By using tDCS, results from neuroimaging work can be further explored – for example, the left DLPFC is found to be active during prototype learning in an fMRI study (e.g., Seger et al., 2000); if when using tDCS to stimulate this area a change in learning is found, the DLPFC can be more confidently linked to performance in prototype distortion tasks. However, as in the COVIS literature field, there are very few tDCS studies that investigate the effect of brain stimulation on the prototype effect. The current studies that utilize tDCS during prototype distortion learning present contradictory conclusions.

Ambrus et al. (2011) stimulated the left DLPFC with 1mA of cathodal or anodal stimulation for 10 minutes, or for 30 seconds (the sham conditions). This stimulation began 8.5 minutes before the training phase in an (A, not A) prototype distortion task. Participants were trained on high and low distortion exemplars of category A. During test they were then presented with high and low distortion exemplars as well as the prototype from category A and random dot patterns. Participants indicated if the stimulus was a member of the trained category. Categorization accuracy was lower in participants who had received anodal stimulation compared to those in sham or cathodal conditions. However, Ambrus et al. (2011) also found that in the anodal condition, participants did not display the standard prototype effect that was present in sham participants and that they were significantly worse than sham participants in accuracy of responding to the prototype. Anodal stimulation abolished the prototype effect that was present in sham participants.

Although Ambrus et al. (2011) acknowledged that the tDCS could simply be adding neural noise to the systems operational during task performance, the timing of the stimulation in Ambrus et al. (2011) could potentially account for the pattern of results found. Kuo et al. (2008) discovered that when participants received tDCS to the motor cortex during a motor task their performance was increased compared with controls. However, if the stimulation was received before the motor task then their performance was worse than controls. This could account for the impaired performance of anodal participants' in Ambrus et al. as the majority of the stimulation was given prior to training, with just 1.5 minutes synonymous with the task.

However, when presenting participants with an (A, B) prototype distortion task with training beginning after just 1.5 minutes stimulation, McLaren et al. (2016; Experiment 2a) found a different pattern of results. Participants received 1.5mA of

This image has been removed by the author of this thesis for copyright reasons.

Figure 1.4. Results of McLaren et al. (2016, Experiment 2a). The graph shows mean accuracy during test for responding to old and new exemplars and prototype stimuli. The chequerboards represent typical exemplars or the prototype. Figure taken from McLaren et al. (2016).

anodal or sham stimulation to the left DLPFC during training. Participants learnt to differentiate between two chequerboards categories derived from two prototypes. They were then tested on their categorization accuracy of previously seen and unseen (novel) exemplars and the two prototype chequerboards. McLaren et al. (2016) found, as in Ambrus et al., that participants in the anodal condition performed worse on overall categorization accuracy than sham participants (see Figure 1.4). However, unlike Ambrus et al., those receiving anodal stimulation also displayed a prototype effect. Participants responded more accurately in categorizing the previously unseen prototype than unseen and seen exemplars. This effect was not apparent in sham participants where no significant prototype effect was found (however, the magnitude of the prototype effect was not found to be significantly different between sham and anodal

participants). McLaren et al. concluded that there was a prototype effect present in anodal participants that was not apparent in sham participants. However, it should be noted that there is an apparent ceiling effect in McLaren et al. (2016; see Figure 1.4). In the sham condition of Experiment 2a performance is highly accurate in categorizing previously seen, unseen and prototype exemplars. It is therefore possible that there was a prototype effect present in the sham conditions that was masked due to the excellent performance in the exemplar and prototype trials.

Although the results of these two studies are contradictory, the research cannot be truly compared. While Ambrus et al. (2011) use an (A, not A) task, McLaren et al. employ an (A, B) task. The fundamentally different task structures could account for the discrepancies in results, and there are concerns whether the (A, not A) task genuinely displays learning (Palmeri & Flanery, 1999). These two studies highlight the necessity of further research to investigate whether the prototype effect is enhanced or decreased by anodal stimulation to the left DLPFC. Gaining more conclusive results can give greater insight into the neural substrates that are responsible for the prototype effect.

Research aims of Strand 2

Strand 2 of this thesis, presented in Chapter 6, aims to investigate the effect anodal tDCS has on the prototype effect by using a method that considers the critique of Ambrus et al. (2011) that the timing of the stimulation may account for the decline in performance (Kuo et al., 2008). The potential ceiling effect in McLaren et al. (2016; Experiment 2a) will also be accounted for. The paradigm used in Strand 2 aims to make the task more complex to better investigate the potential of a prototype effect in both anodal and sham stimulation conditions, with the majority of stimulation occurring during the training phase of the task.

The final experiment of Strand 2 will be a joint tDCS and fMRI study investigating the effect of anodal stimulation on a prototype distortion task. Such a study has not been performed in the current research field and will not only inform the neural areas active to prototypes in comparison to exemplars, but also compare neural activation between those receiving anodal tDCS and those in sham conditions. Also, one area, to my knowledge, seemingly not investigated by previous fMRI studies using the prototype distortion task is the neural response to the presentation of the prototype in comparison with the previously seen and unseen category exemplars. It is surprising that the prototype effect, an effect so robust in the behavioural literature, has not been directly examined using fMRI methods. For example, in Seger et al. (2000) participants were not presented with the prototype at all. Neural activation was analysed by comparing all categorization trials to a baseline task (button press to a random stimulus). In Zeithamova et al. (2008) the prototype was presented, however the neural activation during this was not analysed. The prototype for each category was only presented once at test, so there would not be enough power in the analysis to infer if any areas were more or less active during prototype presentation than exemplar presentation. Therefore, this study aims to investigate whether there is a distinctive neural response to the prototype effect by presenting the prototype repeatedly to participants during test.

Chapter 2. Investigating the neural correlates of rule-based and information-integration category learning with fMRI

As discussed in Chapter 1, there are currently very few studies that directly compare neural activation during RB and II category learning. One study that examined the neural areas implicated during II and RB category learning was conducted by Nomura et al. (2007). Participants in the unidimensional RB condition demonstrated more MTL activation than II participants. In contrast, II participants had greater activation in the body of the caudate nucleus than RB participants. However, as discussed in Chapter 1, there were non-essential differences between the RB and II conditions that may potentially account for these neural dissociations. Specifically, the learning conditions were not matched on category separation (Stanton & Nosofsky, 2007), with the RB stimuli falling closer to the decision boundary than the II stimuli (see Figure 1.3). This may account for the greater MTL activation present in the RB condition as this area is suggested to be involved in the placement of the decision boundary (Nomura & Reber, 2008). Similarly, the II condition had two dimensions relevant for category learning compared with just one necessary in the unidimensional RB condition, which may have resulted in the increased caudate body activation in the II condition as the basal ganglia is found to be more engaged when learning complex category structures (e.g., Ell et al., 2010). Both of these methodological artefacts may have contributed to the neural activation patterns presented in Nomura et al. (2007) as discussed on pages 27-28.

Experiment 1³ aims to complete a comparison of the neural correlates of RB and II learning by employing category structures that control for the non-essential differences highlighted above which may have been driving the differential pattern of

³The experiment presented in Chapter 2 has been published in 'Human Brain Mapping' (Carpenter, Wills, Benattayallah, & Milton, 2016).

This image has been removed by the author of this thesis for copyright reasons.

Figure 2.1. The category structures used in Filoteo et al. (2010); (a) the conjunctive rule-based condition; (b) the information-integration condition. Solid lines indicate the decision boundary separating category A and category B.

activation in Nomura et al. (2007). In the present fMRI study participants will learn one of the conjunctive RB or II categories presented in Figure 2.1 (Filoteo et al., 2010). The conjunctive RB and II categories are matched in terms of category separation, number of relevant dimensions necessary for optimal category learning as well as rates of learning, therefore controlling for difficulty differences between conditions (Filoteo et al., 2010) better than the stimuli used by Nomura et al. (2007). This will enable a more rigorous assessment of COVIS's predictions.

Another notable aspect of Nomura et al.'s (2007) study is their use of incorrect trials as the baseline comparison for correct responses. While this is a convenient baseline to use and has been employed in other categorization research (e.g., Milton & Pothos, 2011) it may not be the most effective due to difficulties in interpreting what is driving the incorrect response. First, participants may have been using the correct general strategy but not identified the relevant dimension/precise category structure; for example, participants used a rule-based strategy but categorized by orientation rather

than line length. Second, participants might have used the appropriate stimulus dimension but placed the decision boundary in the incorrect place. Third, and less commonly, participants may have classified correctly but pressed the wrong button. Fourth, participants may have been guessing or not fully engaged on the trial and fifth, participants could have used a completely different strategy to what was appropriate. It is likely that the errors are a combination of these (and potentially other) mistakes but it is not possible at the individual trial level to determine the source of the error. The first three of these error types appear particularly problematic as they would result in similar brain activation to correct trials meaning that this is unlikely to be a sensitive baseline. Furthermore, comparing correct and incorrect trials is likely to be confounded with degree of learning as there will be more incorrect trials early in training than later in training. This is particularly an issue when wishing to make inferences across the whole of training, as is typically the case.

While the main analyses will be presented with this "incorrect" baseline to aid comparison of these results with Nomura et al. (2007), an "odd-or-even" task will be the principal baseline in this study. This type of control is increasingly being used in imaging studies of categorization (e.g., Davis, Love, & Preston, 2012a; Davis, Love, & Preston, 2012b; Davis, Xue, Love, Preston, & Poldrack, 2014) and while it may superficially seem similar to a RB task (albeit one that is highly automated and engages limited neural resources, Stark & Squire, 2001) its main advantage is that it is well established that it does not recruit the MTL or indeed the frontal lobes (Stark & Squire, 2001), the pivotal regions of COVIS's rule-based, explicit system. Equally, activation in the striatum, the key site of COVIS's implicit system, is also readily identified with an odd-or-even baseline task (Zink, Pagnoni, Chappelow, Martin-Skurski, & Berns, 2006). Therefore, Experiment 1 will implement this baseline task interleaved with category learning blocks, which should provide a clear measure of the regions engaged in both

RB and II categorization without the involvement of key regions being obscured by their activation in the baseline task.

According to how COVIS is often conceptualized (e.g., Nomura et al., 2007), one might predict greater activation in the caudate head, the anterior cingulate, prefrontal cortex, and the MTL (and in particular the hippocampus) when learning a conjunctive RB structure compared to learning an II structure (Ashby & Valentin, 2005). In contrast, greater activation should be found in the body/tail of the caudate, the putamen, and the substantia nigra for the II condition compared to the RB condition (Ashby & Valentin, 2005). Conversely, if Nomura et al.'s (2007) results were driven by one of the non-essential differences between the RB and II structures outlined above then, when these variables have been better controlled, one might expect that these neural differences would disappear leaving an extensive overlap of activation. Further, given that the II structure is harder to verbalize than the RB structure and yet categorization accuracy is the same (Filoteo et al., 2010), greater activation might be expected in the prefrontal cortex for the II compared to the RB condition to reflect the greater processing demands of finding and applying a less easy to verbalize rule. Alternatively, or perhaps additionally, the MTL may be more engaged in the II condition than the RB condition if the lower levels of verbalizability lead to an increase in memory demands to store exceptions in decision space to the rule that is utilized (Davis, Love, & Preston, 2012a).

Method

Participants and design

45 right handed University of Exeter students (26 female, 19 male) with normal or corrected vision completed the experiment for £5 remuneration. Participants were

randomly allocated to one of two between-subject conditions (RB or II). One participant from the RB condition was excluded for failing to reach 50% (chance) accuracy in the final run (although the inclusion of this participant does not alter any of the conclusions of this study), leaving 22 participants in each condition. Participants gave informed consent according to procedures approved by the University of Exeter's School of Psychology Ethics Committee.

Stimuli

The stimuli (see Figure 2.2) were a subset of the two-dimensional II stimuli and conjunctive RB stimuli (where short, upright lines belong in category A, and the rest in category B) employed by Filoteo et al. (2010). In the original data set there were 600 stimuli in both conditions; in the present imaging study, 320 of these stimuli were randomly selected (160 stimuli in each category) for each category structure. This number of stimuli was the same as used by Nomura et al. (2007). Each stimulus was a black line varying on two dimensions: length and orientation. The stimuli used are likely to be separable (where dimensions can be processed independently) rather than integral (where the dimensions are perceived as a unitary whole) according to Garner's (1974) criteria. The stimuli were originally constructed by Filoteo et al. to be of approximately equal salience and although this is not something that I directly examined, the similar prevalence of unidimensional strategies using the length and orientation dimensions suggests that this is indeed the case⁴.

⁴ The grt modelling analysis demonstrates the stimulus dimension that people are using when implementing unidimensional strategies to form a decision boundary. Overall, 6 people used line orientation to form a decision boundary, and 4 people line length when considering all learning trials in the analysis. However, when modelling decision boundaries used during blocks 3 and 4 alone, 4 people used line orientation and 6 people used line length to make category decisions.

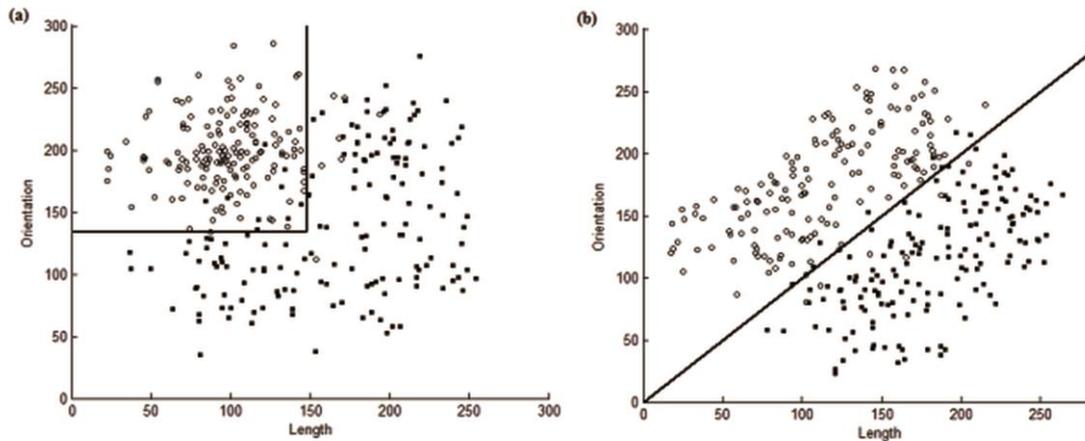


Figure 2.2. The category structures used for the present study; (a) the conjunctive rule-based condition; (b) the information-integration condition. Solid lines indicate the decision boundary separating category A (unfilled circles) and category B (filled squares).

As in Filoteo et al. (2010), there was 5% overlap between the categories so that the maximum accuracy attainable was 95%.

FMRI imaging

A 1.5-T Phillips Gyroscan magnet, equipped with a Sense coil, was used to collect images from each participant in one scanning session. A T2*-weighted echo planar sequence (TR = 3000ms, TE = 45ms, flip angle = 90°, 36 transverse slices, 3.5 x 2.5 x 2.5mm) was used. Upon entering the scanner, the participant's head was secured in place with foam pillows inside the coil to prevent excessive head movement. Participants completed four runs, each containing 205 scans. Five “dummy scans” were completed before every run prior to presentation of the first trial. After the functional scans, standard volumetric anatomical MRI was completed using a 3-D T1-weighted pulse sequence (TR = 25ms, TE = 4.1ms, flip angle = 30°, 160 axial slices, 1.6 x 0.9 x 0.9mm).

Procedure

In each scanning run, participants performed two interleaved tasks - the category learning task and an “odd-or-even” baseline task. Each run began with 15 odd-or-even trials, followed by two blocks of 40 categorization trials. Each run then concluded with another block of 15 odd-or-even trials. After each block there was a blank screen of 8000ms during which time participants were asked to rest. In total, there were 320 category learning trials, presented in a random order, and 120 odd-or-even trials. The stimuli were presented on a back projection screen positioned at the foot end of the MRI scanner and viewed via a mirror mounted on a head coil. Responses were measured using a fiber-optic button box held in the participants’ left and right hands. E-Prime (Psychological Software Tools, 2002) was used for the presentation and timing of stimuli and collection of response data.

In the *category learning task*, participants were informed that they had to learn into which of two categories a series of stimuli belonged. The trial-by-trial procedure for the RB and II conditions was identical. Each trial began with a blank screen lasting a variable interval between 500ms and 4000ms, followed immediately by a black fixation cross presented in the centre of the screen for 250ms. A stimulus then appeared in the middle of the screen for 2000ms during which time participants were required to respond by pressing the far right button on the button box with their right hand if they thought the item belonged to category A or the far left button with their left hand if they thought the item was a member of category B. Feedback ("Correct" or "Incorrect") was then displayed for 500ms. If participants did not respond in time the message "Time out!!!" appeared on the screen for 500ms instead. The next trial then immediately began.

The odd-or-even task was closely modelled on that used by Stark and Squire (2001; see also Davis et al., 2012a) and had a similar trial-by-trial structure to the category learning task. Each trial began with a blank screen lasting between 500-4000ms, followed by a black fixation cross for 250ms. A randomly generated number from one to nine then appeared in the middle of the screen for 2000ms during which time participants had to press the far left button if the number was even or the far right button if it was odd. Following this, feedback ("Correct" or "Incorrect") was presented for 500ms or if participants did not respond in time a message saying "Time out!!" appeared during this interval.

FMRI data analysis

Data analysis was performed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>). Functional images were corrected for acquisition order, realigned to the mean image, and resliced to correct for motion artifacts. The realigned images were coregistered with the structural T1 volume and the structural volumes were spatially normalised. The spatial transformation was applied to the T2* volumes which were spatially smoothed using a Gaussian Kernel of 8mm full-width half maximum. Data were high-pass filtered (128s) to account for low frequency drifts.

Random effect whole brain analyses were completed using the general linear model with a combined statistical threshold of $p < .001$ (uncorrected) and a voxel threshold of 27 contiguous voxels, which together produce an overall corrected threshold of $p < .05$, according to AlphaSim, as implemented in the REST toolbox (Version 1.8; Song et al., 2011). Correct trials, incorrect trials, and timeouts were all included as separate regressors in the model. A canonical hemodynamic response function (HRF) together with temporal and dispersion derivatives was used to model the blood oxygen level-dependent (BOLD) response and the six head movement parameters

were included as covariates. My analyses focused on comparing correct categorization trials (for the RB and II groups separately) to the odd-or-even baseline task (although for the principal analyses correct categorization trials are also compared to incorrect trials). In addition, to measure common activation between the RB - baseline contrast and the II - baseline contrast, a conjunction analysis was performed. The contrasts were combined using a logical 'and' function through the minimum statistic to the conjunction null hypothesis (MS/CN; Nichols, Brett, Andersson, Wager, & Poline, 2005) technique implemented in SPM8. Both these contrasts were again conducted with a combined threshold of $p < .001$ (uncorrected) and a cluster threshold of 27; note that this approach is highly conservative because it reveals only those regions significantly activated for both the RB ($p < .05$, corrected) *and* the II ($p < .05$, corrected) conditions. Normalised MNI space coordinates were transformed to Talairach space (<http://imaging.mrcctu.cam.ac.uk/imaging/MniTalairach>) to establish activation sites as per the atlas of Talairach and Tournoux (1988).

Results

Behavioural analysis

The mean categorization accuracy across all runs for both the RB and II conditions is displayed in Figure 2.3.a. A 4 x 2 mixed-design analysis of variance (ANOVA) was conducted; the within-subjects factor was 'run' (4 levels) and the between-subjects factor was categorization task (RB/II). There was a highly significant effect of run ($F(3,126)=12.47$, $p<.001$, $\eta^2_p = .229$) indicating that performance improved with practice. There was, however, no significant difference between the II and RB conditions in accuracy ($F(1,42)=.14$, $p=.708$, $\eta^2_p = .003$, $BF = 1.04$) and no significant interaction between run and categorization task ($F(3,126)=1.71$, $p=.169$, $\eta^2_p = .039$, $BF = 2.00^5$).

COVIS would predict that RB participants should improve in their performance markedly when the correct rule is hypothesized and applied (Ashby et al., 1998). The behavioral analysis was therefore also completed using only the data from participants who used the optimal learning strategy in the RB (a conjunctive strategy) and the II (an information-integration strategy) conditions. The average accuracy of responding for each run is shown in Figure 2.3.b. Again, there was a main effect of run suggesting participants improved as the study proceeded ($F(3,69)=9.39$, $p<.001$, $\eta^2_p = .29$). As in the analysis containing all subjects, the main effect of category condition ($F(1,23)=.94$, $p=.342$, $\eta^2_p = .04$, $BF = 1.47$) and the interaction between run and category learning

⁵ Bayes Factor analysis requires an estimate of the mean expected difference under the experimental hypothesis; I estimated this from Filoteo et al.'s (2010) study, which used the same stimuli and category structures, using plot digitizer (<https://sourceforge.net/projects/plotdigitizer/>). Following Dienes (2011), the expected difference was modelled as a two-tailed normal distribution with a standard deviation equal to half the mean. By convention, a Bayes factor of over three is interpreted as providing substantial evidence for the experimental hypothesis (Jeffreys, 1961), while a Bayes factor below a third provides substantial evidence for the null (Dienes, 2011). A value in between a third and three is indeterminate, providing no clear evidence either for the null or the experimental hypothesis.

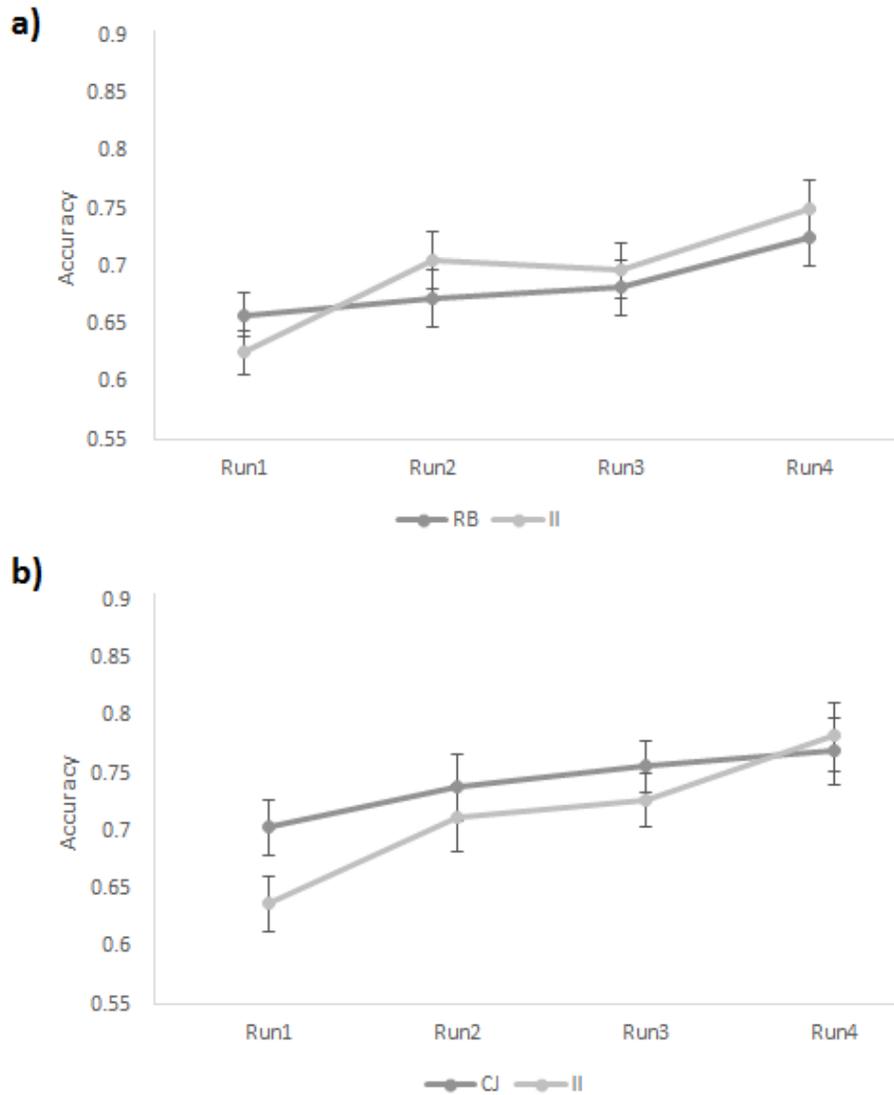


Figure 2.3.(a) Mean performance across runs in the RB and II conditions; (b) Mean performance across runs in the RB and II conditions for participants who used the ‘optimal’ learning strategy (conjunctive (CJ) in the RB condition and information-integration in the II condition). Error bars show standard error.

condition ($F(3,69)=1.28$, $p=.289$, $\eta^2_p = .05$, $BF = 1.4$) were non-significant. It would appear that participants using the conjunction strategy do not display marked improvement in categorization in comparison to those using an II strategy.

Imaging analysis - ‘Odd-or-even?’ baseline measure

Analysis of all runs

Whole brain activation across all runs of the category learning task was first analysed for participants in the RB and II conditions separately. Correct RB categorizations led to an extensive pattern of activation (Figure 2.4.a) including diverse areas of the frontal cortex (including BA’s 6, 8, 10, 45, 46, and 47), the anterior cingulate, posterior cingulate, the MTL, the bilateral caudate head/body, the putamen, the bilateral inferior and superior parietal lobes, the right superior temporal gyrus, bilateral inferior temporal gyrus and the bilateral occipital lobes. II category learning also activated these same brain regions (Figure 2.4.b).

I also examined whether there were any changes in activation across time for both RB and II learning. To assess this, activation in the first half of the experiment (runs 1 and 2) was directly compared to activation the second half (runs 3 and 4) for the RB and II conditions separately. No brain regions were more activated in the first half of training compared with the second half of training in either condition. No brain areas were engaged more in the second half of training than the first half in the RB condition either. However, in the II condition several regions including the right MTL (BA 30; see Table 2.1) were activated more in runs 3 and 4 than in runs 1 and 2.

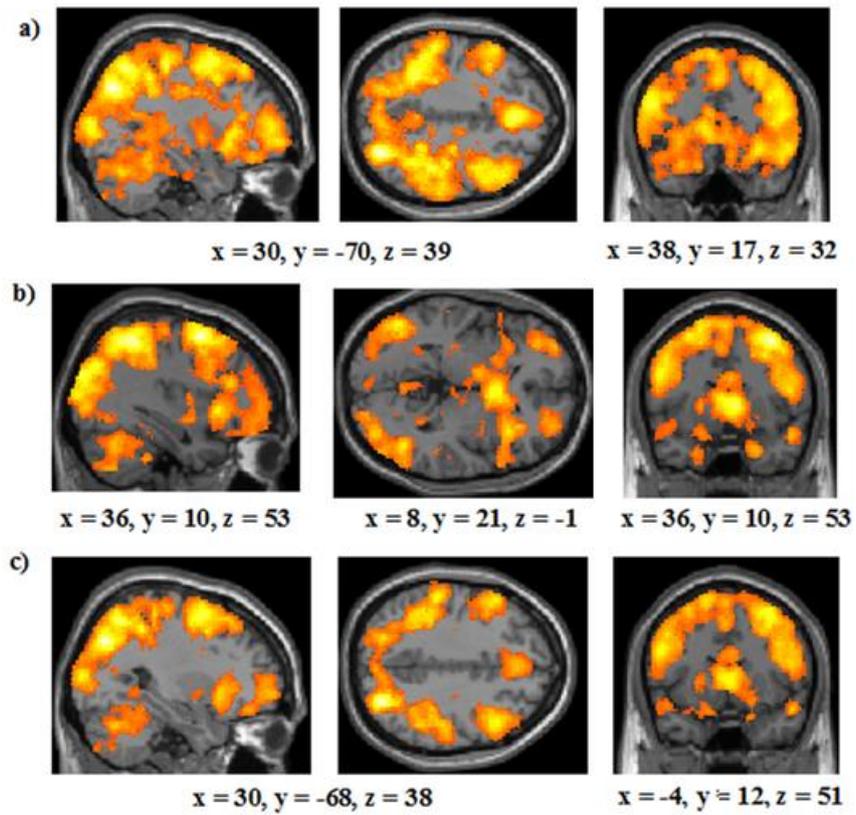


Figure 2.4. Whole brain analyses of all runs of the study; (a) areas of activation in the RB condition; (b) areas of activation in the II condition; (c) a conjunction analysis showing areas commonly activated in the RB and II conditions. All analyses are thresholded at $p < .001$ and 27 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Table 2.1.

Brain Regions Activated More in the Last Runs of the Study (Runs 3 and 4) than the First Runs of the Study (Runs 1 and 2) in the Information-Integration Condition.

Region	Cluster Size	BA	Talairach Coordinates			z-score
			x	y	z	
Right Precentral Gyrus	30	4	26	-25	53	3.84
Right Posterior Parahippocampal Gyrus	36	30	16	-43	4	3.80
Right Thalamus	52	-	18	-21	-1	3.78
Right Substantia Nigra		-	16	-18	-8	3.71
Left Paracentral Lobule	36	3	-18	-38	55	3.56
Left Postcentral Gyrus		5	-20	-41	65	3.48

Note that indented rows indicate voxels in the same cluster as the non-indented row above

The striking overlap in activation between the tasks was confirmed in a conjunction analysis, looking at common activation across the correct RB - odd-or-even contrast and the correct II - odd-or-even contrast (both with thresholds of $p < .001$ and 27 contiguous voxels; Figure 2.4.c). Areas engaged included key regions of both the explicit and implicit systems of COVIS. Regions linked to the explicit system that were recruited in both tasks were the MTL, the bilateral caudate head, diverse bilateral areas of the prefrontal cortex (including BA's 6, 8, 10, 46, and 47) and the bilateral anterior cingulate (right BA 25, left BA 33). Areas implicated in the implicit system that were engaged included the bilateral caudate body and the bilateral putamen. When contrasting incorrect trials to the odd-or-even task, a similar, if somewhat less extensive, pattern of activation was found including the right caudate body, right putamen and bilateral caudate head (Figure 2.5). This pattern of activation is consistent with the argument put forward in the introduction that incorrect trials are likely to share a similar neural substrate to correct trials.

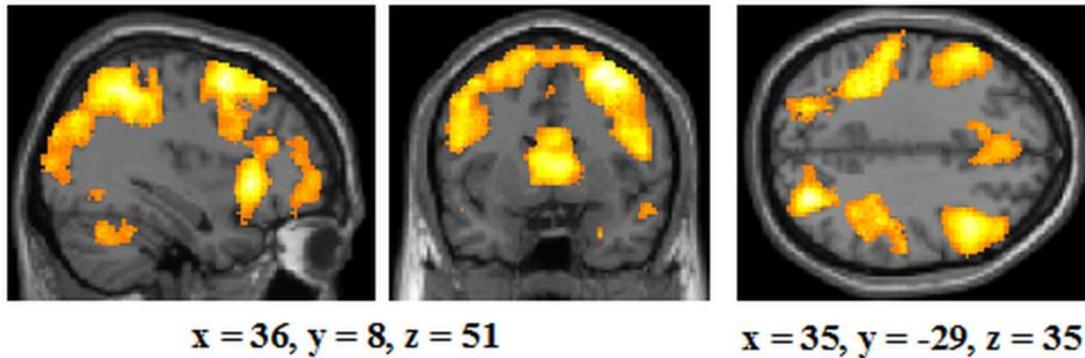


Figure 2.5. Whole brain analysis of all runs comparing incorrect trials to the ‘odd-or-even’ baseline, showing areas of common activation in the II and RB conditions thresholded at $p < .001$ and 27 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Next, areas engaged in the RB and II conditions were directly contrasted to examine whether there was evidence for the neural dissociations observed by Nomura et al. (2007). No regions were more active in the RB condition than the II condition (calculated by subtracting correct II trials - the odd-or-even trials from correct RB trials - the odd-or-even trials). However, diverse regions were more active in the II condition than the RB condition (calculated by subtracting correct RB trials - the odd-or-even trials from correct II trials - the odd-or-even trials; see Table 2.2, Figure 2.6.a). Critically, this included extensive activation in the left MTL (hippocampus/ posterior parahippocampal gyrus; 131 voxels; see Figure 2.6.b for areas of the MTL engaged). The results of these analyses are contrary to the predictions of COVIS, where the MTL is thought to be more critical for RB rather than II learning (e.g., Ashby & Valentin, 2005; Nomura et al., 2007).

Table 2.2.

Brain Regions Activated More in the Information-Integration Condition than the Rule-Based Condition in All Runs.

Region	Cluster Size	BA	Talairach Coordinates			z-score
			x	y	z	
Left Hippocampus	212	-	-22	-39	-1	4.02
Left Parahippocampal Gyrus		36	-26	-39	-5	3.78
Left Posterior Cingulate		29	-14	-46	10	3.30
Left Superior Temporal Gyrus	41	21	-52	-6	-11	3.61
Left Precuneus	80	31	-18	-47	32	3.91
Left Cingulate Gyrus		31	-16	-43	34	3.70
Left Thalamus	56	-	-14	-28	16	3.57

Note that indented rows indicate voxels in the same cluster as the non-indented row above

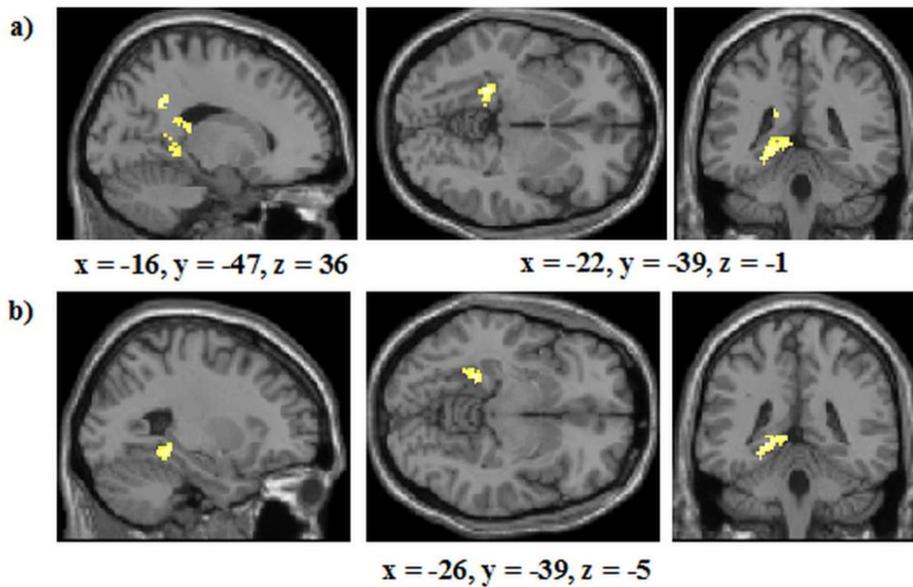


Figure 2.6. Analyses of areas more activated in the II condition compared to the RB condition in all runs of the study; (a) whole brain analysis; (b) regions of the MTL more engaged; non-MTL regions were masked in this analysis but the thresholds remained $p < .001$ and 27 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

However, in spite of the generally greater activation in the II condition compared to the RB condition, no regions associated with the implicit system of COVIS were identified in this analysis. Of course, it is possible that, even though this study had almost double the number of participants that Nomura et al. (2007) used (they had 13 in their II condition and 12 in their RB condition), this activation might have been present but below the a priori statistical thresholds. To provide greater sensitivity I, therefore, conducted a region of interest (ROI) analysis using the WFU PickAtlas (Maldjian, Laurienti, Burdette, & Kraft, 2003) comprising the caudate body, the putamen, and the substantia nigra with the more liberal thresholds of $p < .005$ and 10 contiguous voxels (the same thresholds used in previous ROI analyses, c.f., Milton, Butler, Benattayallah, & Zeman, 2012; Milton, Muhlert, Butler, Benattayallah, & Zeman, 2011). There was again no evidence found for greater activation in the II condition than the RB condition in these regions. To further confirm this conclusion, the relative percent signal change of correct RB and II responding was examined in the caudate body based on the peak right ($x = 17, y = -11, z = 28$) and left ($x = -20, y = -14, z = 29$) caudate body activations reported by Nomura et al. (2007). These percent signal change values were obtained using the Anatomy toolbox (Eickhoff et al., 2007; Version 2.2). Using independent samples t-tests, no significant difference was found between conditions for either the right caudate body ($t(42)=1.05, p=.300, d = .32^6, BF=.84$) or for the left caudate body ($t(42)=1.00, p=.323, d = .3, BF=.91^7$).

⁶ Throughout this thesis, the calculation of Cohen's d used was $d = M_1 - M_2 / SD_{pooled}$, in which the mean for each trial type was the measure of variability. A pooled standard deviation of the sample is used, as the standard deviations are presumed to be estimates of the same population (Olejnik & Algina, 2000). Here, the comparisons are made where $M_1 = RB$ $M_2 = II$.

⁷ The percent signal change in the right caudate body for the RB and II conditions in Nomura et al.'s (2007) study (shown in their Figure 4.d) was used to calculate the prior (these values were estimated using plot digitizer <https://sourceforge.net/projects/plotdigitizer/>). The expected difference was modelled as a two-tailed normal distribution with a standard deviation equal to half the mean (Dienes, 2011).

Analysis of runs 3 and 4 only

COVIS can potentially explain this pattern of extensive neural activation overlap during RB and II learning by assuming that for the II condition as well as the RB condition the explicit system dominates initially, and participants in the II group only switch to the implicit system once there has been sufficient time for the RB system to be proven ineffective (e.g., Filoteo et al., 2010). Including the initial trials in the analysis could therefore be obscuring neural differences that emerge later in learning. To investigate this possibility, runs 3 and 4 alone were analysed which, according to the results of previous studies (e.g., Filoteo et al., 2010), should provide a sufficient number of trials for participants to switch to the implicit system in the II condition.

A conjunction analysis, using the same thresholds as before, again revealed extensive activation overlap between the RB and II conditions. This included the bilateral putamen, the bilateral caudate body as well as the bilateral caudate head, the prefrontal cortex and the right MTL (Figure 2.7). No regions were more activated in the

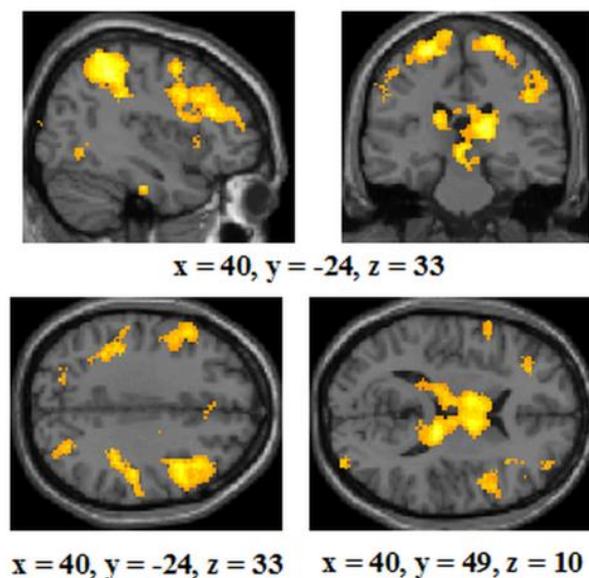


Figure 2.7. Analysis of areas commonly activated in both the RB and II conditions in runs 3 and 4 of the study only, thresholded at $p < .001$ and 27 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

RB condition than the II condition. However, as before, a number of regions were more engaged in the II condition than the RB condition (Table 2.3; Figure 2.8.a). Most prominent amongst these was activation in the bilateral hippocampus/posterior parahippocampal gyrus (left: 207 voxels; right: 44 voxels, Figure 2.8.b). However, as before, in spite of this generally elevated activation in the II condition compared to the RB condition, there was no evidence for the recruitment of regions linked to the implicit system of COVIS. Again a follow up ROI analysis was conducted comprising the caudate body, the putamen and the substantia nigra with a threshold of $p < .005$ and a voxel threshold of 10, but no regions were activated in this analysis.

Table 2.3

Brain Regions Activated More in the Information-Integration Condition than the Rule-Based Condition in Runs 3 and 4.

Region	Cluster Size	BA	Talairach Coordinates			z-score
			x	y	z	
Left Parahippocampal Gyrus	350	36	-24	-37	-7	4.25
Left Culmen		-	-14	-49	-3	4.00
Left Posterior Cingulate		29	-10	-42	10	3.76
Left Parahippocampal Gyrus		37	-32	-43	-8	3.71
Left Parahippocampal Gyrus		19	-34	-41	-5	3.62
Left Fusiform Gyrus		37	-34	-47	-8	3.20
Left Paracentral Lobule	176	5	-6	-36	53	4.06
Right Medial Frontal Gyrus		6	2	-24	57	3.49
Left Paracentral Lobule		3	-16	-36	55	3.36
Left Precentral Gyrus		4	-16	-28	53	3.23
Right Culmen	137	-	8	-53	-1	3.87
Right Parahippocampal Gyrus		30	16	-45	2	3.78
Right Lingual Gyrus		19	14	-49	1	3.46
Right Cingulate Gyrus	32	31	22	-27	36	3.71
Left Middle Temporal Gyrus	52	21	-54	-7	-13	3.62

Note that indented rows indicate voxels in the same cluster as the non-indented row above

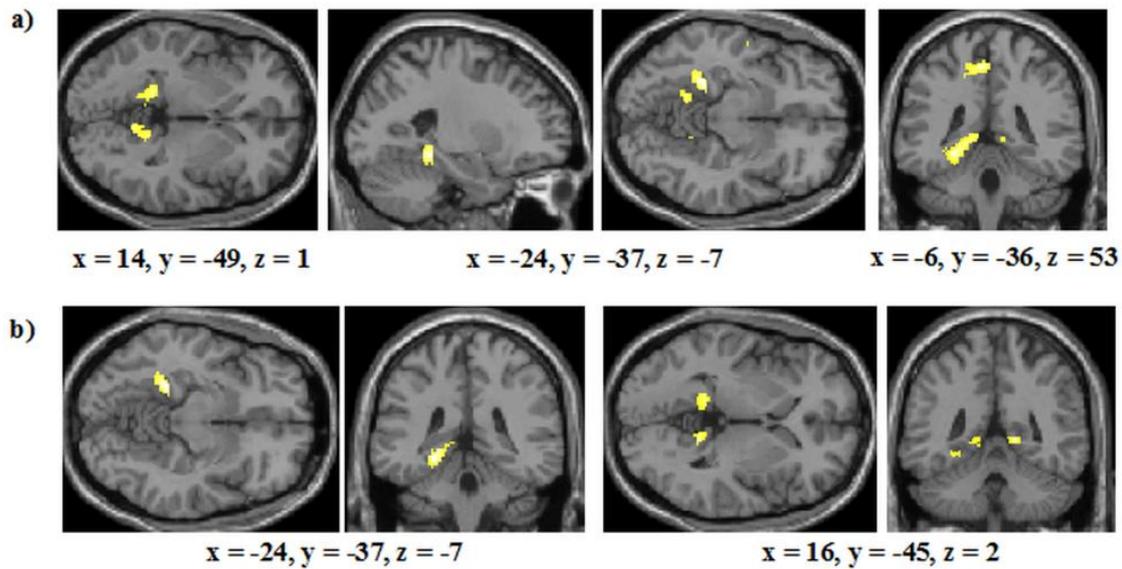


Figure 2.8. Analysis of runs 3 and 4 only for: (a) areas more activated in the II condition than the RB condition; (b) regions of the MTL more active in the II condition compared with the RB condition; non-MTL regions were masked in this analysis but the thresholds remained at $p < .001$ and 27 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Modelling analysis

The predictions made by COVIS are, of course, dependent on the assumption that more participants are using the explicit system in the RB condition than in the II condition and that a greater number of participants are using the implicit system in the II condition than in the RB condition. If participants in the II condition persist with the explicit system throughout learning (or alternatively if participants in the RB condition as well as the II condition use the implicit system) then this might explain why the present results appear inconsistent with the predictions of COVIS. It is harder, though, from a COVIS perspective to explain why participants in the II condition engaged the MTL, a critical region of the explicit system (Ashby & Valentin, 2005; Nomura & Reber, 2008), more than participants in the RB condition unless one assumes that the II category structure was more effective than the RB structure at engaging the explicit system. While this may seem unlikely, it can be tested using model-based analysis based on General Recognition Theory (GRT; Ashby & Gott, 1988) as is commonly

carried out in COVIS related studies (e.g., Ashby, Maddox, & Bohil, 2002; Filoteo et al., 2010; Nomura & Reber, 2008).

For each participant, the GRT analysis determines the decision boundary (from a set of pre-defined alternatives) that provides the best account of that participant's responses. Each participant is then assigned a strategy type (e.g. 'conjunctive') on the basis of the best-fitting model.

The *unidimensional* models assume that the participant determines a criterion along either the line orientation or line length dimension. As an example, for length, this corresponds to a rule such as: 'Assign to category A if the stimulus is long, or category B if short'. The unidimensional models have two parameters: the value of the criterion and the variance of internal (criterial and perceptual) noise.

The *conjunctive* model assumes that the participants make two judgments, one for each stimulus dimension, and then combine these to make a judgment about category membership. The conjunctive rule in the current analysis was: 'Assign the stimulus to category A if it is short and upright, otherwise assign to category B'. The conjunctive model has three parameters: the two criterion values and internal noise.

The *General Linear Classifier (GLC)* model assumes that the decision boundary can be described by a straight line that can vary in gradient and intercept. The unidimensional models are therefore special cases of the GLC model. The GLC model has three parameters: the intercept and slope of the decision boundary, plus internal noise.

The *random* model assumes that participants are responding randomly; it has no parameters.

For each participant, the best fit of each of these models was calculated, and the best-fitting model selected using Akaike’s information criterion (Akaike, 1974) as implemented by others in the COVIS literature (e.g., Ashby et al., 2002; Filoteo et al., 2010; Maddox, Ashby, & Bohil, 2003). The results from this analysis, which was performed using the grt package in the R environment (Matuski, 2014), are reported in Table 2.4. Within the COVIS framework, the unidimensional and conjunctive models are considered to represent explicit, rule-based strategies, while the GLC represents an implicit, information-integration strategy.

The results, displayed in Table 2.4, are generally consistent with previous work indicating that more participants used a conjunctive strategy in the RB condition than in the II condition and that more participants in the II condition used an II (GLC) strategy than in the RB condition. This is the pattern expected and obtained in previous COVIS studies (e.g., Ashby et al., 2002); therefore, the modelling analyses seem to rule out the possibility (at least within the COVIS framework) that the present results were driven by participants not using the intended strategy for their condition. However, the GRT modelling results, as usual, indicate that not all participants are adopting the expected strategy, so the brain activity of participants in the RB condition whose responses were best fit by a conjunctive strategy (12 participants) was compared to the activity of participants in the II condition whose responses were best fit by the optimal GLC model (13 participants). Note, the selection of a subset of participants on the basis of their

Table 2.4.

Proportion of Participants Using the Information-Integration (II), Conjunctive (CJ), Unidimensional (UD), or Random (RND) Strategies in the RB and II Conditions

	II	CJ	UD	RND
Rule-Based	0.23	0.55	0.17	0.05
Information-integration	0.59	0.18	0.23	0.00

GRT modelling results has seldom been carried out in previous COVIS related studies, perhaps because there are limits to the accuracy of these modelling results (see Donkin et al., 2014; Edmunds, Milton, & Wills, 2015, for a discussion), so this analysis should be taken with some caution. Nevertheless, given the nature of the present results, these supplementary analyses appear valuable.

A conjunction analysis using, as before, thresholds of $p < .001$ and 27 contiguous voxels, again revealed an extensive overlap of activation between the RB and II conditions in similar regions to those found in the all-participant analyses (Figure 2.9.a). Regions activated included the left MTL, bilateral caudate head, as well as the bilateral caudate body and right putamen. As in the all-participant analyses, no areas were more active in the RB than in the II condition (RB – II). No regions were activated more in the II than the RB condition either (II – RB) and, in particular, the prominent MTL

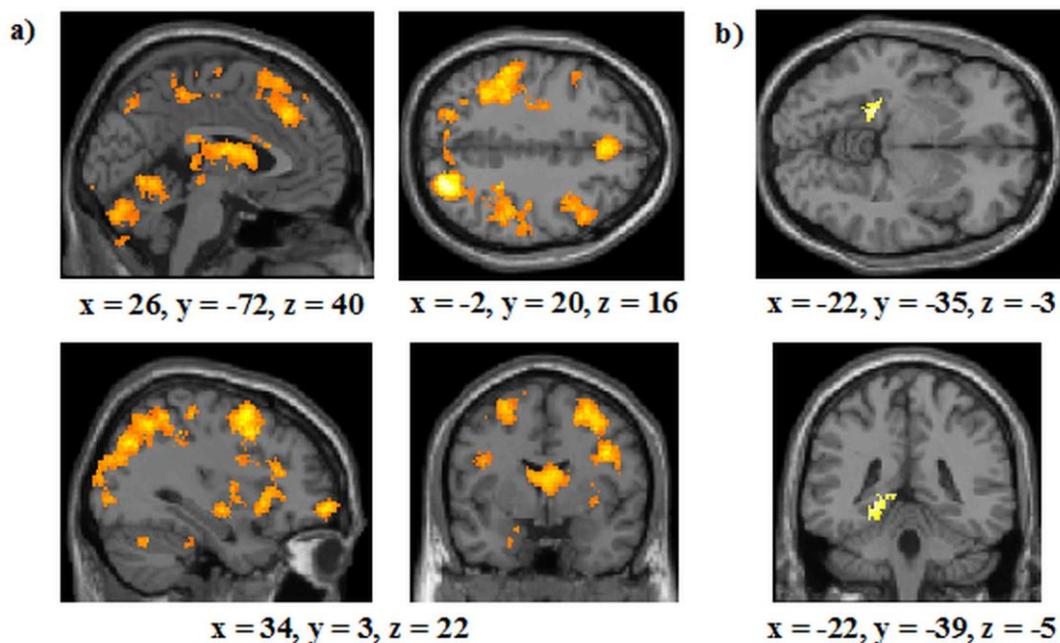


Figure 2.9. Analysis of participants who were shown by the modelling analysis to use the optimal learning strategy overall for all runs of the study: (a) areas commonly activated in the RB and the II condition (with thresholds of $p < .001$ and 27 contiguous voxels); (b) a ROI analysis of areas of the MTL more activated in the II condition compared with the RB condition (with thresholds of $p < .05$ and 79 contiguous voxels). The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

activation found in the all-participant analyses did not emerge. One potential reason for this is simply that the smaller number of participants in this model-based analysis reduced the ability to detect this activation. Therefore, a post-hoc ROI analysis of the MTL (using the WFU PickAtlas; Maldjian et al., 2003) was conducted with a threshold of $p < .05$ (uncorrected) and a cluster threshold of 79 (which combined produce a corrected threshold of $p < .05$ according to AlphaSim). This revealed activation in the same left hippocampus/ parahippocampal gyrus region (cluster size 115; Figure 2.9.b) as previously identified.

Again, no activation was found in regions linked to the implicit system of COVIS in the II - RB analysis. Therefore, another post-hoc ROI analysis was conducted comprising the caudate body, substantia nigra and the putamen in the same manner as for the MTL ROI analysis with cluster thresholds of $p < .05$ (uncorrected) and 41 contiguous voxels (which corresponded to $p < .05$, corrected according to AlphaSim). This also did not produce any significant activation. An additional ROI analysis with these regions using alternative thresholds of $p < .005$ and 10 voxels also yielded no activation. Finally, these modelling analyses were repeated with runs 3 and 4 alone. These produced the same pattern of results as the all-run analyses - there was considerable common activation (see Table 2.5) with no regions more activated in the RB compared with the II condition (RB - II). There was, though, as in the corresponding all-participant analysis, evidence of left MTL activation in a post-hoc ROI analysis of the II - RB contrast with thresholds of $p < .05$ and 79 contiguous voxels (cluster size 157, peak voxel: $x = -12, y = -41, z = 4$), and no evidence for the implicit system of COVIS in either the whole brain or ROI analyses.

Table 2.5.

Brain Regions Commonly Activated in the Rule-Based and Information-Integration Conditions in the Modelling Analysis of Runs 3 and 4.

Region	Cluster Size	BA	Talairach Coordinates			z-score
			x	y	z	
Right Thalamus	1459	-	8	-25	16	4.75
Left Caudate Tail		-	-16	-28	20	4.49
Right Caudate Body		-	2	5	13	4.11
Left Caudate Body		-	-2	7	13	4.10
Right Caudate Tail		-	20	-34	13	3.98
Left Thalamus		-	-14	-19	12	3.74
Left Postcentral Gyrus	873	2	-28	-36	59	4.64
Left Superior Frontal Gyrus		6	-18	-6	69	4.21
Left Precentral Gyrus		4	-26	-24	64	3.88
Left Superior Parietal Lobule		7	-32	-50	58	3.75
Left Precuneus		7	-20	-66	49	3.62
Right Superior Parietal Lobule	237	7	34	-46	56	3.90
Right Inferior Parietal Lobule		40	36	-44	52	3.79
Right Precuneus		7	30	-43	45	3.25
Left Culmen	181	-	-12	-49	-8	3.89
Left Declive		-	-6	-55	-12	3.26
Right Culmen		-	2	-60	-2	3.22
Left Cerebellar Lingual		-	-2	-45	-11	3.19
Right Middle Frontal Gyrus	245	10	44	43	11	3.86
Right Inferior Frontal Gyrus		9	55	15	29	3.74
Right Insula	75	-	30	21	1	3.77
Right Inferior Frontal Gyrus		47	28	27	-5	3.41
Right Postcentral Gyrus	72	3	16	-36	64	3.75
Left Middle Temporal Gyrus	30	37	-52	-64	3	3.70
Right Precuneus	197	7	24	-72	39	3.58
Left Sub-Gyral	49	6	-26	5	59	3.58
Left Middle Frontal Gyrus		6	-20	11	62	3.20
Left Pyramis	63	-	-24	-62	-27	3.37
Right Superior Frontal Gyrus	29	6	24	-5	63	3.51
Right Superior Frontal Gyrus	41	6	32	4	50	3.41

Note that indented rows indicate voxels in the same cluster as the non-indented row above

Imaging analysis - Incorrect trials baseline measure

To complement the analyses just described, the principal analyses were also run using incorrect trials as the baseline. For the all-runs analysis, consistent with previous work (e.g., Cincotta & Seger, 2007), the left caudate head was more active on correct trials than incorrect trials (with thresholds of $p < .001$ and 27 contiguous voxels) for both the RB (peak voxel: $x = -16, y = 20, z = 5$) and II (peak voxel: $x = -8, y = 13, z = -6$) groups. There was again, using the same conjunction analysis approach as before, large overlap of activation between the II and RB conditions including bilateral putamen, left caudate body, right MTL and frontal lobe (including BA's 8, 9, 10 and 11) (Table 2.6). However, there were not any differences detected between II and RB learning. A similar pattern emerged when considering all blocks in the modelling analysis with activation overlap in the conjunction analysis, but no significant differences detected between II and RB conditions.

Table 2.6.

Brain Regions Commonly Activated in the Rule-Based and Information-Integration Conditions in All Runs When Using Incorrect Trials as the Baseline.

Region	Cluster Size	BA	Talairach Coordinates			z-score
			x	y	z	
Left Medial Frontal Gyrus	693	11	-6	50	-9	5.54
Right Medial Frontal Gyrus		10	10	50	-8	4.81
Left Anterior Cingulate		32	-16	37	-2	3.61
Left Middle Temporal Gyrus	677	39	-52	-65	22	4.36
Left Angular Gyrus		39	-42	-74	31	4.29
Left Inferior Parietal Lobule		39	-46	-68	38	3.79
Right Putamen	70	-	18	9	-9	4.29
Left Superior Parietal Lobule	363	7	-20	-52	58	4.26
Left Postcentral Gyrus		3	-20	-23	47	4.13
Left Precentral Gyurs		4	-14	-24	64	3.92
Right Parahippocampal Gyrus	132	28	16	-5	-13	4.25
Left Putamen	74	-	-14	9	-9	4.20
Right Cuneus	67	18	26	-77	17	4.15
Right Middle Occipital Gyrus		19	24	-85	12	3.29
Right Precuneus	361	7	18	-44	56	4.13
Right Sub-Gyral		40	24	-38	55	3.97
Right Precentral Gyrus		4	18	-19	58	3.78
Right Middle Temporal Gyrus	193	37	54	-66	5	4.10
Right Superior Temporal Gyrus	263	42	63	-21	12	4.10
Right Postcentral Gyrus		40	65	-20	21	3.69
Left Superior Frontal Gyrus	288	8	-16	39	42	4.02
Left Medial Frontal Gyrus		9	-20	33	30	3.95
Left Middle Frontal Gyrus		8	-28	27	39	3.76
Left Caudate Body	43	-	-20	11	22	3.96
Right Middle Frontal Gyrus	163	11	24	28	-15	3.94
Left Cingulate Gyrus	186	31	-12	-41	33	3.90
Right Cingulate Gyrus		31	8	-41	33	3.80
Left Precuneus	52	31	-14	-57	25	3.87
Left Posterior Cingulate		23	-8	-54	16	3.27
Left Inferior Frontal Gyrus	47	47	-24	33	-10	3.76
Left Middle Occipital Gyrus	112	19	-32	-81	8	3.71
Left Middle Temporal Gyrus	62	21	-54	-47	-3	3.53

Note that indented rows indicate voxels in the same cluster as the non-indented row above

Looking at runs 3 and 4 alone, there was again common activation in the frontal, parietal and temporal lobes (Table 2.7) and no significant differences between RB and II learning in whole brain analyses. However, in a similar ROI MTL analysis to before (thresholds $p=.05$ and 79 contiguous voxels, corresponding to $p<.05$, corrected), greater activation was observed in the II condition than the RB condition in two right hippocampus/parahippocampal gyrus regions (cluster size 215, peak coordinate: $x = 24$, $y = -7$, $z = -13$; cluster size 180, peak coordinate: $x = 34$, $y = -34$, $z = -12$; Figure 2.10), with the posterior cluster being in the same area as observed in the corresponding odd-or-even comparison. There was, though, again no evidence for activation in regions associated with the implicit system of COVIS, even when the analogous ROI analyses to those previously conducted were performed. This same pattern emerged when considering runs 3 and 4 alone in the modelling analysis.

Table 2.7.

Brain Regions Commonly Activated in the Rule-Based and Information-Integration Conditions in Runs 3 and 4 of the Study Using Incorrect Trials as a Baseline

Region	Cluster Size	BA	Talairach Coordinates			z-score
			x	y	z	
Left Medial Frontal Gyrus	443	10	-2	50	-9	4.38
Right Medial Frontal Gyrus		10	6	54	-6	4.32
Left Superior Frontal Gyrus		10	-8	60	-3	4.16
Left Middle Temporal Gyrus	562	39	-52	-63	22	4.04
Left Angular Gyrus		39	-42	-74	31	4.01
Left Inferior Parietal Lobule		39	-46	-66	37	3.98
Left Superior Temporal Gyrus		39	-55	-63	25	3.77
Left Superior Frontal Gyrus	177	8	-34	26	47	3.83
Left Middle Frontal Gyrus		8	-28	27	37	3.74
Left Middle Frontal Gyrus		9	-26	27	32	3.58
Left Superior Frontal Gyrus		9	-24	35	33	3.23
Right Parahippocampal Gyrus	27	28	16	-10	-10	3.72
Left Posterior Cingulate	54	30	-6	-52	19	3.71
Left Precuneus		31	-12	-57	25	3.37

Note that indented rows indicate voxels in the same cluster as the non-indented row above

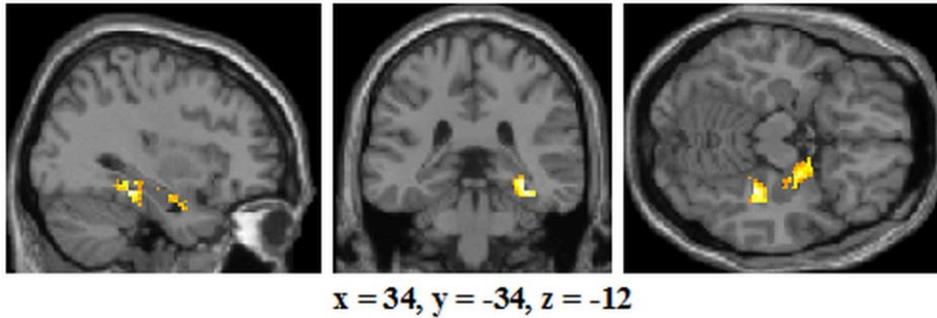


Figure 2.10. Analysis of correct trials contrasted against incorrect trials across all participants, showing a ROI analysis of MTL activation greater in the II compared with the RB condition in runs 3 and 4 (thresholded at $p < .05$ and 79 contiguous voxels). The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Discussion

Previous work has found that there is differential brain activity in the learning of RB and II categories with the MTL preferentially recruited for RB compared to II learning while the caudate body is engaged more for II than RB learning (Nomura et al., 2007). However, Experiment 1 finds no evidence for this pattern of dissociable neural activation. In particular, the most noteworthy finding was that the MTL was significantly more activated in the II condition than the RB condition. In addition, there was a striking overlap of activation between RB and II category learning emphasizing the extensive common neural processes that are engaged in learning both category structures. Common activation included regions thought to be recruited both in the explicit system such as the prefrontal cortex (including BA's 8, 10, 46, and 47), the anterior cingulate, the caudate head, and the MTL, and regions implicated in the implicit system including the posterior caudate, the putamen, and the substantia nigra. This pattern persisted when the second half of training was analysed alone. The same basic findings were also observed when including only those participants who had used the intended strategy as indicated by GRT modelling analyses.

The extensive activation overlap found in the present study between RB and II conditions could represent a similar system of category learning being applied when participants are presented with either category structure. However, it should be considered that some of this overlap in activation could represent processes necessary in both conditions, but not specific to category learning. For example, attentional demands, stimuli processing, response selection and feedback monitoring will be necessary in both conditions regardless of the category structure and may therefore contribute to the neural overlap in activation found in the present study. A multiple-demand brain network similar to that proposed by Duncan (2010) could also be operational in both tasks that coordinates and integrates the specific sub-tasks necessary to perform category learning. COVIS could also offer an explanation for the activation overlap found between the two conditions. The work of Ashby and Crossley (2010) suggests that there should be extensive overlap between II and RB category learning (as is presently observed) and no neural differences between conditions because both systems are operating simultaneously. The dominant learning system (for example the explicit system in RB learning) then inhibits the responding of the competing system. Nevertheless, what is particularly striking about the present results and a challenge to COVIS, as it is currently formalized, is not so much the overlap in activation between the conditions but that regions of the MTL were found to be activated *more* in the II condition than the RB condition, with this pattern more pronounced in later learning, when COVIS appears to make the reverse prediction that there should be *less* activation in the II condition than the RB condition in this region.

One important question, therefore, is why a markedly different pattern of results from Nomura et al. (2007) was observed? It is unlikely that the present study's use of an odd-or-even baseline measure for MRI analysis can account for this discrepancy between studies because the same basic pattern of results was observed when incorrect

trials were used as the baseline (albeit less pronounced, perhaps for the reasons outlined in the introduction).

Hence, it seems plausible that it is the change in RB task from a unidimensional task in Nomura et al. (2007) to a conjunctive task in the present study that is causing the difference in results. The decision boundaries of both unidimensional and conjunctive RB tasks have been found to be easily verbalizable in relation to II categories which have been found to be more difficult to verbalize (Edmunds et al., 2015). However, the conjunctive category structure appears to be a better control for the II category structure on extraneous variables known to be relevant to category learning, such as the number of relevant dimensions for categorization (Edmunds et al., 2015). The unidimensional structure used in Nomura et al. (2007) has only one dimension relevant for category learning compared with the two dimensions relevant in the II condition.

Multidimensional structures have been shown to implicate the basal ganglia more so than categories with simpler structures (e.g. Ell et al., 2010). This could have led to the larger levels of caudate body activation in the II condition of Nomura et al. (2007) compared with the RB condition.

Related to this, because unidimensional classifications are generally easier to learn than multidimensional classifications (e.g., Ashby, Maddox, & Bohil, 2002; Maddox, Ashby, & Bohil, 2003), Nomura et al. reduced the category separation of the unidimensional structure compared to the II structure. While this enabled error rates to be successfully matched between conditions, this manipulation effectively replaced one confound with another (Stanton & Nosofsky, 2007). The MTL is assumed to be critical for storing the precise location of the decision boundary (Nomura & Reber, 2008) and it seems plausible that this would be more demanding in Nomura et al.'s unidimensional

structure, where the decision boundary is more difficult to perceptually discriminate than the II structure which could have been driving the differential activation in this region.

The comparison of the present results and those of Nomura et al. (2007) highlights the influence that extraneous variables between conditions can have on categorization. Controlling for these extraneous factors allows stronger inferences to be drawn from the data. Concerns regarding extraneous variables have been raised regarding behavioural dissociation work (e.g., Edmunds et al., 2015), and the present research extends these concerns to previous imaging data. In light of this, the present results indicate that COVIS may be in need of revision to accommodate the greater level of observed MTL activation in II categorization compared to RB categorization.

Current research concerning the role of the MTL in category learning is inconsistent. While some, such as Nomura et al. (2007) find it active during RB learning, it has also been found active by others in more complex learning tasks akin to II category structures (e.g., Cincotta & Seger, 2007; Milton & Pothos, 2011). Moreover, some studies have found the MTL engaged only during initial category learning (Seger & Cincotta, 2006) whilst others have found the area unengaged by category learning tasks (e.g., Milton, Wills, & Hodgson, 2009; Seger & Cincotta, 2002). However, such diversity in findings is unsurprising given the multitude of differing methods between studies. Perhaps some of these discrepancies can be accounted for by methodological aspects such as choice of baseline measure. For example, Seger and Cincotta (2002) and Milton et al. (2009) compared activation in category learning trials to a resting state baseline measure of neural activity during a fixation point or blank screen. This could explain the findings of no MTL activation as the default network has been found to implicate the MTL (Buckner, Andrews-Hanna, & Schacter, 2008) therefore making such a baseline an insensitive measure of task related MTL recruitment and resulting in

an underestimation of activation during category learning. The use of a more sensitive baseline to MTL activation suggests that this region is a crucial area in category learning.

The present finding of MTL activity during RB and II category learning (albeit greater during the II condition) could be accounted for by the theory proposed by Love and Gureckis (2007), who suggest that the MTL is integral in organising and storing category stimuli. Love and Gureckis (2007) theorise that the hippocampus is fundamental in forming abstract codes (known as clusters) which represent stimulus configurations that pertain to category membership (Davis et al., 2012a). If a novel stimulus is encountered that is similar to a previously stored configuration of stimuli, it will be ‘captured’ by a pre-existing cluster and categorized as such. However, if a stimulus is not similar to a pre-existing cluster then the MTL generates a new cluster in which to represent it as a configuration (Love & Gureckis, 2007). This theory was supported by work of Davis et al. (2012a; see also Davis et al., 2012b) who created categories of schematic beetles that could be separated on a single stimulus dimension, but with a few beetles that were exceptions to the rule. It was found that the MTL was more activated during the presentation of exceptions to the rule than to stimuli that were non-exceptions. It was concluded that this represented the MTL generating a new cluster in which to store these exception stimuli. This theory can account for the MTL activation found during II and RB learning in the present study. MTL activation could represent the formation of new clusters consistent with the presentation of novel stimuli and category structures, as well as perhaps the retrieval of the category ‘label’ in later learning once the category is established. However, the present study also found greater MTL activation during II compared with RB learning. Again, this is consistent with the theory proposed by Love and Gureckis (2007) as this category is more complex and difficult or impossible to describe with a verbal rule, and therefore contains more

exception stimuli. This requires the need for the formation of more clusters, hence the more active role of the MTL in this condition. The MTL was also found to be more active in the second half of learning compared with the first half of the paradigm in the II condition. This pattern of activation could also be accounted for by this model as the MTL may now also be functioning to ‘retrieve’ the learnt cluster label created to categorize a novel stimulus.

The ATRIUM model (Erickson & Kruschke, 1998) could also potentially account for the MTL activity in the II and RB conditions. This model proposes two systems of category learning. The first is a RB system similar to the explicit system proposed by COVIS (Ashby et al., 1998). The other system is also assumed to be explicit but is exemplar-based and is responsible for learning when rules are not easily applicable. This occurs when error feedback indicates that rule-based performance is suboptimal. It may be that in the II condition of the present study, participants are using a suboptimal explicit RB strategy but supporting this with storage of stimuli that do not fit with this generated rule (controlled by the exemplar-based system). Although the greater activation of the MTL in the II condition of the present study is surprising from the perspective of the COVIS model (Ashby et al., 1998), the findings are less surprising when considering alternative theories which can explain the engagement of the MTL during a task with an intrinsically more complicated structure.

The basal ganglia have also been previously linked with category learning, especially during feedback processing. The caudate head has been found by many to be active during positive feedback processing (e.g., Seger & Cincotta, 2002; Cincotta & Seger, 2007). In the present study the caudate head is active in both RB and II category learning conditions. This finding remains when comparing activation during correct categorization to activity found during incorrect trials, which supports the notion that

this area may be fundamental in processing positive feedback (Seger & Cincotta, 2002; Cincotta & Seger, 2007). The present study also finds the caudate body operational in both RB and II conditions. This supports research conducted by Seger and Cincotta (2006) who found successful learners engaged this area more than poor learners in an RB task, as well as Lopez-Paniagua and Seger (2011) who associated greater caudate body activity with successful stimulus-response association formation using Granger causality mapping. The putamen was also observed in the current study to be commonly activated in the RB and II conditions, which could represent the need for motor planning in both tasks as proposed by Cincotta & Seger (2007). While there are clearly differences in the procedures used in these studies and the present experiment, it is plausible that these regions serve the same role in RB and II learning. If this is the case then my results would also be consistent with Nosofsky, Stanton and Zaki's (2005) claim that RB categorization as well as II categorization has a procedural component.

The present results also resonate with those of Schnyer et al. (2009) who found that patients with ventromedial prefrontal cortex (VMPFC) lesions were impaired compared to controls on both II and RB tasks. Schnyer et al. suggested that the VMPFC is responsible for feedback processing in both RB and II learning and is involved in the selection and maintenance of the optimal learning strategies. The present study finds the medial prefrontal cortex more active in correct categorization trials compared with incorrect trials offering imaging support for this previous neuropsychological evidence.

There is a temptation to consider the data from Experiment 1 in the context of whether it is more supportive of single system or dual system accounts. The extensive activation overlap between RB and II conditions could be indicative of a single category learning system operational during RB and II tasks that is more taxed (and therefore more active) when performing an II task to the same standard as a more simply structured RB task. This would account for the greater activation in the II condition

compared to the RB condition, especially in the MTL. However, the results could also fit with dual-process accounts of learning. The explanation of the greater MTL engagement during the II condition has already been considered in the light of dual-process predictions made by Erikson and Kruschke (ATRIUM; 1998) and can equally account for the extensive neural activation overlap seen as well as the increased activity in II learning.

The present results could also be looked at from the same viewpoint as advocated by Davis et al. (2012a) who note that given that the criteria for establishing truly qualitatively separable systems are often underspecified, a more profitable way of viewing category learning may be to link brain function to the particular processes required. Therefore, in the present study, taking into account previous research, the prefrontal cortex may be operational in rule generation and selection, the caudate head in feedback processing, the caudate body for stimulus-response associations and the MTL for storing decision boundaries. These ideas are compatible with the predictions of the COVIS model (see Ashby & Valentin, 2005). However, the greater MTL activation found in the II condition of the present study could also represent a role for this area in the storage of exemplars and the increased reliance on this area in non-verbal tasks not easily explained by rules (Davis et al., 2012a). To test this explanation of the present results, it would be valuable in future to investigate patients with MTL lesions to see if they, as would be predicted by ATRIUM and Davis et al. (2012a), perform worse in acquiring II categories than RB categories. As far as I am aware, this hypothesis has not yet been investigated. A further prediction derivable from this hypothesis is that people should have enhanced memory for exemplar instances after II category learning than after RB learning as they are more likely to have stored individual exemplars as exceptions in this condition.

While fMRI provides excellent spatial resolution, it is well known to have limited temporal resolution. One consequence of this is that the present study, like Nomura et al.'s (2007), and virtually all existing imaging studies of categorization cannot determine whether the activation identified is driven during the response or feedback processing stages. While greater MTL activation in the II compared to the RB condition appears unexpected from COVIS's perspective regardless of when it occurs in the category learning process it would, nonetheless, be valuable in follow-up studies to understand at what stage in the process this difference is occurring. Such a follow up study will be discussed in Chapter 3, where an extra variable ITI after the category response was included to identify activation differences between the response and feedback stages.

In conclusion, Experiment 1 aimed to advance knowledge about the neural pathways engaged in RB and II category learning tasks by building on the limited research currently published. There was extensive neural commonality between RB and II learning structures supporting the findings of Milton and Pothos (2011). Little support for the dual system model of COVIS (Ashby et al., 1998) was found, as when controlling for category separation, error rates, and number of relevant dimensions the MTL was specifically more activated during II compared with RB learning. This is contrary to the predictions of COVIS where the MTL is associated with explicit memory processes (e.g., Scoville & Milner, 2000; Squire, Stark, & Clark, 2004) proposed to be optimally engaged during RB learning rather than II learning, which should preferentially implicate the implicit category learning system. Experiment 1 extends our understanding of the neural processes that underlie RB and II learning and poses a challenge for COVIS as it is currently instantiated.

Chapter 3. Investigating the feedback processing systems of rule-based and information-integration learning with fMRI

Experiment 1 used fMRI to compare neural areas activated during stimulus presentation in RB and II learning. The results found were contrary to the predictions of COVIS (Ashby et al., 1998) as the MTL, an area COVIS implicates in the explicit system, was more activated during II learning than in RB learning, when COVIS would expect the implicit system to be dominant in II conditions. Therefore, Chapter 2 finds little evidence to support the presence of separable explicit and implicit neural systems during categorization decision processes. However, while this analysis focused on the decision making process (as neural activation during stimulus presentation was measured), COVIS also predicts that there are large differences in feedback processing during II and RB learning (e.g., Ashby et al., 2002), so it could be that more dissociations emerge when considering this aspect of the classification process.

As discussed in Chapter 1, Ashby and Valentin (2005) propose that the procedural, implicit, system is optimally applied when learning II category structures, and relies upon immediate feedback to create temporally dependent stimulus-response associations. In this system the caudate body and tail are richly connected with the visual and motor cortices, and receive representations of the visual stimulus and the motor response performed. If the response made is followed by positive feedback then the substantia nigra releases dopamine, which strengthens the neural synapses between representations of the stimulus and response in the posterior caudate. With repeated correct responses, the stimulus-response associations made in the caudate body and tail become stronger and the category is learnt without conscious awareness.

In contrast, COVIS theorists propose that the explicit system optimally learns RB categories, and engages a feedback processing system that relies on WM and

executive functioning, processes that are dependent on the frontal lobes (e.g., Ashby & Valentin, 2005; Filoteo et al., 2010). If feedback shows the current rule to be wrong, a switch to another rule is implemented through the prefrontal cortex, anterior cingulate and caudate head (Ashby & Valentin, 2005).

COVIS, therefore, predicts that feedback processing in II and RB conditions should be behaviourally and neurally dissociated. These potentially separable feedback processing systems have been investigated behaviourally using delayed feedback in RB and II category learning procedures. For example, Maddox, Ashby, and Bohil (2003) delayed corrective feedback after responding to stimuli for either 10, 5 or 2.5 seconds during II and RB category learning. RB participants were unaffected by feedback delays; however, II learning was impaired in participants who received delayed feedback of 2.5 seconds or more compared with II participants who received immediate feedback. Maddox and Ing (2005) noted, though, that the RB and II conditions used by Maddox et al. were not matched on the number of relevant dimensions for categorization (the RB condition was unidimensional whereas the II condition was two-dimensional). Therefore, they postulated that the dissociation found could be attributable to the II condition being more complex. This could suggest that as the II condition was harder to learn, it was more susceptible to disruption from less optimal delayed feedback. Maddox and Ing employed conjunctive RB and II categories matched on the number of relevant stimulus dimensions by separating stimuli into four categories (see Figure 3.1). They replicated the finding of Maddox et al. as II learning was selectively impaired by delaying feedback for 5 seconds whereas RB learning was unaffected. However, as discussed in Chapter 1, the use of a mask visually very similar to the categorization stimuli during the delay period could have affected II learning,

This image has been removed by the author of this thesis for copyright reasons.

Figure 3.1. The category structures used by Maddox and Ing (2005); (a) the RB category; (b) the II category. The solid lines represent the decision boundaries separating the stimuli into four categories. Figure adapted from Maddox and Ing (2005).

because a stimulus-response association could have mistakenly formed to the mask stimulus rather than the stimulus that was actually categorized. Indeed when Dunn et al. (2012) used a mask between stimulus presentation and feedback that was dissimilar to the stimulus presented, the dissociation between conjunctive RB and II learning disappeared (see Chapter 1, page 22).

The dissociation in RB and II performance reported by Maddox et al. (2003) and Maddox and Ing (2005) has also been supported by Worthy, Markman, and Maddox (2013) who presented participants with either delayed feedback or immediate feedback in either unidimensional RB or II category conditions, and therefore, as in Maddox et al. (2003), employing category structures which varied on the number of relevant stimulus dimensions. II participants were more impaired by a 1000ms feedback delay than a 500ms delay whereas RB participants were unaffected by feedback timing. Interestingly, II participants also performed worse when there was a 0 second feedback delay than II participants who experienced a 500ms delay. The authors suggest that II feedback presentation is optimal at about 500ms as the dopamine levels responsible for learning peak at this point after responding.

RB learning has also been found to be impaired by a concurrent load during feedback presentation. Filoteo et al. (2010) presented participants with II or conjunctive RB category structures to learn. Half of the participants also completed a digit probe WM task, whereby after feedback was presented participants were shown an array of digits for 500ms. Then, after a delay of 1000ms, participants saw a random digit and they had to indicate if it was in the array previously shown. There was a significant interaction between category structure and concurrent load condition. When performing a WM task concurrent with feedback processing II participants performed significantly better than II participants who did not perform the concurrent task, whereas RB participants were numerically impaired in the load condition compared to those learning the RB structure without the concurrent load. COVIS predicts that early stages of both RB and II learning are supported by the explicit system. However, when the explicit system has been shown to be ineffective at learning the II category structure, participants will switch to the implicit system which is more optimal for learning in this condition. Under a load, the explicit system will operate less effectively and should be shown to fail quicker, and therefore participants will transition to the implicit system earlier, meaning that more of the learning process is under the control of the optimal learning system. This leads to the greater learning of the II structure under load conditions than no load conditions. In contrast, the RB structure is learnt best using the explicit system which, as noted previously, requires extensive WM capacity that is likely compromised under load conditions. This means that, if anything, COVIS should predict worse performance under load conditions than no load conditions.

However, as was discussed in Chapter 1 (page 23) participants in Filoteo et al.'s WM load condition experienced a longer ITI than those without a concurrent task, and this consequently increased the time participants could think about the feedback received. This could account for performance being higher in the II load condition

compared to the no load condition; indeed, Newell et al. (2013) found that increasing the length of time available to process feedback in an ITI improved II participants' performance. This led the authors to conclude that, contrary to the predictions of COVIS (Ashby et al., 1998), II participants benefit from explicitly thinking about the feedback they have received.

Previous studies have shown that some brain regions, such as the caudate head, are associated with positive feedback by comparing neural activation during trials receiving positive feedback compared with negative feedback (e.g., Cincotta & Seger, 2007; Seger & Cincotta, 2006; Seger & Cincotta, 2005; and in Chapter 2 of this thesis). However, one feature of these studies, together with Experiment 1 from this thesis, Nomura et al. (2007), and virtually all existing imaging studies of categorization, is that they do not directly measure brain activation during the feedback processing stage. Instead, the focus of the analyses is nearly always on the decision processing period, and the activation likely contains a mixture of both the decision and feedback processing stages. Experiment 1 reports greater activation in the MTL for the II condition than the RB condition. This finding is unexpected from a COVIS perspective, but nevertheless it is unclear whether this activation is driven by decision processes, feedback processing or is present in both the decision processing and feedback processing stages. Experiment 1 cannot provide direct insight into this because not only was the focus of the statistical analyses centred on the stimulus onset (decision processing stage) but feedback was also provided immediately after this. The perfect temporal correlation between the decision making and feedback stages makes the neural correlates of the different processes impossible to disentangle.

Only two published studies have, to my knowledge, attempted to unravel the neural processes of the decision making and feedback processing stages, and neither of

which investigated the predictions of COVIS. Specifically, Lopez-Paniagua and Seger (2011) separated brain activation during stimulus presentation (decision making) and feedback presentation, by not only including a variable ITI (often known as a jitter) at the beginning of each trial, as is standard, but also between the decision and feedback stages to temporally decorrelate these two processes. Participants were presented with a weather prediction task whereby they learnt to associate six different visual pattern stimuli with a 'sunny' or 'rainy' outcome. Each stimulus had a different probability of being associated with each outcome. Again the caudate head was associated with positive feedback processing along with the MTL; however, no brain areas were more engaged during feedback processing than during stimuli presentation. Aron et al. (2004) also used a similar 'jittered' procedure to examine neural areas recruited during stimulus presentation and during negative and positive feedback, and found that the midbrain was more active during stimulus presentation than in feedback presentation when participants were least certain of category assignment.

Experiment 3 will similarly implement a variable time period delay between stimulus and feedback presentation to temporally decorrelate the processes. This allows for the investigation of untested COVIS assumptions that RB and II learning conditions will implement dissociable feedback systems. The category structures of Filoteo et al. (2010) will be implemented to control for category difficulty, separateness and number of relevant stimulus dimensions. By doing this, it can be inferred whether RB or II conditions engage different feedback processing systems, as would be predicted by COVIS, or whether similar neural processes underlie each task. However, as discussed, delaying feedback presentation (as the current approach requires) has been found to hinder II performance but not RB performance (e.g., Maddox & Ing, 2005) which could mean that II participants performing a category learning task with moderately delayed feedback may not engage the potential implicit neural system. Instead they could rely on

suboptimal explicit learning which is, according to COVIS, unaffected by delays. In order to rule out this possibility, Experiment 2 was conducted in order to confirm that delayed feedback would not differentially affect RB or II performance under the conditions to be employed in Experiment 3.

Experiment 2

The addition of a random delay of 250ms-1500ms between stimulus presentation and feedback is necessary to decorrelate neural activation to stimulus and feedback to research neural areas correlated with feedback processing. However, previous research has reported that delaying feedback by 2.5 seconds differentially slows II learning compared with RB learning (e.g., Maddox et al., 2003). The present study only delays feedback by a maximum of 1500ms to decrease the possibility of the delay differentially influencing RB and II performance. Nevertheless, it is still important to confirm that this delay period does not selectively impair II learning compared with RB learning, as this would suggest that the II category learning system cannot function adequately in delayed conditions. In a 2 (category structure) X 2 (feedback type) design, Experiment 2 will employ the same conjunctive RB and II categories as Experiment 1 with half of the participants also receiving feedback with a variable delay of 250ms to 1500ms between stimulus and feedback presentation.

Method

Participants

Participants were 80 University of Exeter students (male $n = 10$, mean age = 19.76, $SD = 3.77$). Participants gave informed consent and received course credit for participating. All participants performed above chance on the category learning task.

Design

The study used a 2 (category structure) X 2 (feedback condition) between-subjects design. There were four conditions: RB with immediate feedback, RB with delayed feedback, II with immediate feedback, II with delayed feedback. There were 20 participants randomly allocated to each of these conditions.

Stimuli

The stimuli used, as in Experiment 1, were a sub-set of stimuli taken from Filoteo et al. (2010; see Figure 3.2). The conjunctive category structure was optimally learnt by a verbal rule (short, upright lines belong in category A, and the rest in category B) and the II condition had a diagonal decision boundary. In both conditions, 480 of the original 600 lines of Filoteo et al. (2010) were randomly selected (240 stimuli in each category). As in Filoteo et al. (2010), there was 5% overlap between the categories so that the maximum accuracy attainable was 95%.

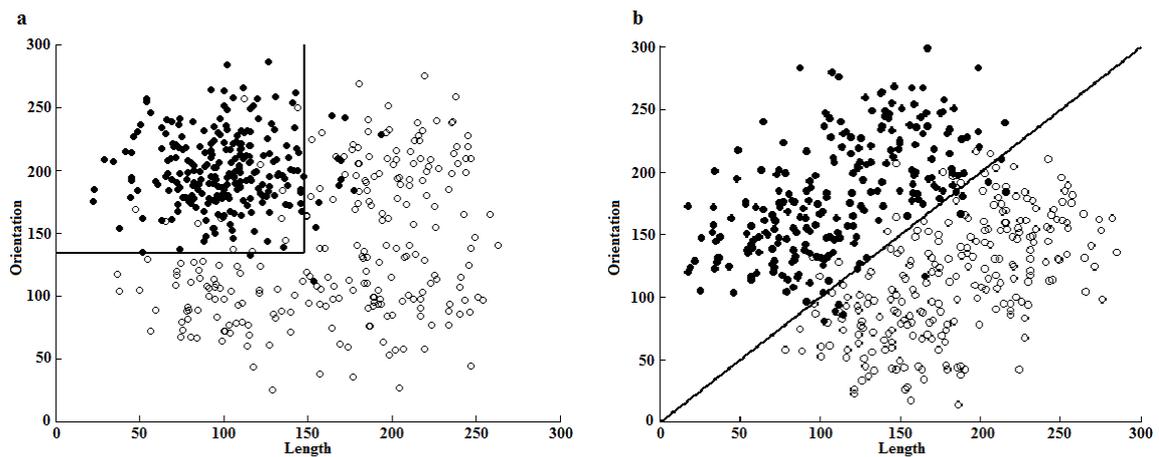


Figure 3.2. The category structures used for the present study; (a) the conjunctive rule-based condition; (b) the information-integration condition. Solid lines indicate the decision boundary separating category A (filled circles) and category B (unfilled circles).

Procedure

In order to fully mimic the learning task participants would be faced with in Experiment 3, the odd-or-even baseline task of Experiment 1 was employed to complement the categorization phase. Each run began with 15 odd-or-even trials, followed by one block of 60 categorization trials and a further 15 odd-or-even trials. After every 20 trials in the category learning block there was a 10 second break during which time participants were asked to rest. In total there were 8 runs with 480 category learning trials presented in a random order and 240 odd-or-even trials. E-Prime (Psychological Software Tools, 2002) was used for the presentation and timing of stimuli and collection of response data.

In the *category learning task (immediate feedback)* participants were informed that they had to learn into which of two categories a series of stimuli belonged through trial and error and feedback presentation. The procedure for the RB and II conditions was identical. Each trial began with a blank screen lasting a varied interval between 500ms and 4000ms, followed immediately by a black fixation cross presented in the centre of the screen for 250ms. One of the stimuli then appeared on the screen for 1500ms during which time participants made their response using the 'x' button (for category A) and the 'm' button (for category B) on the computer keyboard. Corrective feedback ("Correct", "Incorrect" or "Time out!!") was then presented for 500ms. The next trial then immediately began.

The *category learning task (delayed feedback)* condition employed the same trial-by-trial procedure as the immediate condition. However, after the 1500ms of stimulus presentation there was a blank screen lasting 250-1500ms before the corrective feedback screen was presented. This feedback delay was chosen to reduce the impact

delaying feedback could potentially have on II learning. Delays of 2500ms or more have previously been shown to disrupt learning of an II category structure (e.g., Maddox et al., 2003; but see also Dunn et al., 2012) whilst there is no evidence that there is a selective impairment in II learning for the delay interval that we have used in this experiment.

As in Experiment 1, *the odd-or-even task* was closely modelled on that used by Stark and Squire (2001) and had a similar trial structure to the category learning task. Each trial began with a blank screen lasting between 500-4000ms, followed by a black fixation cross for 250ms. A randomly generated number from one to nine then appeared in the middle of the screen for 1500ms. Participants made their odd or even response using the 'x' button (for odd) and the 'm' button (for even) on the computer keyboard. Feedback ("Correct"; "Incorrect" or "Time out!!") was then presented for 500ms. In the delayed feedback condition again the blank screen lasted 250-1500ms before the corrective feedback screen was presented.

Results

Mean accuracy of responding across all 8 blocks can be seen in Figure 3.3. A mixed design ANOVA was completed with accuracy of responding during all 8 blocks entered

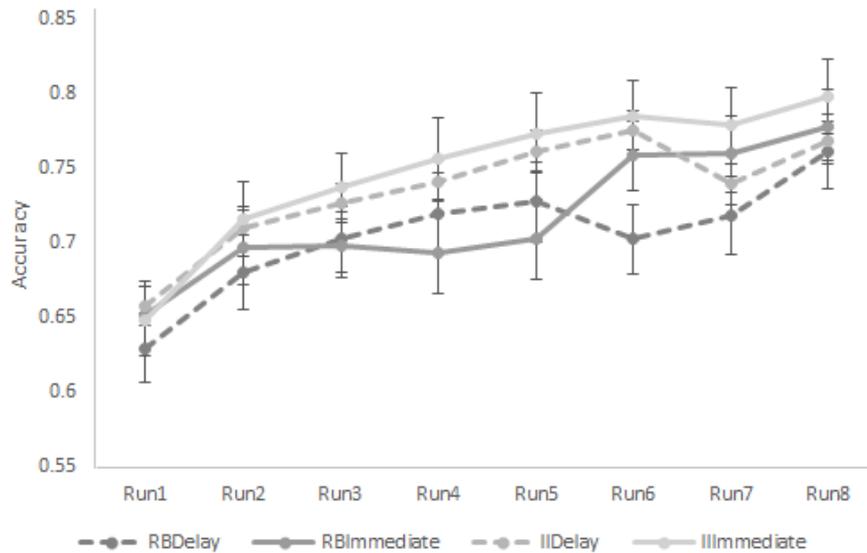


Figure 3.3. Proportion of correct category responses across blocks in Experiment 2. Error bars represent standard error.

as a within-subjects factor. The category structure (RB or II) and the feedback condition (immediate or delayed) were entered as between-subject variables. There was a main effect of block ($F(7, 532)=19.99, p<.001, \eta^2_p=.21$) in a linear direction ($F(1, 76)=53.09, p<.001, \eta^2_p=.41$) suggesting that participants were improving across blocks.

There was no significant main effect of feedback condition on accuracy of responding ($F(1, 76)=.51, p=.479, \eta^2_p=.01, BF = 0.17^8$), but there was a trend towards a main effect of category structure on performance ($F(1, 76)=2.83, p=.097, \eta^2_p=.04, BF = 0.07$) with II participants performing better than RB participants. However, the Bayes factor suggests that there is substantial evidence to accept the null hypothesis that there is no significant difference between RB and II performance. Critically, the interaction

⁸ Bayes Factor analysis requires an estimate of the mean expected difference under the experimental hypothesis; I estimated this from Maddox and Ing (2005), using plot digitizer (<https://sourceforge.net/projects/plotdigitizer/>). Following Dienes (2011), the expected difference was modelled as a two-tailed normal distribution with a standard deviation equal to half the mean.

between category structure and feedback condition was non-significant ($F(1, 76)=.003$, $p=.958$, $\eta^2_p<.001$, $BF = 0.13$) suggesting that the effect of feedback on learning was not modulated by category structure.

It was also examined whether longer time periods between stimulus and feedback presentation in the delayed condition hindered learning more than shorter feedback delays in either the RB or II conditions. The average accuracy of responding was calculated for trials with feedback delays smaller than 500ms (RB mean accuracy=0.7, SD=0.1; II mean accuracy=0.73, SD=0.07), between 501-1000ms (RB mean accuracy=0.71, SD=0.09; II mean accuracy=0.74, SD=0.09), and between 1001-1500ms (RB mean accuracy=0.71, SD=0.1; II mean accuracy=0.73, SD=0.08). The average accuracy for each delay period was entered as a repeated measures variable in a mixed measures ANOVA. Category structure was entered as a fixed factor. There was no significant main effect of delay period ($F(2, 76)=.88$, $p=.418$, $\eta^2_p=0.023$) or category structure ($F(1, 38)=1.24$, $p=.272$, $\eta^2_p=0.032$) on trial accuracy, and no interaction between delay period and category structure ($F(2, 76)=.17$, $p=.847$, $\eta^2_p=0.004$). Participants in neither condition were impaired as a function of the length of delay period.

Discussion

Experiment 2 examined if the fMRI procedure of Experiment 1 could be adapted by implementing a variable delay between stimulus presentation and feedback processing without differentially impairing performance in the II condition. There was no significant effect of feedback delay on RB or II performance and critically, no significant interaction between feedback type and category structure. There was also no evidence for the predictions of COVIS that II participants would be differentially affected by the addition of a variable delay compared to RB participants. However, the

present study does not implement delays as long as previous research supporting this proposal (e.g., Maddox et al., 2003; Maddox & Ing, 2005) which could account for this discrepancy in results. This pattern of results indicates that the current procedure is appropriate for comparing the neural correlates of II and RB feedback processing.

Experiment 3

Experiment 1 questioned the predictions of COVIS as results showed extensive areas of common neural activation between RB and II conditions during decision making (stimulus presentation). Moreover, the MTL, an area implicated in the explicit system of COVIS (Nomura et al., 2007) was found to be more activated in the II condition than the RB condition. However, COVIS also predicts substantial differences between the neural systems underlying RB and II performance during feedback processing, and therefore Experiment 3 aims to test this prediction. In the current literature there are, to my knowledge, no previous studies that have examined neural activation during feedback processing in RB or II category tasks.

If there are separable feedback processing systems functioning in RB and II conditions then RB participants should demonstrate activation in areas associated with the feedback processing aspects of the explicit system of COVIS, such as the frontal lobes, caudate head and anterior cingulate. Conversely, II participants should activate areas implicated in implicit feedback processing such as the caudate body/tail and putamen (e.g., Ashby & Valentin, 2005). However, if, as Experiment 1 indicates, II learning places greater demands on the explicit system than RB learning, areas such as the MTL and frontal lobes might be more active in the II condition compared with the RB condition. In Experiment 3, participants completed two fMRI scanning sessions

separated by 24 hours to minimize fatigue effects, and performed either the RB or II tasks of Experiment 2.

Method

Participants

Participants were 40 University of Exeter students (male $n = 16$, mean age = 19.55 years, $SD = 1.00$). However, one participant was excluded for excessive head movements inside the MRI scanner. Participants gave informed consent according to procedures approved by the University of Exeter's School of Psychology Ethics Committee and were remunerated £10 for participation. 20 participants were randomly allocated to one of two between-subject conditions (RB and II). The excluded participant was from the II condition.

Stimuli and procedure

The category learning procedure was identical to the 'delayed' condition in Experiment 2, and the same stimuli were employed. However, instead of responding using a computer keypad participants were required to respond in the MRI scanner by pressing the far left button on a button box with their left hand if they thought the item belonged to category A or the far right button with their right hand if they thought the item was a member of category B.

FMRI imaging

Images were collected using a 1.5-T Phillips Gyroscan magnet, equipped with a Sense coil over two scanning sessions separated by 24 hours. A T2*-weighted echo planar sequence ($TR = 3000\text{ms}$, $TE = 45\text{ms}$, flip angle = 90° , 36 transverse slices, $3.5 \times 2.5 \times 2.5\text{mm}$) was used. When participants entered the scanner their head was secured in place with foam pillows inside the coil to prevent excessive head movement. In each

scanning session participants completed the four blocks over four scanning ‘runs’, each of which contained 194 scans. Five “dummy scans” were completed before every run prior to presentation of the first trial. After the functional scans, standard volumetric anatomical MRI was completed using a 3-D T1-weighted pulse sequence (TR = 25ms, TE = 4.1ms, flip angle = 30°, 160 axial slices, 1.6 x 0.9 x 0.9mm).

FMRI data analysis

As in Chapter 2, data analysis was performed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>). Functional images were corrected for acquisition order, realigned to the mean image, and resliced to correct for motion artifacts. Realigned images were coregistered with the structural T1 volume and the structural volumes were spatially normalised. The spatial transformation was applied to the T2* volumes which were spatially smoothed using a Gaussian Kernel of 8mm full-width half maximum. Data were high-pass filtered (128s) to account for low frequency drifts.

Random effect whole brain analyses were completed using the general linear model with combined statistical thresholds of $p < .001$ (uncorrected) and 23 contiguous voxels, which together produce an overall corrected threshold of $p < .05$, according to AlphaSim, as implemented in the REST toolbox (Version 1.8; Song et al., 2011). A canonical hemodynamic response function (HRF), together with temporal and dispersion derivatives, was used to model the blood oxygen level-dependent response and the six head movement parameters were included as covariates. The analyses focused on comparing brain activation from the onset of feedback after categorization (for the RB and II groups separately) to the onset of feedback presented during the odd-or-even baseline task. Also, to measure common activation between feedback processing in the ‘RB – baseline’ contrast and the ‘II – baseline’ contrast, a conjunction analysis was conducted. This was performed by a logical ‘and’ function through the

minimum statistic to the conjunction null hypothesis (MS/CN; Nichols et al., 2005) technique implemented in SPM8. Both of these contrasts were again conducted with a combined threshold of $p < .001$ (uncorrected) and a cluster threshold of 23. Normalised MNI space coordinates were transformed to Talairach space (<http://imaging.mrcmbu.cam.ac.uk/imaging/MniTalairach>) to establish activation sites as per the atlas of Talairach and Tournoux (1988).

Results

Behavioural analysis

Figure 3.4 shows the mean categorization accuracy in the RB and II conditions across all runs. A mixed design ANOVA was completed with accuracy in the 8 blocks entered as the within-subjects factor and categorization task (RB/II) entered as the between-subjects factor. There was a significant effect of block ($F(7,259)=12.38$, $p < .001$, $\eta^2_p = .251$) and the linear contrast was significant ($F(1,37)=32.02$, $p < .001$, $\eta^2_p = .464$) indicating that categorization accuracy increased with practice. However, there was no significant difference between the II and RB conditions in categorization accuracy

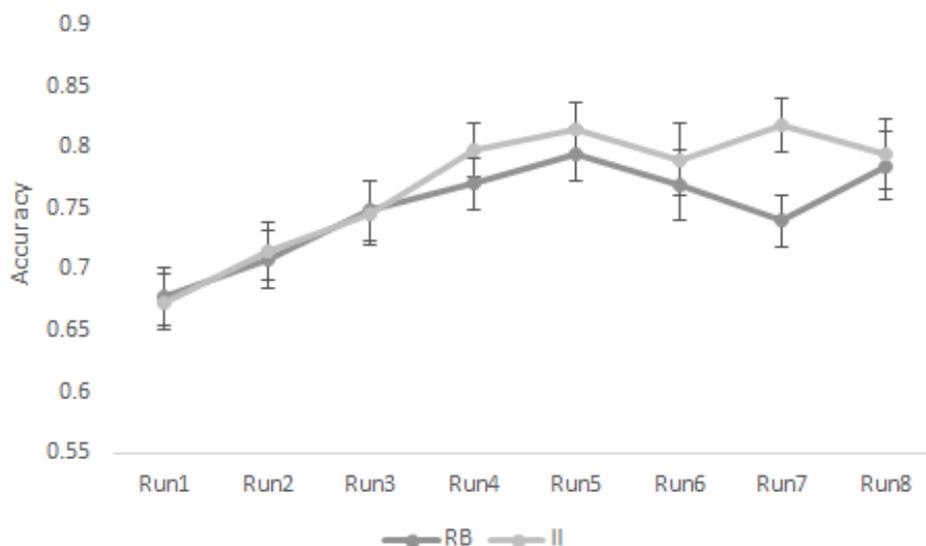


Figure 3.4. Proportion of correct category responses across all runs of the study

($F(1,37)=.59$, $p=.446$, $\eta^2_p = .016$, $BF = 1.18^9$), and no significant interaction between block and categorization task, ($F(7,259)=1.07$, $p=.384$, $\eta^2_p = .028$, $BF=0.91$).

Modelling analysis

The learning strategies used by participants were analysed in the same way as Experiment 1. For each participant, the GRT analysis determined the decision boundary (from a set of pre-defined alternatives) that provides the best account of responses made. Each participant was assigned to one of four strategy types. Participants using a *unidimensional* strategy based category membership on one dimension. The *conjunctive* model assumed participants used both stimulus dimensions to make a category judgment. The *general linear classifier (GLC)* model represented an II strategy whereby the decision boundary used by participants could be described by a straight line that varies in gradient and intercept. Finally, there was a *random* responding model. The learning strategies implemented by participants in the RB and II conditions are shown in Table 3.1. In both the RB and II conditions, the optimal strategy for category learning was the most used. This converges with the assumptions of Filoteo et al. that the present stimuli manipulate the decision strategies implemented by participants while controlling for non-essential differences between conditions¹⁰.

⁹The Bayes factor was calculated in the same way as Chapter 2, using an estimate of the mean difference from Filoteo et al., (2010), and the expected difference was modelled as a two-tailed normal distribution with a standard deviation equal to half the mean.

¹⁰ Following the method of Edmunds et al. (2015), an analysis of verbal reports given by each participant about the strategy they used to respond was completed. Responses were rated as to whether the participant used a unidimensional or two-dimensional explicit strategy, an implicit non-verbal strategy or a miscellaneous strategy (e.g., trying to memorise the lines in this stimuli set). In the RB condition, 10 participants reported using a two-dimensional verbal strategy, 6 a unidimensional verbal strategy, 4 reported using miscellaneous strategies, one of whom reported sometimes answering on instinct. In the II condition 8 participants reported using a two dimensional verbal strategy, 9 a unidimensional verbal strategy, and 2 reported using miscellaneous strategies.

Table 3.1

Proportion Of Participants Using The Conjunctive (CJ), Information-Integration (II), Unidimensional (UD), Or Random (RND) Strategies In The RB And II Conditions In Both Scanning Sessions

	CJ	II	UD	RND
Rule-Based	0.50	0.40	0.10	0.00
Information-integration	0.21	0.63	0.11	0.05

However, the accuracy of these GRT modelling results are limited (see Donkin et al., 2014; Edmunds, Milton, & Wills, 2015, for a discussion) and COVIS studies rarely carry out these analyses (but see Ashby, Maddox & Bohil, 2002). Also, this model-based approach leaves a very small sample size leading to poorly powered imaging results. Therefore, the present study will only include whole-group comparisons rather than using only the subset of participants who used the optimal strategy according to the GRT modelling.

Imaging analysis of both learning sessions

The following analyses focus on neural activity during the feedback processing stage as this was the main interest of the present chapter (please see Appendix 3.1 for supplementary analyses showing brain activity during the decision making phase). A conjunction analysis was completed to investigate common activation between the correct RB - odd-or-even and the correct II - odd-or-even contrasts (both with thresholds of $p < .001$ and 23 contiguous voxels). Consistent with Experiment 1, the conjunction analysis revealed extensive areas of common activation between RB and II conditions during feedback processing in areas implicated in the explicit COVIS system including the bilateral hippocampus, right parahippocampal gyrus (BA 27), right caudate head and areas of the bilateral frontal lobes comprising BA's 6, 8, 9, 10, 46 and 47. There were also areas of activation corresponding to the implicit system of COVIS,

including the bilateral caudate body/tail and the putamen (Figure 3.5.a). Activation in the RB and II conditions was also directly contrasted by subtracting correct II trials - the odd-or-even trials from correct RB trials - the odd-or-even trials and vice versa. There were no areas more active in RB participants compared with II participants. Converging with the results of Experiment 1, the left hippocampus (cluster size 27) was more activated in the II condition than the RB condition (Figure 3.5.b), but the differences between conditions were otherwise restricted. As in Experiment 1, a ROI analysis was performed to more sensitively investigate areas of activation corresponding to the proposed implicit regions of COVIS (the caudate body and tail, the putamen and the substantia nigra). However, no regions were activated above threshold ($p=0.005$, 10 contiguous voxels) in II learning compared with RB learning.

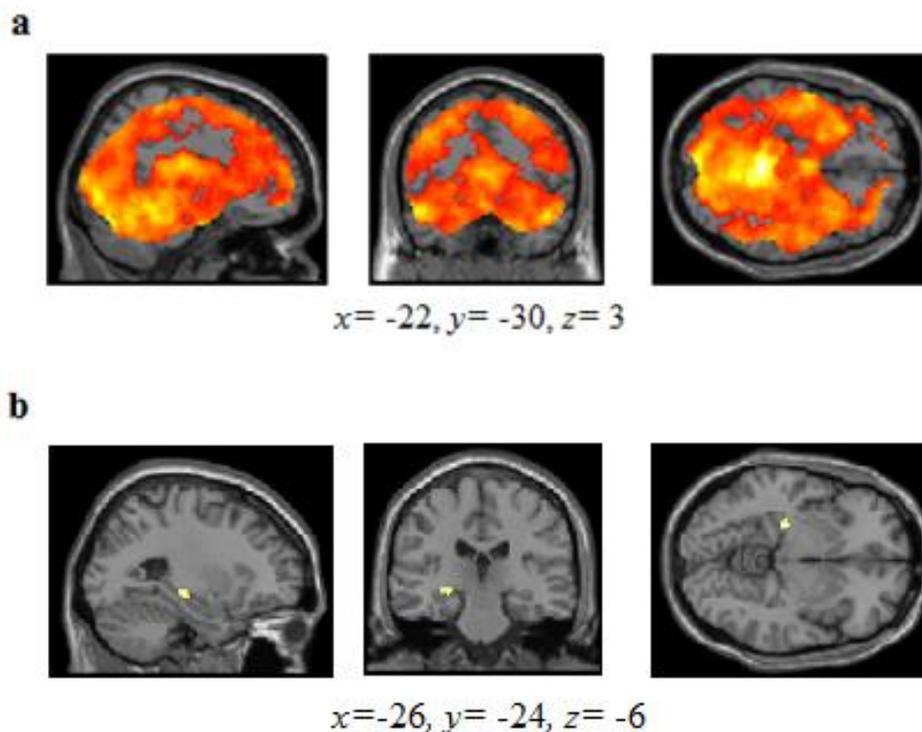


Figure 3.5. (a) Whole brain analysis comparing correct categorization trials to baseline trials in all runs of the study. Areas of common activation in the II and RB conditions during feedback processing; (b) whole brain analysis of the neural areas more activated in II participants than RB participants. All analyses thresholded at $p < .001$ and 23 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Imaging analysis of session 1

The neural activation for the first session alone was then investigated, where the effect of feedback should be of particular importance as participants begin to learn the category structure. A conjunction analysis showing areas of common activation in RB and II trials revealed areas engaged that are associated with the explicit system, including the bilateral caudate head, parahippocampal gyrus (bilateral BA 30 & 16; left BA 35 & 19) and the left hippocampus. Areas associated with the implicit system were also significantly activated in both conditions, including the bilateral putamen, caudate body and caudate tail (see Figure 3.6.a).

Activation in the RB and II conditions was directly compared; however, no regions were more activated in the RB condition than the II condition. In contrast, there were regions more engaged in II participants than in RB participants during feedback processing (see Figure 3.6.b), including the bilateral superior and medial frontal gyrus (BA 6), the left middle occipital gyrus (BA 18) and the right posterior cingulate. An a priori ROI analysis comprising the MTL ($p=0.005$, 10 contiguous voxels) was also conducted based on the finding that this region was more activated in the II condition than the RB condition in Experiment 1. The bilateral hippocampus and parahippocampal gyrus were active more in the II condition than the RB condition (see Figure 3.6.c). As in the analyses of all runs, a ROI analysis comprising of the implicit system of COVIS (the caudate body and tail, the putamen and the substantia nigra) was performed. However, no regions were activated above threshold ($p=0.005$, 10 contiguous voxels) in II learning compared with RB learning.

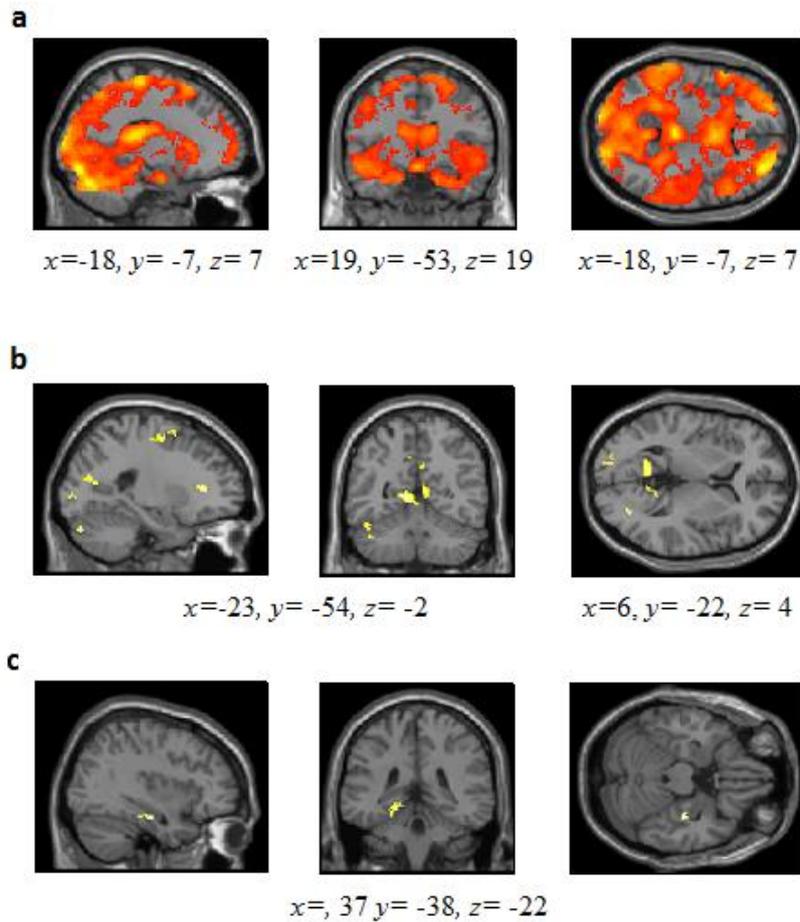


Figure 3.6. Whole brain analysis comparing correct categorization trials to baseline trials in the first session of the study; (a) areas of common activation in the II and RB conditions during feedback processing; (b) areas more activated in II participants compared with RB participants during feedback processing. Analyses thresholded at $p < .001$ and 23 contiguous voxels; (c) ROI analysis showing the areas of the MTL more activated in II participants compared with RB participants, analysis thresholded at $p = .005$ and 10 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Imaging analysis of session 2

COVIS predicts that while the explicit system is optimal for RB learning, it should also be engaged during initial II learning until this system is proven ineffective, and a switch is made to the implicit system (Filoteo et al., 2010). Therefore, neural activation during the last 240 trials (scanning session 2) alone was also analysed to investigate whether evidence for the pattern of results predicted by COVIS could then be observed.

A conjunction analysis showed areas of common activation between the RB and II conditions (Figure 3.7.a) including areas associated with the explicit system of COVIS as the bilateral caudate head, and left hippocampus/parahippocampal gyrus were active along with areas of the frontal lobes (e.g., bilateral BA 6, 8, and 47; right BA 9, and 46; and left BA 44). As found in Experiment 1, the bilateral caudate body, an area implicated in the implicit system, was also commonly active in the RB and II conditions (see Figure 3.7.b).

Activation in the RB and II conditions was directly compared as above. As in the analyses of session 1, no regions were more engaged in the RB condition than in the II condition. However, numerous areas were recruited more in the II condition compared with the RB condition (Table 3.2) including the left hippocampus and bilateral parahippocampal gyrus (left BA 27 and right BA 35), and the left putamen. In an a priori ROI analysis of the MTL (applied as before at a threshold of $p=.005$, 10

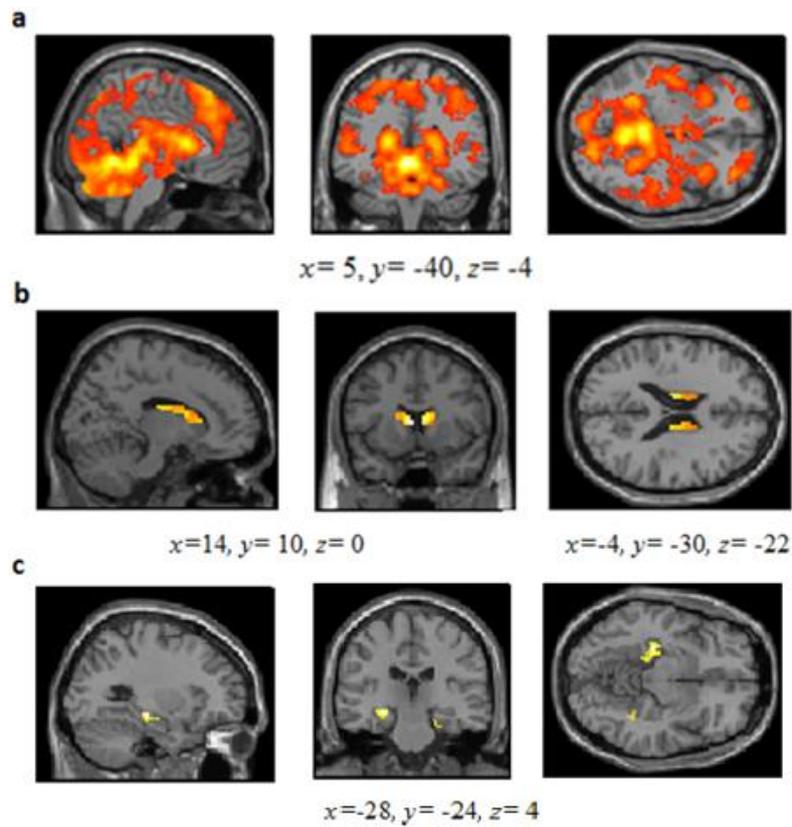


Figure 3.7. Whole brain analysis comparing correct categorization trials to baseline trials in the second run of the study; (a) areas of common activation in the II and RB conditions during feedback processing; (b) areas of the caudate body commonly activated during II and RB conditions during feedback processing. Analyses thresholded at $p < .001$ and 23 contiguous voxels; (c) ROI analysis of areas of the MTL more active in II participants compared with RB participants thresholded at $p = .005$, 10 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Table 3.2.

Brain Areas More Active in the II Condition Compared with the RB Condition in Session 2 of the Experiment

Region	Cluster size	BA	Talairach Coordinates			z-score
			x	y	z	
Right Claustrum	169	-	32	-1	15	4.65
Right Insula		13	40	4	7	4.13
Left Putamen	247	-	-26	-20	-4	4.36
Left Lateral Geniculum Body		-	-20	-27	-2	4.14
Left Claustrum		-	-34	-17	1	4.12
Left Putamen		-	-28	-18	-8	4.25
Left Hippocampus		-	-28	-24	-7	4.12
Left Parahippocampal Gyrus		27	-20	-31	-5	3.31
Right Superior Temporal Gyrus	46	41	50	-28	16	3.50
Right Postcentral Gyrus		40	61	-26	20	3.17
Left Middle Temporal Gyrus	38	39	-38	-71	22	4.12
Right Inferior Frontal Gyrus	77	47	38	31	-5	3.99
Left Posterior Cingulate	23	31	-26	-61	23	3.98
Right Parahippocampal Gyrus	47	35	20	-22	-17	3.84
Left Inferior Parietal Lobule	36	40	-50	-36	26	3.79
Left Insula		13	-44	-32	22	3.42
Right Amygdala	23	-	24	-7	-23	3.69

Note that indented rows indicate voxels in the same cluster as the non-indented rows above

contiguous voxels) the left hippocampus and bilateral parahippocampal gyrus (left BA 27; right BA 19, 28, 35 & 36; see Figure 3.7.c) were, as expected, more active in the II condition than the RB condition. However, when applying the implicit ROI mask (including the caudate body/tail, the putamen, and the substantia nigra) the left putamen was also more active in II participants than in RB participants (two separate clusters of 33 and 27 voxels).

Imaging analysis comparing session 1 and session 2

To investigate the difference in activation between each scanning session, the first 240 trials (session 1) and last 240 trials (session 2) were first compared for the RB and II conditions separately.

During feedback processing in the RB condition, the right hippocampus, bilateral parahippocampal gyrus (right BA 36; left BA 30 and 19), left middle frontal lobe (BA 6), and putamen (Figure 3.8.a) were more active during session 1 compared with session 2. However, in the II condition there were no areas more engaged in session 1 than session 2. Neural areas more active in session 1 compared with session 2 were also contrasted between RB and II participants, but no significant differences emerged.

When assessing areas more active in session 2 than session 1 a different pattern of results was observed. There were no areas more active during session 2 compared with session 1 in RB participants during feedback processing. However, the left hippocampus and right parahippocampal gyrus (BA 28), and the left inferior frontal gyrus (BA 45; see Figure 3.8.b) were engaged more in session 2 than session 1 in the II condition. No significant differences, though, were identified when directly contrasting the RB and II conditions.

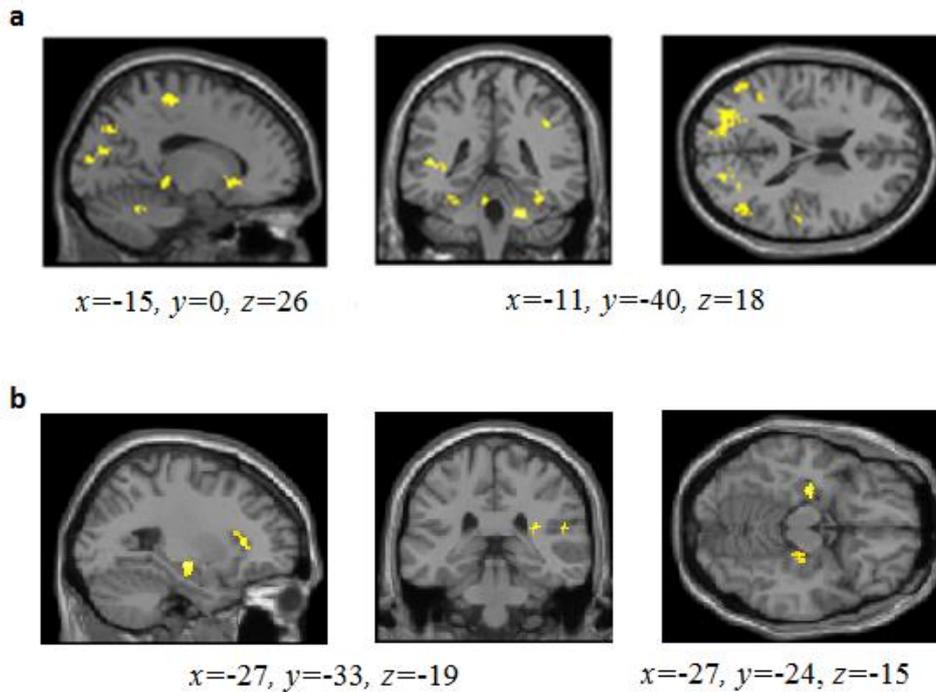


Figure 3.8. Whole brain analyses of areas more activated when comparing correct categorization trials to baseline trials; (a) in first session of the study compared with the second session in the RB condition; (b) in the second session of the study compared with the first session in the II condition. Analyses thresholded at $p < .001$ and 23 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Analyses of Correct and Incorrect Trials

Session 1

Experiment 1 found the right hippocampus more activated during correct trials than incorrect trials in II participants compared with RB participants. In order to investigate if this pattern of activity was also present in Experiment 3, a comparison of areas more active during positive feedback processing compared with negative feedback processing in the first learning session was completed for the RB and II conditions separately. While comparing activation during correct and incorrect trials allows for inferences about neural areas potentially more engaged in positive or negative feedback, as discussed in Chapter 2, using incorrect trials in neuroimaging work makes

interpretation problematic as it is not possible to infer why participants failed each incorrect trial, so these analyses should be taken with some caution.

Figure 3.9.a shows areas more engaged during positive compared with negative feedback in RB participants. Widespread activation was found including recruitment of the bilateral parahippocampal gyrus (right cluster size 61; left cluster size 38); right caudate body/tail (cluster size 32); and the right putamen, along with areas of the frontal lobes (bilateral BA 7, 8, and 10, left BA 47, right BA 6 and 9), and the temporal lobes. Similarly, in II participants the bilateral hippocampus/parahippocampal gyrus (left cluster size 872; right cluster size 476), the right caudate head (cluster size 26), left putamen (cluster size 872), the frontal lobes (left BA 8, 11 and 47; right BA 4, 5, 6, and 10), the occipital gyrus and the temporal lobes were more active during positive feedback than negative feedback processing (Figure 3.9.b).

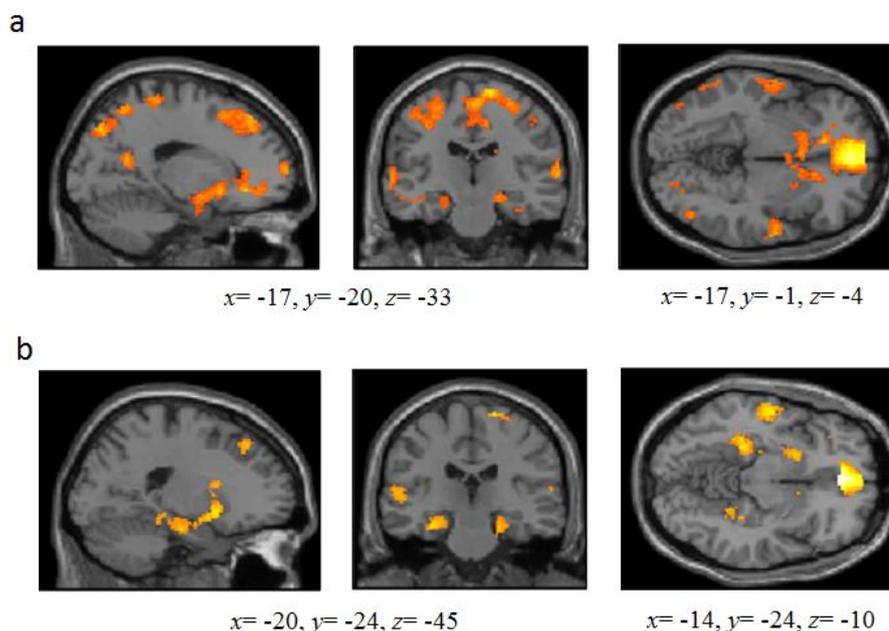


Figure 3.9. Whole brain analysis of areas more active in correct trials compared with incorrect trials in the first scanning session; (a) areas activated in the RB condition; (b) areas activated in the II condition. Analyses thresholded at $p < .001$ and 23 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

An a priori MTL ROI analysis (with thresholds of $p=.005$, 10 contiguous voxels) revealed no regions recruited more in the II condition than in the RB condition during positive feedback compared with negative feedback, and the whole brain analysis of this contrast also revealed no areas more engaged. When reversing this whole brain analysis to investigate brain regions recruited more in RB than II participants, areas of the parietal and frontal lobes were found to be more active during positive feedback than negative feedback (Table 3.3.a).

When analysing areas more engaged during incorrect feedback presentation compared to the baseline task (incorrect activation – odd-or-even activation), II participants demonstrated areas of greater recruitment compared to RB participants in the left parahippocampal gyrus (BA 36), and the left superior temporal gyrus (BA 19; see Table 3.3.b). However, this analysis revealed no areas more activated in RB participants compared with II participants.

Areas more active during incorrect feedback processing than correct feedback processing (incorrect activation – correct activation) were also examined. There were no brain areas more engaged in II participants than RB participants in this analysis. However, RB participants demonstrated greater activity than II participants in negative feedback processing compared with positive processing in the bilateral frontal lobes (BA 6), and the left superior parietal lobes (BA 7; Table 3.3.c).

Table 3.3.

(a) Areas More Activated In The RB Condition Than The II Condition During Correct Feedback Compared With Incorrect Feedback; (b) Areas More Activated In The II Condition Than The RB Condition During Incorrect Feedback Compared With Baseline Trials; (c) Areas More Activated In The RB Condition Than The II Condition During Incorrect Feedback Compared With Correct Feedback.

Region	Cluster size	BA	Talairach Coordinates			z-score
			x	y	z	
<i>a. Correct - Incorrect (RB-II)</i>						
Left Precentral Gyrus	262	6	-30	-9	48	4.39
Left Middle Frontal Gyrus		6	-36	-6	41	4.24
Left Precuneus	256	7	-12	-62	46	4.25
Left Superior Parietal Lobe		7	-24	-66	48	4.00
Left Superior Frontal Gyrus	68	6	-24	14	49	4.09
Right Medial Frontal Gyrus	48	6	8	-12	63	3.99
<i>b. Incorrect - Baseline (II-RB)</i>						
Left Parahippocampal Gyrus	40	36	-22	-35	-8	4.33
Left Declive	33	-	-20	-83	-19	4.09
Left Cuneus	35	18	-24	-67	18	3.81
Right Precentral Gyrus	43	4	57	-10	32	3.77
Right Culmen	91	-	26	-47	-11	3.67
Left Superior Temporal Gyru	53	22	-44	-53	18	3.67
Left Tuber		-	-44	-54	-26	3.49
Left Posterior Cingulate	30	30	-10	-52	6	3.63
<i>c. Incorrect - Correct (RB-II)</i>						
Left Precentral Gyrus	336	6	-30	-9	48	4.32
Left Superior Frontal Gyrus		6	-24	14	49	3.93
Left Sub-Gyral		6	-22	-3	54	3.93
Right Medial Frontal Gyrus	37	6	8	-12	63	3.81
Left Precuneus	137	7	-12	-62	46	3.78
Left Superior Parietal Lobe		7	-24	-66	48	3.68

Note that indented rows indicate voxels in the same cluster as the non-indented rows above

Session 2

The above analyses were then completed for session 2. Figure 3.10.a shows areas more active during positive feedback than negative feedback in RB participants. Widespread activation was found including engagement of the bilateral hippocampus/parahippocampal gyrus (right BA 28 and 36, cluster size 1348; left BA 35, cluster size 1348), the left caudate head, the right putamen (cluster size 1348), along with areas of the bilateral inferior and superior, and left middle frontal lobes (left BA 3, 4, 6, 8, 10, and 47), the bilateral middle occipital gyrus, and the bilateral superior, middle, and inferior temporal lobes. Similarly, in II participants the bilateral hippocampus/parahippocampal gyrus (cluster size 4205), the right caudate body/tail (cluster size 241), the bilateral middle and superior, and left inferior frontal lobes (bilateral BA 6, 8, and 10; left BA 9 and 47; right BA 3, 4, and 5), the occipital gyrus and the bilateral middle and right inferior and superior temporal lobes were more active during positive feedback than negative feedback processing (Figure 3.10.b).

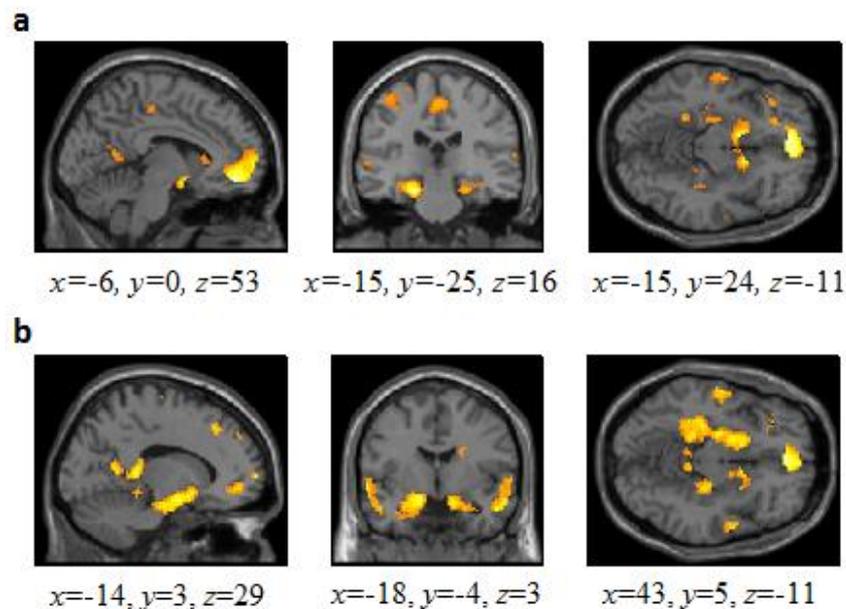


Figure 3.10. Whole brain analysis of areas more active in correct trials compared with incorrect trials in the second scanning session; (a) areas activated in the RB condition; (b) areas activated in the II condition. Analyses thresholded at $p < .001$ and 23 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

As before, a ROI analysis of the MTL (with thresholds of $p=.005$, 10 contiguous voxels) revealed the right hippocampus was more activated in the II condition than in the RB condition during positive feedback compared with negative feedback (peak voxel: 34, -13, -21; cluster size 24). However, in a whole brain analysis no areas were more active in positive feedback than negative feedback processing in II participants compared with RB participants. When reversing this analysis to investigate brain regions more engaged in RB participants than II participants no areas were identified.

When analysing areas more implicated during incorrect feedback presentation compared to the baseline task (incorrect activation – odd-even-activation), RB participants demonstrated areas of greater activation compared to II participants in the left hippocampus, left putamen, and right middle occipital gyrus (BA 19; see Table 3.4). However, this analysis revealed no areas more activated in II participants compared to RB participants.

Areas more active during incorrect feedback processing than correct feedback processing (incorrect activation – correct activation) were also examined but no significant differences were found between the RB and II conditions.

Table 3.4.

Areas More Activated in the RB Condition than the II Condition During Incorrect Feedback Compared with Baseline Trials.

Region	Cluster size	BA	Talairach Coordinates			z -score
			x	y	z	
Left Hippocampus	70	-	-26	-24	-4	4.01
Left Thalamus		-	-22	-27	1	3.83
Left Putamen		-	-28	-16	-8	3.44
Right Fusiform Gyrus	44	19	34	-64	-4	3.92
Right Middle Occipital Gyrus		37	36	-66	0	3.60
Right Claustrum		-	38	-7	-1	3.60
Right Inferior Occipital Gyrus		19	36	-70	-2	3.30
Left Culmen	23	-	-10	-29	-27	3.92

Note that indented rows indicate voxels in the same cluster as the non-indented rows above

Discussion

Experiment 3 investigated the neural correlates of feedback processing during RB and II category learning. Behaviourally it was found that there were no significant differences in accuracy between category structures, again confirming that learning rates are the same for the II and conjunctive RB category structures. When analysing activation present during feedback compared with the odd-or-even baseline there was widespread activation across the brain in both the RB and II conditions, suggesting that the neural patterns present in decision making in Experiment 1 generalise to feedback processing. Concurring with Experiment 1 the bilateral MTLs were again more activated in the II condition during correct feedback processing compared with the RB condition in both sessions of the study. However, as in Experiment 1 no areas were more active in the RB condition compared with the II condition.

Also, as found in Experiment 1, the right hippocampus was more activated in II participants than RB participants in correct trials compared with incorrect trials. This supports the conclusions made by Cincotta and Seger (2007) that this area is activated when receiving corrective feedback in II conditions. The MTL was more activated in the first learning session than the second learning session in RB participants during feedback processing. However, in II participants the MTL was found to be more activated in learning session 2 compared with session 1. This could suggest that the MTL is involved in feedback processing during both RB and II conditions, but is potentially engaged in different learning stages in each task. However, given this interaction was not significant, these results should not be over interpreted.

General Discussion

Chapter 3 aimed to assess the predictions of COVIS that RB and II category structures engage different neural feedback processing systems (Ashby et al., 1998; Ashby & Valentin, 2005; Maddox et al., 2003). While Chapter 2 investigated neural responding to decision making, the design of Experiment 1 meant that the neural response to stimulus and feedback presentation could not be distinguished, as the neural activity of the processes occurring at stimulus presentation were temporally correlated with the neural activation occurring during feedback presentation. Experiment 2 was a behavioural study designed to assess whether the addition of a varied time delay between stimulus and feedback processing would impair II performance. Previous research has found that delaying feedback by 5 seconds can hinder II learning compared with RB learning (e.g., Maddox et al., 2003; Maddox & Ing, 2005), and while the present work only delayed feedback by a maximum of 1500ms in order to account for this concern, it was still necessary to confirm that this smaller delay did not impair II participants. Neither RB nor II participants demonstrated worse performance in the delayed condition compared to an immediate feedback condition, and critically, the effect of delay was not modulated by the category structure participants learnt. This was an essential requisite for Experiment 3 as if the II condition were hindered by the feedback delay then this may inhibit the use of any potential implicit system. Instead, in these circumstances, according to COVIS, both structures would engage the explicit system and one would not expect to detect any neural differences.

While finding no evidence of selective II impairment after delayed feedback is contrary to the predictions of COVIS and previous work (e.g., Maddox et al., 2003; Maddox & Ing, 2005; Worthy et al., 2013), comparisons between these studies and the present research should be made cautiously. Experiment 2 had a delay between stimulus

and feedback presentation that was at least 1 second shorter than Maddox et al. (2003) and 3.5 seconds shorter than Maddox and Ing (2005) to minimise the potential disruption to II learning. Nevertheless, it does undermine the basic contention of COVIS theorists that II learning is more sensitive to delayed feedback than RB learning and, in this regard, our results are in line with those of Dunn et al. (2011), who found that II categories were not impaired by feedback delay when category learning procedures do not contain methodological artefacts.

Experiment 3 employed the same basic procedure as Experiment 2 to investigate whether separable neural systems of feedback processing were present in RB or II learning. There was extensive neural overlap between both conditions during feedback processing, with RB and II participants recruiting regions associated with the explicit COVIS system, namely the bilateral caudate head and the MTLs, as well as areas implicated in the implicit system of COVIS including the bilateral putamen and caudate body. This pattern was found when assessing areas of activation in all runs of the study, as well as in the first and second half of training separately. These results mirror the common activation found between conditions in Experiment 1, which focused on activation during the decision process. The similar results support and extend the conclusions drawn from Experiment 1, and indicate that both RB and II learning share many common neural mechanisms.

However, as discussed in Chapter 2, Ashby and Crossley (2010) could predict extensive neural overlap during both stimulus and feedback presentation in RB and II learning. They suggest that both the implicit and explicit systems would be simultaneously activated in each condition, but the suboptimal system would then be suppressed by the dominant system when the response to a stimulus is made. On the other hand, the predictions of Ashby and Crossley do not explain why the MTL, an area

associated with the explicit system (Nomura & Reber, 2008), is found in both Experiments 1 and 3 to be more active in II conditions than RB conditions. Experiment 3 consistently found the MTL (hippocampus/parahippocampal gyrus) more activated in II feedback processing than RB feedback processing. Therefore, there is converging evidence for the findings of Experiment 1 that the MTL is more engaged during II trials than RB trials, and supports the conclusion that the MTL activation found in Nomura et al. (2007) resulted from the use of a unidimensional RB category that had a perceptually less discriminable decision boundary than the II condition. Experiment 3 also generalises this neural dissociation in MTL activity to feedback processing, and can be considered notable as a number of COVIS studies have indicated that feedback processing in II learning differs from RB learning (e.g., Maddox et al., 2003; Maddox & Ing, 2005; Worthy et al., 2013). The present results also converge with conclusions of Dunn et al. (2011) and Newell et al. (2013) who argue that the apparent behavioural dissociations between RB and II performance from feedback manipulations are driven by methodological problems.

However, unlike Experiment 1, Experiment 3 found the putamen, an area associated with the implicit system proposed by COVIS (Waldschmidt & Ashby, 2011), more active in II compared with RB feedback processing in the second half of training (however, not when analysing both sessions or session 1 alone). There has been some debate over whether the putamen is active only in II learning, as it has been found that focal lesions to the putamen impair RB but not II conditions, suggesting that this area plays more of a role in explicit category learning (Ell, Weinstein, & Ivry, 2010). So whilst this result could be taken as providing moderate support for COVIS, the broad pattern of results including the greater MTL activation in the II condition than the RB condition, are, as in Experiment 1, highly problematic for COVIS as it is currently instantiated.

In the first half of training there were also areas outside the MTL more active in II participants than RB participants including the frontal lobes (BA 6). This frontal lobe activation converges with previous studies which have found analogous results in this brain region (e.g., Milton & Pothos, 2011). This activation may reflect the recruitment of rule-based processes, and the greater engagement of this region in the II than in the RB condition may plausibly reflect the more complex rules that would be required for learning the II structure.

One noteworthy finding was that the MTL was engaged differently in the first and second half of training in RB and II participants. In RB participants the MTL was more activated in session 1 than session 2. In RB learning, there is a greater reliance on an abstract rule and this MTL activation may reflect the involvement of this region in the formulation of particular rules. However, as learning progresses and a rule proven effective there is less need to formulate such rules as participants' can simply apply the successful one they generated earlier. However, in II participants the MTL was more active in session 2 than session 1. The greater activation in the latter half of training in II participants could represent the storage of category clusters by the MTL (Love & Gureckis, 2007) as discussed in Chapter 2. In the II condition there are more stimuli that are exceptions to a rule throughout learning that must be stored in new clusters. By the nature of II category structures, in later learning participants would retrieve information from more clusters than RB participants, as well as continue to generate more clusters for exception stimuli. Therefore, in later II learning the MTL is more engaged than in early learning to deal with this larger storage and retrieval demand. Yet, as there were no significant differences in activation between RB and II conditions when directly comparing brain activation between scanning sessions, this interpretation should be taken with caution.

Experiment 1 of this thesis found that while there was extensive commonality in brain regions active in RB and II learning, contrary to the predictions of COVIS there was greater MTL activation in II participants compared with RB participants. One caveat of Experiment 1 is that the activation measured likely included decision making and feedback processing stages (with an emphasis on the decision making stage). Experiment 3, therefore, decorrelated the presentation timings of the decision and feedback stages, and made feedback processing the measured event to find a converging pattern of results that extended the findings of Experiment 1 to feedback processing. The advantage of this procedure was that it was the first study to isolate neural activation during RB and II feedback processing, allowing for a more nuanced understanding of the neural mechanisms of this process in category learning. The present results offer little evidence for the predictions of COVIS.

Appendix 1 Chapter 3: Imaging analysis of activation during decision making

Imaging analysis of both learning sessions

Neural activity at the onset of stimulus presentation was analysed to investigate the brain areas engaged during category decision making (as in Experiment 1). All analyses implemented thresholds of $p < .001$ and 23 contiguous voxels. A conjunction analysis examining areas commonly activated in the ‘correct RB - odd-or-even’ contrast and the ‘correct II - odd-or-even’ contrast revealed extensive activation overlap between conditions (Figure S.3.1). Regions activated included areas associated with the explicit system of COVIS, namely the bilateral anterior cingulate and parahippocampal gyrus (BA 27, left BA 19 and 30), the left hippocampus and right caudate head, along with areas of the frontal lobes (right BA 2, 6, and 8; left BA 47). Areas associated with the implicit system of COVIS including the bilateral caudate body/tail, the right putamen and the bilateral substantia nigra, were also recruited in both learning conditions.

To investigate potential neural dissociations between category structures, the brain activity in the RB and II conditions were directly contrasted. This was performed in the same way as in Experiment 1 by subtracting ‘correct II trials - the odd-or-even trials’ from ‘correct RB trials - the odd-or-even trials’ to reveal areas more active in the RB condition than in II learning. This calculation was reversed to demonstrate areas more active in II learning than RB learning. In a whole brain analysis, the left hippocampus (cluster size 27) was more active in the II condition compared with the RB condition. In a ROI analysis, a mask comprising the caudate body/tail, putamen, and substantia nigra examined whether any areas associated with the implicit system were more active in II learning than RB learning at the threshold of $p = .005$ and 10 contiguous voxels, but none emerged.

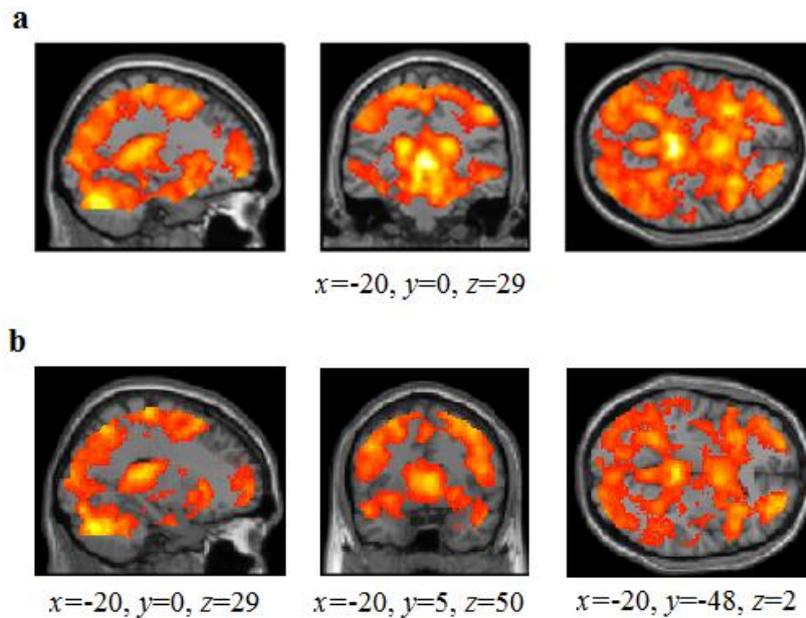


Figure S.3.1. Analysis of activity during decision making in correct trials compared with the baseline task; (a) whole brain analysis of areas commonly activated in the II condition compared with the RB condition in all runs of the study; (b) whole brain analysis of areas commonly activated in the II condition compared with the RB condition in scanning session 1. Analyses thresholded at $p < .001$ and 23 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Imaging analysis of learning session 1

A conjunction analysis revealed widespread common activation between the RB and II conditions during stimulus presentation in learning session 1 (Figure S.3.1.b). Regions engaged included areas associated with the explicit system; the bilateral parahippocampal gyrus and right hippocampus, the bilateral anterior cingulate, and the right caudate head were active in both conditions along with areas of the frontal lobes (bilateral BA 47, right BA 8, 9, 10 and 45, and left 46 and 44). Areas associated with the implicit system were also found to be active including the bilateral caudate body/tail and the bilateral putamen.

Unlike the feedback analysis, there were areas more active in RB participants than II participants in the first learning session. In a ROI analysis focused on the MTL, the left parahippocampal gyrus was recruited more in RB participants compared with II

participants (BA 36; cluster size 20). In a whole brain analysis of areas more engaged in RB participants than II participants the right precuneus was significantly active. When reversing these analyses and investigating areas more active in the II condition compared to the RB condition there were no regions significantly activated in the whole brain analysis, or in the ROI analyses when applying the MTL and implicit ROI masks.

Imaging analysis of learning session 2

A conjunction analysis revealed areas of extensive activation overlap in the RB and II conditions during decision making in learning session 2 (Figure S.3.2.a). The left parahippocampal gyrus (BA 36; cluster size 61) and the frontal lobes (right BA 8; left BA 10, 13, and 44; cluster size 824), areas implicated in the explicit system, were active in both conditions, as were the bilateral caudate body (right cluster size 305; left cluster size 315) and bilateral caudate tail (right cluster size 23; left cluster size 15), regions proposed to be active in the implicit category learning system.

When directly comparing activation in the II and RB conditions during decision making, there were no areas more activated in the RB condition compared with the II condition. However, when looking at areas more active during II learning than RB learning, a MTL ROI analysis implicated the bilateral parahippocampal gyrus and left hippocampus (see Figure S.3.2.b). In a whole brain analysis of the areas more engaged in the II condition compared to the RB condition, the right inferior frontal gyrus (BA 47) was significantly activated (Figure S.3.2.c) as well as areas of the bilateral parahippocampal gyrus (right BA 35 and 28; left BA 28) and right hippocampus. An ROI analysis comprising the regions associated with the implicit system of COVIS revealed that the left putamen (cluster size 33) was more active in II participants than RB participants.

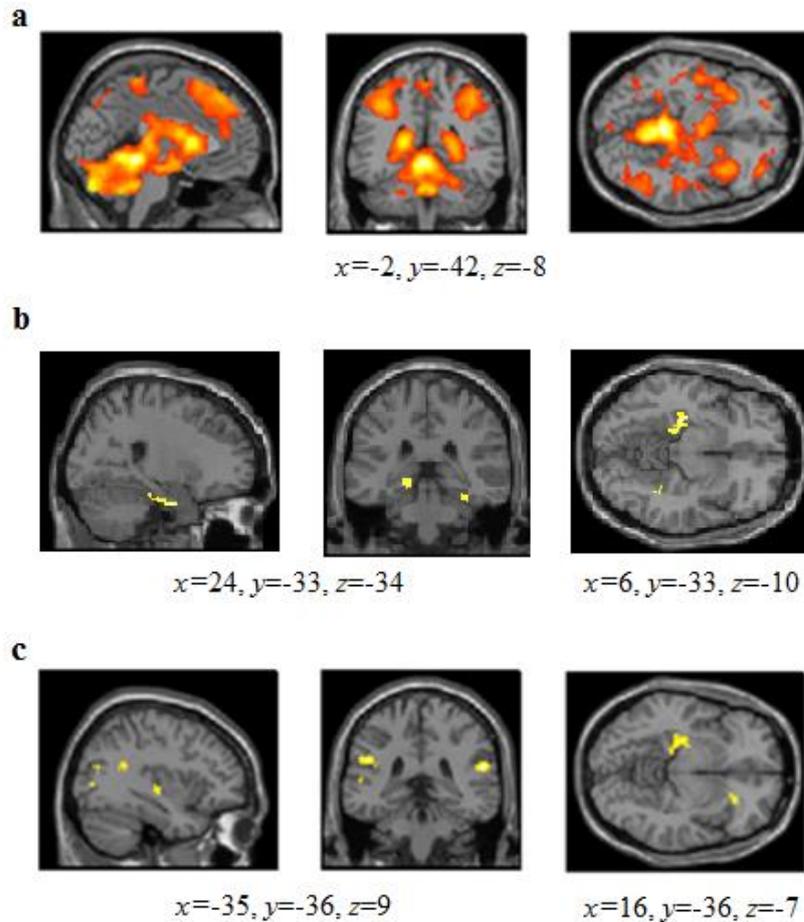


Figure S.3.2. Whole brain analyses comparing correct categorization trials to baseline trials in the second run of the study; (a) areas of common activation in the II and RB conditions. Analysis thresholded at $p < .001$ and 23 contiguous voxels; (b) a ROI analysis of areas more activated in the MTL in the II condition compared with the RB condition. Analysis thresholded at $p = .005$ and 10 contiguous voxels; (c) areas of activation greater in the II condition compared with the RB condition. Analysis thresholded at $p < .001$ and 23 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

As found in Experiment 1, category decision making in the RB and II conditions displayed common activation in areas hypothesised to be implicated in the separable explicit and implicit systems proposed by COVIS. This effect prevailed when analysing sessions one and two alone. Converging with evidence reported in Experiment 1, the left hippocampus was more activated in II participants than in RB participants when analysing both learning sessions together, and when assessing learning session two alone. However, in learning session one, an ROI suggests that a small region of the left

parahippocampal gyrus was more active in RB participants than in II participants. The pattern of MTL activation could reflect RB participants finding and storing a successful category learning rule earlier on in training than II participants, who may engage the MTL more in later learning to store and retrieve information about exemplars that are exceptions to a rule (e.g., Love et al., 2012). This analysis replicates the key findings of Experiment 1, supporting the assumption that the explicit system is operational not only during RB categorisation but also during II learning, a process COVIS considers purely driven by an implicit system.

Chapter 4. The effect of deferring corrective feedback on rule-based and information-integration category learning

Chapters 2 and 3 of this thesis evaluated COVIS from a neuroimaging perspective. Chapter 4 will look at this debate from a behavioural angle by re-examining the recent claim put forward by Smith, Boomer, Zakrzewski, Roeder, Church, and Ashby (2014) that deferring feedback adversely affects II learning but not RB learning. Smith et al. (2014) propose that this finding “constitutes one of the strongest RB-II dissociations yet seen” (Pp. 454) and cannot be accommodated by single process theories. The focus of this chapter is to scrutinise the claims made by Smith et al. in the first follow-up investigation of their paper.

Smith et al. (2014) presented participants with traditional unidimensional RB or II category structures to learn. Participants received either immediate feedback after every trial, or deferred feedback collectively presented after every six categorization trials. There was a significant interaction - II participants receiving deferred feedback were impaired compared to II participants who received immediate feedback, yet RB participants were unaffected by the feedback manipulation. This pattern of results was further supported by modelling analyses. The majority of RB participants used an optimal unidimensional learning strategy regardless of the feedback condition. II participants in the immediate condition also learnt using an optimal II strategy; however, the majority of participants in the deferred II condition used a suboptimal unidimensional RB strategy to learn the category structure.

COVIS predicts this precise pattern of results. Learning in RB participants should remain intact when receiving deferred feedback as performance in this condition would be optimally performed by the explicit system which stores the responses made to each trial, as well as the category learning rule, in WM. Through reliance on WM this

explicit system is able to hold category learning information for later use, and therefore is unimpaired by deferred feedback as information can be stored until the later feedback presentation. This assumption led the authors to predict that RB learning would ‘flourish’ (Pp. 450) in deferred feedback conditions. However, COVIS also predicts that II learning would be disadvantaged by deferred feedback as immediate feedback is needed for the optimal learning of the II structure (see also Chapter 3, pages 73-74). According to COVIS, optimal II learning is performed by the implicit system, which procedurally associates the representations of the perceived stimulus and the response made through a dopamine release initiated when presented with positive feedback. However, if feedback is delayed or deferred then the neural representation of the response made and the stimulus seen will have decayed, and therefore the representations cannot be associated when dopamine is released (Ashby et al., 1998).

Although Smith et al. (2014) present evidence that II categories are differentially affected by the feedback manipulation compared to unidimensional RB categories, there was just a single, un-replicated experiment demonstrating an interaction between feedback type and category structure, which was only significant at $p=.048$. A number of behavioural dissociations consistent with COVIS do not replicate well (e.g., Zeithamova et al., 2006; Newell, Dunn & Kalish, 2010) and therefore there is a need to verify the findings of Smith et al. and the conclusions they draw from the data reported given the strength of the claims that they make. The first aim of Chapter 4 is to try and replicate the behavioural dissociation found between unidimensional RB and II learning when deferring feedback.

However, as discussed in Chapter 2 (see pages 66 - 67), there are studies in the literature that argue that dissociable performance in RB and II conditions can be attributed to using unidimensional categories rather than conjunctive RB categories as a

comparison for II performance. For example, strong evidence for this argument is presented by Edmunds et al. (2015), who reassessed a behavioural dissociation reported by Ashby, Maddox, and Bohil (2002). Ashby et al. found that trial-by-trial feedback training led to better performance than observational training (where participants were presented with a stimulus after being shown the category assignment of that trial) in II learning, but with equal learning between the observational and feedback conditions in unidimensional RB category learning. Ashby et al. argued that optimal II performance relied upon the implicit system in which stimulus-response associations are necessary for effective learning. Therefore, II learning was hindered by observational feedback as seeing the category assignment before the stimulus did not initiate the dopamine reward response necessary to form stimulus-response associations. On the other hand, as RB learning relied upon a WM based, explicit feedback system the disruption of this association process was irrelevant to RB performance.

However, Edmunds et al. argued that feedback training is of greater assistance than observational training more generally, but as the number of dimensions relevant to category learning increases, so too does the benefit of trial-by-trial feedback, as the task becomes more effortful, and therefore more reliant on optimal feedback presentation. To support this the authors found that when controlling for the number of relevant dimensions in the RB and II conditions, and matching them on relative difficulty by using the conjunctive RB and II category structures presented in Chapters 2 and 3, feedback learning was superior to observational learning in both conditions. This led Edmunds et al. to conclude that the use of an easy to learn unidimensional RB task as a comparison for II learning in Ashby et al. was driving the dissociation reported. Similarly, Chapter 2 found that when accounting for the number of dimensions relevant for category learning in RB and II conditions, neural dissociations reported between unidimensional RB and II learning (e.g., Nomura et al., 2007) disappear.

This argument also appears to apply to the category structures employed by Smith et al. (2014), and by the same rationale could be the reason for the dissociation which Smith et al. argue is down to the engagement of qualitatively separate learning systems. Therefore, the present chapter aims to re-evaluate the bold claims made by Smith et al. (2014) that the procedure implemented in their work results in some of the strongest evidence of separable explicit and implicit category learning systems, and ascertain whether the dissociation they report could be due to the specific category structures implemented.

Experiment 4

The aim of Experiment 4 was to use the same category structures and stimuli as Smith et al. (2014) to try and replicate their finding that II learning was impaired by deferring feedback whereas RB learning was not. In a between-subjects design, participants learnt either the unidimensional RB or the II category structure of Smith et al. (2014) when receiving either immediate feedback after every response or deferred feedback presented after each block of six categorization trials.

Method

Participants and design

86 University of Exeter students completed the experiment (19 males, mean age, 20.24 years, SD=5.32). Participants received course credit or £5 remuneration for participation in the study. In Smith et al. (2014) participants were excluded if they performed significantly lower in the last 100 trials compared with the first 100 trials. I wished to replicate this aspect of their experiment too; however, Smith et al. did not specify how they calculated this exclusion criterion so I did this using a chi-square

analysis which seemed the most appropriate approach. One participant was excluded for performing significantly worse in the last 100 trials of the study compared with the first 100 trials, leaving 85 participants in the study.

The experiment used a 2 (category structure) X 2 (feedback type) design creating four conditions: RB learning with immediate feedback (RB immediate, n=20); RB learning with deferred feedback (RB deferred, n=21); II learning with immediate feedback (II immediate, n=20) and II learning with deferred feedback (II deferred, n=24).

Stimuli

The ‘dot distribution’ stimuli (as used by Smith et al., 2014) were unframed rectangles of green dots (see Figure 4.1). Each stimulus varied in the size of the rectangle and density of green dots within the rectangle. In the unidimensional RB condition (Figure 4.2.a), the size of the rectangle was the relevant dimension for learning, but for the II condition (Figure 4.2.b) both dimensions were relevant in a standard II category structure as used in Chapters 2 and 3. As in Smith et al. (2014) each

This image has been removed by the author of this thesis for copyright reasons.

Figure 4.1. An example of the stimuli used by Smith et al. (2014). The stimuli varied on two dimensions: dot density and size of rectangle. Figure from Smith et al. (2014).

This image has been removed by the author of this thesis for copyright reasons.

Figure 4.2. Illustrative of the category structures used in Experiment 4 and Smith et al. (2014); (a) the unidimensional RB structure; (b) the II structure. Figure from Smith et al. (2014).

set of 600 stimuli were generated individually for each participant. This meant that while participants in each condition experienced the same category decision boundaries, the precise visual stimuli presented varied between each individual (see Table 4.1 for the values used to construct each category).

Procedure

Participants were randomly allocated to one of the four conditions and completed all 600 trials of the experiment. Participants were informed that on each trial they would see a novel stimulus which could be categorized as either A or B, and they were asked to learn into which of two categories a series of stimuli belonged through trial and error and feedback presentation.

Table 4.1.

Distributional Characteristics of the Unidimensional RB and II Categories. Each Dimension (X and Y) Varied On 100 Levels; The Numbers in the Table Refer to The Bivariate Distribution of Each Category Out of 100. Table from Smith et al. (2014).

<i>Task</i>	<i>Category</i>	<i>MeanX</i>	<i>MeanY</i>	<i>VarX</i>	<i>VarY</i>	<i>CovarXY</i>
<i>RB</i>	A	35.86	50.00	16.33	355.55	0
	B	64.14	50.00	16.33	355.55	0
<i>II</i>	A	40.00	60.00	185.94	185.94	169.61
	B	60.00	40.00	185.94	185.94	169.61

RB = rule-based; II = Information-integration;
 X = rectangle size; Y = density of green dots;
 Var = variance; Covar = covariance

Each trial began with a white fixation cross in the middle of a black screen lasting 500ms which was replaced by a stimulus. This remained on the screen until the participant pressed either the ‘z’ key on the keyboard if they thought the stimulus belonged in category A, or the ‘m’ key to indicate that it was in category B. As in Smith et al.’s study, the feedback was audial. In the *immediate feedback* conditions participants heard a feedback tone after every trial. If the participant correctly categorized the stimulus they heard a high-pitched tone. There was then a 500ms ITI before the next trial began. However, if the participant incorrectly categorized the stimulus, then the participant heard a low pitched tone and incurred a 4 second ITI. Participants received a self-paced break after every 6 trials.

In the *deferred feedback* condition, participants responded to a block of six stimuli with an ITI of 250ms. Once these six trials were completed feedback was collectively presented. Participants received feedback for correct trials first, hearing a high-pitched tone for every correct categorization response made, each separated by 500ms. A low-pitched tone was then played for every incorrect response made, with

each incorrect tone followed by a 4 second ITI. The tones presented were the same as in the immediate feedback condition. The next block then followed after a self-paced break.

Modelling analysis

Using the same method as detailed in Chapter 2, a modelling analysis based on the General Recognition Theory (GRT; Ashby & Gott, 1988) was performed using the *grt* package in the R environment (Matsuki, 2014). This was used as an indication of the decision boundary implemented by participants. However, the accuracy of these modelling results is debated (see Donkin et al., 2014; Edmunds, et al., 2015, for a discussion) so these analyses should be interpreted cautiously.

For each participant, the GRT analysis determines which learning strategy from a set of four pre-defined models provides the best fitting account of that participant's responses (indicated by the Akaike's information criterion; Akaike, 1974). As in Chapter's 2 and 3, the first model was a *unidimensional* decision boundary, where participants responded based on a decision criterion along either the rectangle size or dot density dimension. The *conjunctive* model assumed that the participants combined judgments about the two stimulus dimensions to make a response. The *General Linear Classifier (GLC)* model assumed that participants generated a decision boundary that could be described by a straight line that could vary on gradient and intercept. The final model assumed that participants responded *randomly*.

Results

Mean accuracy in all blocks of the category learning task can be seen in Figure 4.3.a. The analyses were conducted on the proportion of correct categorization trials in the last 100 trials of the study (to be consistent with Smith et al., 2014). The mean accuracy

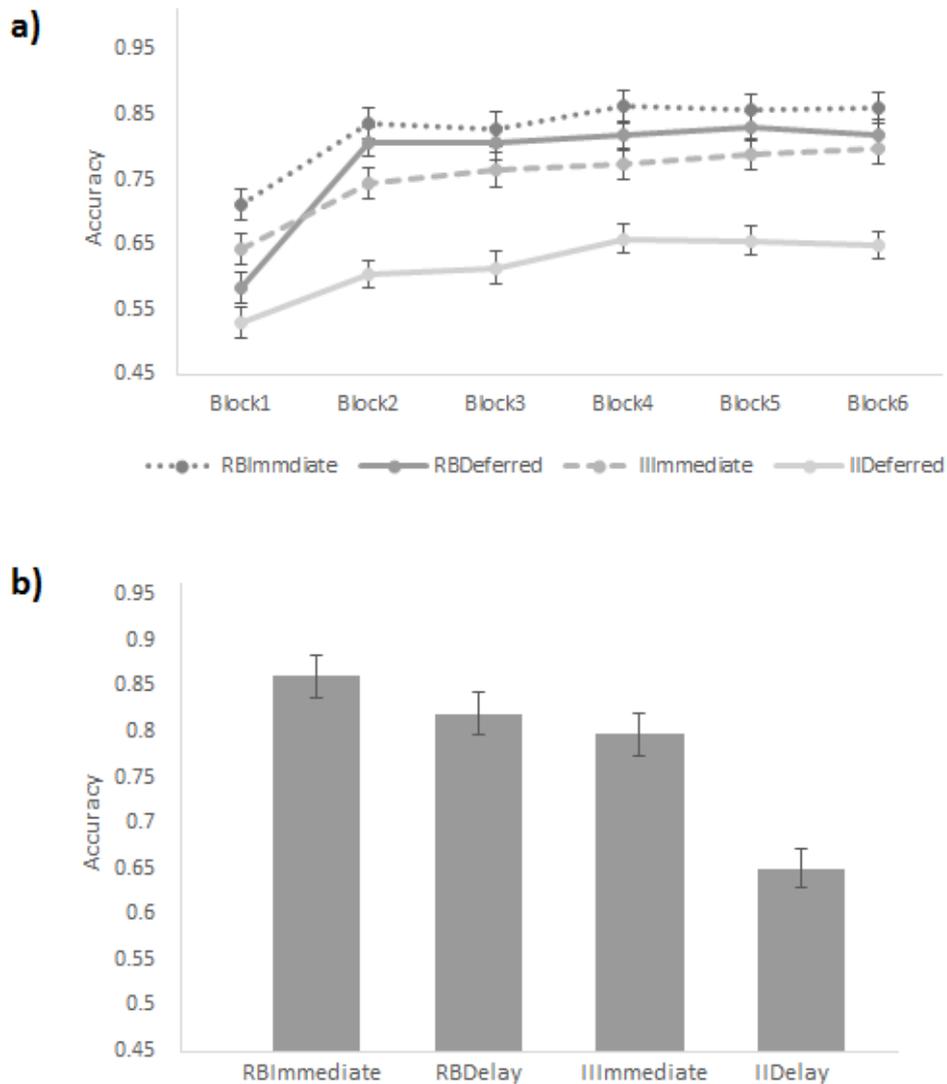


Figure 4.3. (a) Mean accuracy of responding across all blocks in Experiment 4; (b) mean accuracy of responding in the last 100 trials of Experiment 4. Error bars represent standard error.

for each condition in the last 100 trials can be seen in Figure 4.3.b.

A 2x2 between-subject ANOVA was completed with category structure (RB/II) and feedback type (immediate/deferred) entered as the fixed factors. The dependent variable was the proportion of correct answers for each participant on the last 100 trials of the study. There was a main effect of category structure on accuracy ($F(1,81)=26.93$, $p<.001$, $\eta^2_p=.25$) where participants in the unidimensional RB condition were more accurate than II participants. There was also a main effect of feedback type

($F(1,81)=17.35$, $p<.001$, $\eta^2_p=.18$) as participants in deferred feedback conditions performed less accurately than participants in immediate feedback conditions. Furthermore, there was also a significant interaction between category structure and feedback type ($F(1,81)=5.58$, $p=.021$, $\eta^2_p=.06$). Paired comparisons, assessing the nature of this interaction, showed that performance in the II deferred feedback condition was significantly worse than those in the II immediate feedback condition ($t(42)=4.23$, $p<.001$, $d=1.31^{11}$). However, performance in the RB immediate condition was not significantly different from that of the RB deferred condition ($t(39)=1.45$, $p=.156$, $d=0.45$, $BF=0.38^{12}$). A test of simple main effects shows that II participants are performing marginally worse than RB participants in both the immediate conditions ($p=.055$, $\eta^2_p=.05$), and in the deferred conditions ($p<.001$, $\eta^2_p=.27$).

Modelling analysis

The modelling analysis was completed on responses from the last 100 trials for each participant individually. Table 4.2 shows which learning strategy participants used in each condition of the study. As was found in Smith et al. (2014) the majority of participants in the II immediate condition were best fit by an II strategy (GLC). However, in the II deferred condition the majority of participants were classified as using a UD strategy. Unlike Smith et al. (2014), though, a large majority of RB participants in the immediate and deferred feedback conditions also used an II learning strategy. The low number of RB participants using unidimensional learning strategies is perhaps surprising. This finding could reflect the potential instability of the modelling

¹¹ Again, the calculation of Cohen's d used was $d = M_1 - M_2 / SD_{pooled}$. In Chapter 4, the comparisons are made where $M_1 = \text{Immediate}$ $M_2 = \text{deferred}$.

¹² An estimate of the mean expected difference under the experimental hypothesis was calculated from Smith et al.'s (2014) study. Following Dienes (2011), the expected difference was modelled as a two-tailed normal distribution with a standard deviation equal to half the mean. This procedure was used throughout Chapter 4.

Table 4.2.

Proportion of Participants in Experiment 4 Using the Conjunctive (CJ), Information-Integration (II), Unidimensional (UD), or Random (RND) Strategies in the RB and II Conditions

	CJ	II	UD	RND
RBIm	0.00	0.50	0.50	0.00
RBDe	0.00	0.57	0.39	0.04
IIIm	0.00	0.65	0.30	0.05
IIIE	0.05	0.08	0.74	0.13

analysis which only included the last 100 trials of the study. Indeed, when including all blocks of the study in the modelling analysis evidence for a greater use of unidimensional learning strategies was found (see Appendix 4.1, Table S.4.1).

Discussion

Experiment 4 replicated the dissociation in unidimensional RB and II performance reported by Smith et al. (2014). There was a significant interaction between category structure and feedback type, with II performance significantly impaired by deferring feedback whereas unidimensional RB accuracy was unaffected by feedback type. However, participants in the unidimensional conditions also performed marginally better than II participants in the immediate conditions. This fits with the conclusions made in Chapters 2 – 3, and in the work of others such as Edmunds et al. (2015), that unidimensional categories are, by nature, easier to learn than multidimensional category structures.

Experiment 5 aims to investigate whether the behavioural dissociation between unidimensional RB and II performance under deferred feedback can be generalised to the conjunctive RB and II stimuli implemented in Chapters 2 and 3.

Experiment 5

The unidimensional RB structure used by Smith et al. (2014) can be considered as less complex than the II condition used as only one stimulus dimension is relevant for learning as opposed to the two necessary for learning in the II structure, and therefore is likely to require less cognitive resources to learn optimally (e.g., Edmunds et al., 2015; Nosofsky & Kruschke, 2002; Milton, et al., 2008; Wills, et al., 2015). This difference in difficulty between the II and unidimensional RB conditions in Smith et al. (2014) and Experiment 4 could mean that the II condition is more susceptible to the effects of suboptimal feedback. Experiment 5 used a conjunctive RB category structure matched to the II structure in difficulty and the number of relevant dimensions to assess whether the difference in relevant dimensions for learning between conditions accounts for the dissociation found in Smith et al. and Experiment 4. Based on Edmunds et al. (2015) and the results of Chapters 2 and 3, it is predicted that, contrary to the predictions of COVIS and Smith et al., conjunctive RB and II category structures will not be differentially affected by deferring feedback.

Method

Participants and design

86 University of Exeter students completed the experiment (16 males, mean age = 19.15 years, SD = 2.77). However, the same exclusion criteria to Smith et al. (2014) was applied again, and 4 participants were excluded for performing significantly worse in the last 100 trials of the study compared with the first 100 trials. Participants received course credit or £5 remuneration for participation in the study.

The study employed a 2 (category structure) X 2 (feedback type) design creating four conditions: RB immediate (n=21); RB deferred (n=20); II immediate (n=20); and II deferred (n=21).

Stimuli

Experiment 5 used the full sets of 600 conjunctive RB and II stimuli first employed by Filoteo et al. (2010; see Figure 4.4). These were black line stimuli on white backgrounds that varied on line length and line orientation – both dimensions were relevant for learning in the RB and II conditions. The conjunctive rule was “short, upright lines belong in category A, and the rest in category B” and the II condition was separated by a diagonal decision boundary. As in Filoteo et al. there was a 5% overlap between the categories so that the maximum accuracy that could be achieved was 95%.

Procedure and analysis

The procedure used and the analyses performed were the same as in Experiment 4.

This image has been removed by the author of this thesis for copyright reasons.

Figure 4.4. The category structures used in Filoteo et al. (2010) and the present study; (a) the conjunctive rule-based condition; (b) the information-integration condition. Solid lines indicate the decision boundary separating category A and category B. Figure taken from Filoteo et al., (2010).

Results

Mean accuracy in all blocks of the category learning task can be seen in Figure 4.5.a.

Again, the analyses were completed on the proportion of correct categorization trials for each participant in the last 100 trials of the study (see Figure 4.5.b).

A two-way ANOVA was completed in the same way as in Experiment 4. There was no significant effect of category structure on accuracy ($F(1,78)=.52$, $p=.472$, $BF=0.05$, $\eta^2_p=.007$). There was, however, a main effect of feedback type ($F(1,78)= 46.07$,

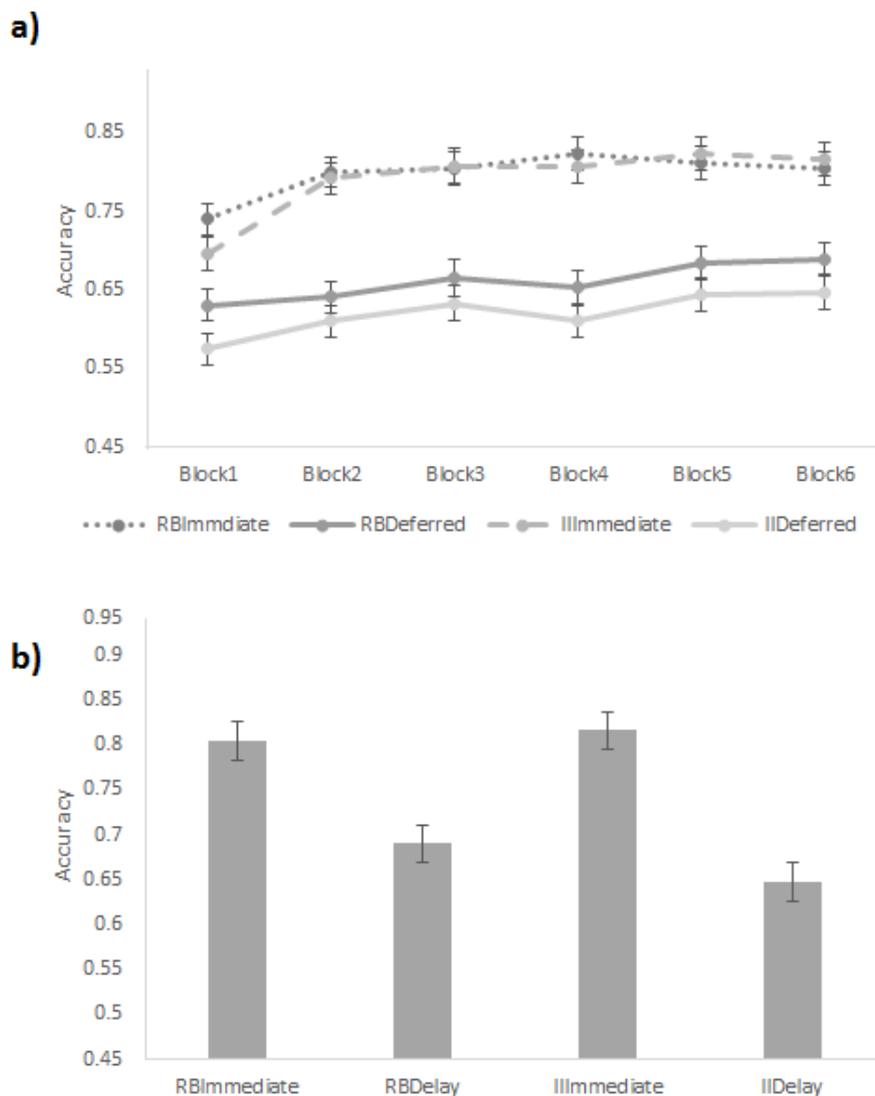


Figure 4.5. (a) Mean accuracy of responding across blocks in Experiment 5; (b) the mean accuracy of responding in the last 100 trials of Experiment 5. Error bars represent standard error.

$p < .001$, $\eta^2_p = .37$) as participants in deferred feedback conditions performed less accurately than participants in immediate feedback conditions. Yet, the interaction between category structure and feedback type found in Smith et al. (2014) and Experiment 4 was not significant ($F(1,78) = 1.71$, $p = .195$, $BF = .69$, $\eta^2_p = .02$).

As in Experiment 4, participants receiving deferred feedback in the II condition performed significantly worse than those in the II immediate condition ($t(39) = 6.42$, $p < .001$, $d = 2$). However, unlike Experiment 4, participants in the deferred RB condition also performed less accurately than participants in the immediate RB condition ($t(39) = 3.53$, $p = .001$, $d = 1.08$).

Modelling analysis

The modelling analysis was again performed on the last 100 trials for each participant individually. Table 4.3 shows the proportion of participants using the different learning strategies during each condition. As was found in Smith et al. (2014), the most used learning strategy by participants performing the RB and II category structures with immediate feedback was optimal for the condition. However, in the deferred conditions, the majority of both II and RB participants used a suboptimal unidimensional RB strategy to learn the categories. The modelling analysis across all blocks can be found in Appendix 4.1, Table S.4.2.

Discussion

Unlike Experiment 4 and Smith et al. (2014), Experiment 5 did not find a significant interaction between category structure and feedback type. Instead, both conjunctive RB participants and II participants were significantly impaired when receiving deferred feedback compared with their immediate feedback counterparts. This contradicts the prediction of Smith et al. (2014) that RB learning should ‘flourish’ (Pp. 450) under

Table 4.3.

Proportion of Participants in Experiment 5 Using the Conjunctive (CJ), Information-Integration (II), Unidimensional (UD), or Random (RND) Strategies in the RB and II Conditions

	CJ	II	UD	RND
RBIm	0.48	0.42	0.05	0.05
RBDe	0.10	0.00	0.80	0.10
IIIm	0.00	0.86	0.14	0.00
IIDE	0.08	0.17	0.63	0.12

deferred feedback conditions. The finding that participants in the conjunctive RB condition were significantly impaired with deferred feedback compared to those in the immediate condition appears fairly intuitive. For example, as deferring feedback requires information about six stimuli to be held in memory before corrective feedback is presented, this condition necessitates a greater engagement of WM processes compared with immediate feedback conditions to successfully store information. This information must also be engaged for longer than in immediate conditions, again adding a further WM strain on deferred participants. Furthermore, as in Smith et al., in the deferred feedback conditions Experiment 5 did not present corrective feedback corresponding to the order of responding, as all ‘correct’ feedback was presented before ‘incorrect’ feedback. Thus deferred feedback is also non-specific as it is not associated with a single response. Therefore, even if participants receiving deferred feedback successfully store all information about the stimuli seen and responses made in WM, the participants will be disadvantaged compared to participants receiving immediate feedback as they do not know specifically which trials they answered incorrectly, but rather receive a general indication of their performance level. Intuitively knowing “I

responded correctly to stimulus X” appears more informative than knowing “I performed at 66% in that block of stimuli”.

Similarly, the present results are in line with recent work by Edmunds et al. (2015) and Nosofsky and Kruschke (2002) who suggest that the simplicity of unidimensional categories in comparison to II categories calls into question dissociations found when implementing these structures. The greater ease of learning the unidimensional structure compared to the II structure as well as the conjunctive RB condition of Experiment 5, would make it less susceptible to behavioural manipulations that tax the learning process, and therefore make it less likely to be impaired by suboptimal feedback. Here, however, because both the conjunctive RB and II conditions are relatively difficult, they are both hindered by deferring feedback. These speculations are supported by the modelling analysis which demonstrates that RB and II participants are more likely to rely on a suboptimal yet less complex unidimensional decision boundary under deferred feedback than immediate feedback.

Nevertheless, just as it appeared important to replicate the initial finding of Smith et al. (2014) in Experiment 4, it seems necessary to establish the robustness of this finding that conjunctive RB learning is impaired under deferred feedback, before drawing firm conclusions on this data. This was the aim of Experiment 6.

Experiment 6

Experiment 5 found, contrary to the results of Experiment 4 and Smith et al. (2014), that the conjunctive RB condition was impaired by deferring feedback during category learning. This result opposes the predictions of COVIS theorists that suggest RB learning should ‘flourish’ (Pp. 450) with deferred feedback (Smith et al., 2014).

Experiment 6 aimed to further investigate and confirm the finding of Experiment 5 that conjunctive RB learning is impaired by deferring feedback compared to immediate feedback conditions. Participants performed the conjunctive RB task in either the deferred or immediate feedback condition. The II condition was not tested again in this study, as II participants are consistently impaired by deferring feedback in Smith et al. and Experiments 4 and 5. This allowed the number of participants used in each condition in Experiment 6 to double from that of the previous two studies, to increase the reliability and power of the analyses performed. In line with Experiment 5 it was predicted that there would be a significant difference in categorization accuracy between the two conditions with those receiving deferred feedback performing significantly worse than those receiving immediate feedback. This would offer strong evidence that conjunctive RB learning is hindered by deferring feedback and support the conclusion that the dissociation between RB and II learning in Smith et al. (2014) and Experiment 4 was driven by the differing number of relevant dimensions between the category structures employed.

Method

86 University of Exeter students completed the experiment (23 males, mean age = 20.64 years, $SD = 4.04$). However, 1 participant was excluded for performing significantly worse in the last 100 trials of the study compared with the first 100 trials, in line with the exclusion criteria used by Smith et al. (2014). This left 41 participants in the immediate feedback condition and 44 participants in the deferred condition. Participants received course credit or £5 remuneration for participation in the study.

A between-subjects design was implemented with two conditions: conjunctive RB learning with immediate feedback, and conjunctive RB learning with deferred

feedback. The stimuli used were the conjunctive RB stimuli of Experiment 5 (see Figure 4.4) and the procedure was the same as Experiments 4 and 5.

Results

Mean accuracy of responding across all blocks is shown in Figure 4.6. The analysis was completed on the proportion of correct categorization responses for each participant in the last 100 trials.

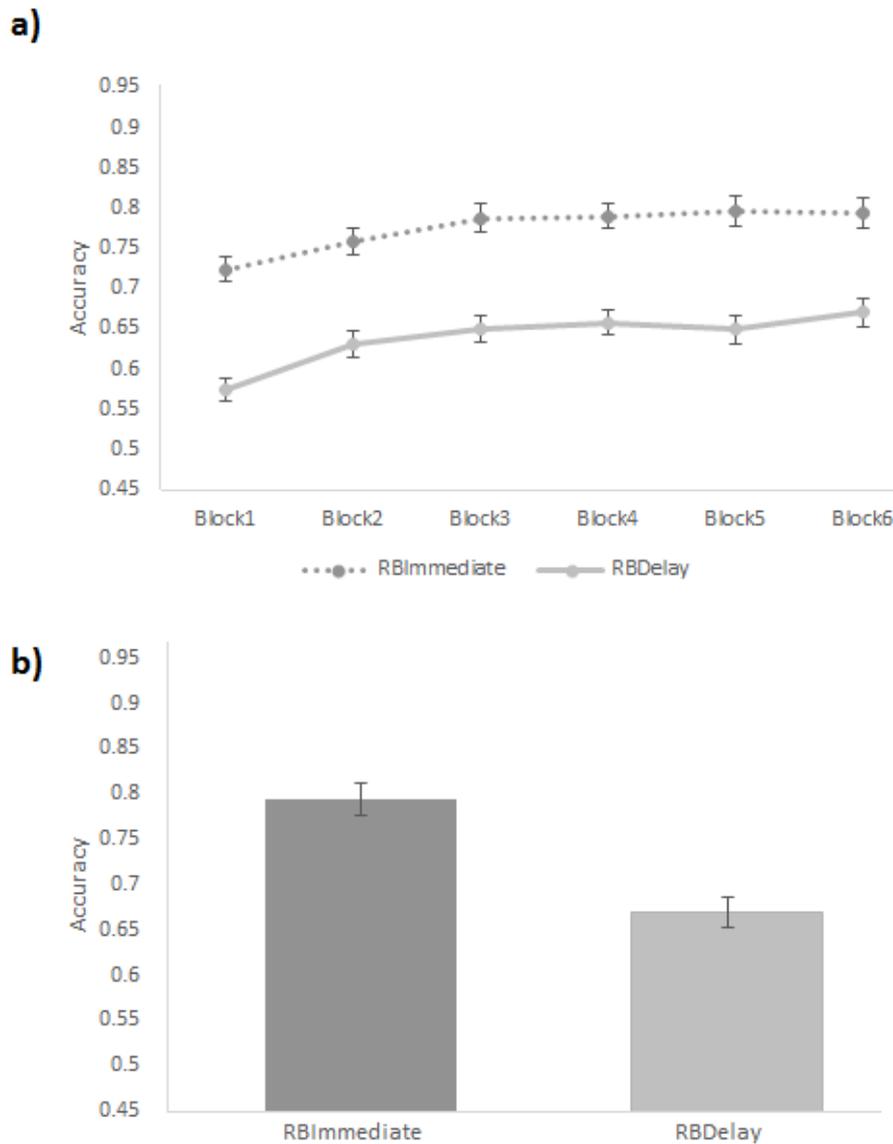


Figure 4.6. Proportion of correct responses across blocks in Experiment 6. Error bars represent standard error.

A one-way ANOVA was completed with feedback type (immediate/deferred) as the fixed factor. The dependent variable was the proportion of correct responses for each participant on the last 100 trials of the study. There was a main effect of feedback type on accuracy ($F(1, 83)=24.91, p<.001, \eta^2_p=.23$) where participants in the deferred condition were less accurate than those in the immediate feedback condition.

Modelling analysis

The modelling analysis was again performed on the last 100 trials for each participant individually. Table 4.4 shows the proportion of participants using each learning strategy in both conditions of the study. As found in Experiment 5, the majority of participants in the immediate condition relied on the optimal conjunctive learning strategy. However, in the deferred condition the majority of participants implemented a suboptimal unidimensional decision boundary to learn the categories. The modelling analysis across all runs can be found in Appendix 4.1, Table S.4.3.

Discussion

The results of Experiment 6 support the findings of Experiment 5. Learning in the conjunctive RB condition was significantly impaired by deferring feedback to the end of a six trial block. Furthermore, with deferred feedback conjunctive RB participants used a suboptimal unidimensional decision boundary to respond more than

Table 4.4.

Proportion of Participants in Experiment 6 Using the Conjunctive (CJ), Information-Integration (II), Unidimensional (UD), or Random (RND) Strategies in the RB Condition.

	CJ	II	UD	RND
RBI _m	0.56	0.29	0.12	0.03
RBD _e	0.11	0.09	0.66	0.14

those who received immediate feedback. This contradicts the predictions of COVIS, and importantly offers further support for the view that the dissociation found in Smith et al. and Experiment 4 is accounted for by the use of a unidimensional category structure to compare with II performance.

General Discussion

Smith et al. (2014) found that II learning is impaired when deferring feedback to the end of a six trial block compared with II participants who receive immediate feedback, an effect not present in RB participants. Smith et al. (2014) claim that this provides some of the strongest evidence for COVIS. This behavioural dissociation was replicated in Experiment 4, which used the same category structures and stimuli as Smith et al. to find a significant interaction between category structure and feedback presentation, as participants in the II deferred condition performed less accurately than participants in the II immediate condition, whilst there was no significant effect of feedback type on RB learning.

Smith et al. (2014) argue that this pattern of results is beyond the scope of single system theories. However, Experiment 5 found the dissociation reported by Smith et al. did not emerge when a conjunctive RB category structure rather than a unidimensional structure was employed as a comparison to the II condition. Instead, there was no interaction between category structure and feedback condition, with both conjunctive RB participants and II participants significantly hindered when receiving deferred feedback compared with participants receiving immediate feedback. The results of Experiment 5 are further supported by Experiment 6 which doubled the number of participants per condition and found that deferring feedback impaired conjunctive RB category learning compared to providing immediate feedback. Complementing the behavioural results, GRT modelling indicated that deferred feedback encouraged both II

and conjunctive RB participants to use a suboptimal unidimensional strategy whilst a greater proportion of participants in the immediate feedback conditions used the appropriate optimal strategy. While COVIS predicts that deferred feedback should impair learning of an II structure, the fact that a similar result emerges for the conjunctive RB structure is contrary to the clear predictions made by COVIS theorists.

An important question raised by Chapter 4 is what causes the discrepancies found between the findings of Smith et al. (2014) and Experiment 4, and Experiments 5 and 6? The most likely explanation for this is the switch from the use of a unidimensional structure in Experiment 4 to a conjunctive structure in the RB conditions in Experiments 5 and 6. As discussed, the unidimensional structure used in Smith et al. (2014) has only one dimension relevant for category learning compared with the two dimensions relevant in the II condition. It is likely that the more complex II category structure places greater demand on cognitive processes such as WM than participants in unidimensional conditions (e.g., Milton et al., 2008; Wills et al., 2015). This inference is supported by Experiment 4 which finds that, overall, unidimensional RB participants are significantly more accurate than II participants. If multidimensional category structures engage more WM and cognitive resources, interrupting feedback may be more detrimental to an already taxed learning system when compared to easier to learn unidimensional categories. Thus, when comparing two multidimensional conditions (as in Experiments 5 and 6) it perhaps is unsurprising from this perspective that the dissociation reported by Smith et al. disappears. This echoes the research presented in Chapters 2 and 3 that concludes that neural and behavioural (feedback delay) dissociations between RB and II performance presented in the literature (e.g., Nomura et al., 2007; Maddox et al., 2003) are likely attributable to the use of RB and II structures not matched on the number of relevant category dimensions.

While there is a greater demand on WM and executive functioning processes during deferred feedback trials in the conjunctive RB condition compared to the more easily processed unidimensional RB condition, the non-specific feedback in the deferred condition could also be contributing to the dissociation in multidimensional and unidimensional performance. Experiment 4 suggests a difficulty confound between unidimensional and multidimensional learning conditions; the more difficult multidimensional RB conjunctive and II conditions therefore may be less resilient to non-specific feedback than unidimensional conditions, which have been found to be easy to acquire even in the absence of feedback (Ashby, Queller, & Berretty, 1999). This could account for the switch to suboptimal unidimensional strategies in the modelling analyses of the II and conjunctive RB conditions in the present chapter, and explain why non-specific feedback is not disrupting learning in the unidimensional RB category structure.

Smith et al. (2014), however, argue that a unidimensional RB structure is an appropriate comparison for the II structure. Indeed, Smith et al. (2014) described unidimensional RB and II tasks as “elegant mutual controls” (Pp. 449) as they are matched on variables such as within-category exemplar similarity, and note that II category structures are simply rotations of unidimensional RB structures (e.g. Figure 4.2) and therefore there should be no a priori difference in error rates between the two conditions. Smith et al. further claim that all of the behavioural dissociations found in the COVIS research field cannot be explained simply by a difficulty confound caused by using unidimensional structures. Although humans can learn unidimensional conditions more quickly than II conditions, Smith et al. (2011), found that pigeons learn unidimensional and II structures equally. This led the authors to suggest that pigeons do not possess an explicit system of category learning, so acquire both categories using the same non-analytic learning system evolved millennia ago in vertebrae. Therefore, Smith

et al. (2014) suggest that there is no a priori difference between unidimensional RB and II conditions, as pigeons, which are less cognitively advanced and do not pose an explicit system optimal for RB learning, learn RB and II conditions equally. Smith et al. (2014) conclude that “If humans’ learning-rate difference [in RB and II conditions] arose because the II task is inherently difficult, then a less cognitively sophisticated species (pigeons) should be more challenged on the II task (relative to the RB task) than humans are” (Smith et al., 2014, Pp. 49). However, Edmunds et al. (2015) consider that there is no evidence that pigeons recognise the two stimulus dimensions as psychologically separable. If pigeons do not recognise the categories as either unidimensional or two-dimensional, there is no evidence that the categories implemented differ in dimensionality between conditions, and therefore the work does not provide transferrable insight as to whether conditions unmatched on dimensionality critically affect performance in humans. It should also be noted that the conclusions drawn from Smith et al. (2011) are based on a null result suggesting no significant difference in RB and II performance, and no Bayesian analysis was performed which could have provided insight into whether there was genuine evidence for the null hypothesis. Therefore, it cannot be inferred whether the data collected was insensitive to true differences between RB and II performance in pigeons, or whether the conditions were indeed learnt at equal rates.

Experiments 5 and 6 present a pattern of results that is consistent with previous work that suggests dissociations in RB and II performance can be accounted for by the comparison of II categories to unidimensional RB categories (e.g., Edmunds et al., 2015; Nosofsky & Kruschke, 2002; and Chapters 2 and 3). However, it cannot be definitively ruled out that the difference in stimuli between Experiment 4 and Experiments 5 and 6 are driving the discrepancy in results. Smith et al. (2014) and Experiment 4 of this thesis both present participants with two dimensional green dot

distribution stimuli, while Experiments 5 and 6 employ two dimensional line stimuli. Although it is not clear from a COVIS perspective why such a change in stimuli would result in the contrasting findings, this possibility cannot be eliminated with the current data. Nevertheless, it does appear unlikely that this is the critical factor given that a myriad of studies now report that dissociations which are present when comparing unidimensional RB structures with II structures, disappear when the unidimensional structure is replaced by a conjunctive one (e.g., Edmunds et al., 2015; Chapter 2). To confirm this, though, future work should address whether the abolished dissociation in Experiments 5 and 6 is attributable to the switch in visual stimuli from dots to lines by replicating the results of Experiments 5 and 6 when maintaining the conjunctive RB and II structures, but using the green dot stimuli employed by Smith et al. (2014).

To conclude, Smith et al. (2014) claimed that their dissociation demonstrating that II learning was selectively impaired by deferred feedback presented some of the strongest evidence to date of separable explicit and implicit systems of category learning, and poses a real challenge to single system accounts. However, when accounting for complexity and relevant stimulus dimensions in RB and II categories, Experiments 5 and 6 find no evidence of this behavioural dissociation. While this is contrary to Smith et al.'s (2014) account, it is entirely consistent with single process accounts of category learning. The present chapter extends our understanding of the impact deferring feedback has on category learning conditions, and poses a challenge for the current predictions of COVIS.

Appendix 1 Chapter 4: Modelling analyses containing all blocks of the experiments.

Table S.4.1.

Proportion of Participants in All Blocks of Experiment 4 using the Conjunctive (CJ), Information-Integration (II), Unidimensional (UD), or Random (RND) Strategies in the RB and II Conditions.

	CJ	II	UD	RND
RBIIm	0.00	0.30	0.70	0.00
RBDDe	0.05	0.52	0.43	0.00
IIIIm	0.00	0.63	0.37	0.00
IIDDe	0.17	0.13	0.65	0.05

Table S.4.2.

Proportion of Participants in All Blocks of Experiment 5 using the Conjunctive (CJ), Information-Integration (II), Unidimensional (UD), or Random (RND) Strategies in the RB and II Conditions.

	CJ	II	UD	RND
RBIIm	0.67	0.23	0.05	0.05
RBDDe	0.25	0.15	0.55	0.05
IIIIm	0.14	0.86	0.00	0.00
IIDDe	0.26	0.18	0.52	0.04

Table S.4.3.

Proportion of Participants in All Blocks of Experiment 6 Using the Conjunctive (CJ), Information-Integration (II), unidimensional (UD), or Random (RND) Strategies in the RB Condition.

	CJ	II	UD	RND
RBIIm	0.70	0.28	0.00	0.02
RBDDe	0.25	0.11	0.59	0.05

Chapter 5. Investigating the effect of anodal tDCS on rule-based and information-integration category learning

In the last few years there has been a surge of interest in transcranial Direct Current Stimulation (tDCS) within the field of cognitive psychology. The application of topical stimulation modulates neural activity under a given electrode site by passing a small current through the scalp and skull (Coffman, Clark, & Parasuraman, 2013). TDCS current can be applied in one of two directions: anodal or cathodal. While anodal stimulation is suggested to increase neural excitability by facilitating neural depolarization, cathodal stimulation is proposed to hyperpolarize neurons, which leads to neural inhibition (Bindman, Lippold, & Redfearn, 1962). In the cognitive research field it has been found that accuracy in WM tasks can be improved by anodal tDCS (Fregni et al., 2005) and that externally driven attention is increased with this technique (Coffman, Trumbo, & Clark, 2012). Performance on tasks requiring spatial memory (Flöel, et al., 2012), object detection (Clark et al., 2012), as well as word list recollection (e.g., Marshall, Mölle, Hallschmid, & Born, 2004) has also been found to be improved by anodal tDCS (for a review of the effects of tDCS on cognitive tasks see Coffman et al., 2013).

TDCS has also been reported to modulate participants' ability to categorize familiar objects. Lupyan, Mirman, Hamilton, and Thompson-Schill (2012) found that 1.5mA of anodal stimulation to the left inferior frontal cortex enhanced the likelihood that participants selected more marginal category members in a categorization task. For example, if asked to select stimuli that are 'soft' from an array of objects, participants undergoing tDCS often select a strawberry as something soft over a pillow, which is arguably more associated with the 'soft' category. The authors speculated that the anodal stimulation increased participants' sensitivity to the different properties of a stimulus, especially those relevant for task performance. Lupyan et al. also asked

participants to separate visual items based on a single stimulus dimension (e.g., pick the blue shapes). Participants who had received cathodal stimulation to the inferior frontal cortex were less accurate in sorting stimuli by a single dimension compared to sham controls whereas anodal stimulation had no significant effect on this measure. These two findings serve to highlight how the polarity of tDCS used can affect behavioural results in brain stimulation studies. Some theorise that anodal stimulation facilitates performance, as it leads to increased neural excitability because it has a positive polarization (Nitsche & Paulus, 2000). However, cathodal stimulation is proposed to suppress neural excitability because it has a negative polarization (Nitsche & Paulus, 2000; but see Ambrus et al., 2011).

Although some tDCS studies such as Lupyan et al. (2012) have examined categorization, there are few that focus on category learning processes. This is a potentially profitable avenue of research as neural predictions about the brain regions involved in category learning can be further examined using a method of brain stimulation that has few side effects, that is inexpensive, and has minimal set up time. For example, if one theory implicates a certain brain region in category learning and tDCS to this area is found to modulate performance, then this could be used as evidence to support the theory. If tDCS to a particular brain region is found to increase learning, then it can be applied in educational settings where category learning is necessary (for example in student radiologists; Filoteo et al., 2010). One such example of tDCS work based on neural predictions is Kincses, Antal, Nitsche, and Paulus (2003) who noted that previous MRI research implicates the neostriate in implicit category learning (e.g., Poldrack, Prabhakaran, Seger, & Gabrieli, 1999). Kincses et al. predicted that since the prefrontal cortex and neostriatum are neurally connected, category learning attributable to neostriate areas could be modulated by tDCS to the prefrontal cortex. Kincses et al. used a weather prediction task that they proposed tapped into an implicit learning

system (but c.f. Lagnado, Newell, Kahan, & Shanks, 2006; Newell, Dunn, & Kalish, 2011, for a criticism of this assumption) supported by the neostriatum. This task required participants to learn to associate visual stimuli (e.g., varying shape montages) with a 'sunny' or 'rainy' outcome. Each shape presented had a different probability of being associated with each outcome. It was found that 1.0mA of anodal stimulation to the left prefrontal cortex (Fp3) for 10 minutes increased participants' accuracy on this task when compared with participants who had received cathodal or no stimulation. The authors concluded that learning in the weather prediction task can be modulated by anodal tDCS to the prefrontal cortex. However, they note that they cannot conclude whether the stimulation to the prefrontal cortex alone is responsible for the increased accuracy, or whether an increase in basal ganglia activity due to its strong connections with the prefrontal cortex is responsible for the effect found.

Vercammen et al. (2011) also manipulated activation in the prefrontal cortex using anodal tDCS stimulation. The researchers first presented schizophrenic patients with a weather prediction task to be performed without brain stimulation, to establish the patients' baseline performance when learning cue-outcome associations. In a within-subjects study participants then performed two more weather prediction tasks over two days, one with anodal tDCS and one without tDCS. Patients who were able to learn the weather prediction task at the baseline test demonstrated improved performance with anodal tDCS compared to the second weather prediction task performed without tDCS. However, if a participant performed poorly on the baseline weather prediction task then tDCS had no impact on their category learning. Participants' accuracy on the baseline task was also positively correlated with performance on the Wechsler Adult Intelligence Scale III (WAIS-III) which was used as a measure of WM capacity. Vercammen et al. (2011) suggest that participants with preserved WM capacity have more 'prefrontal neural reserves' than participants with reduced WM, and therefore, as these reserves are

stimulated, participants with intact WM functioning benefit more from anodal tDCS than those without such reserves. Although there is debate as to whether implicit learning is truly represented by the weather prediction task used in Kincses et al. (2003) and Vercammen et al. (2011) (e.g. Newell et al., 2011), it appears clear that categorization processes, whether explicit or implicit, can be modulated using tDCS¹³.

Prototype distortion category learning tasks are also reported to be affected by tDCS; this work was discussed in Chapter 1, and will be considered in more detail in Chapter 6. Briefly though, Ambrus et al. (2011) presented participants with a prototype distortion task whereby they were shown exemplars of category (A) which were all distortions of a single category prototype. At test participants performed an (A, not A) task where they categorized novel category member and random non-category member stimuli. Anodal tDCS to the left DLPFC reduced accuracy compared to cathodal and sham participants who did not differ significantly in overall performance. A prototype effect (where participants responded more accurately to the category prototype than to exemplars after training) present in cathodal and sham participants was abolished in anodal participants. Similarly, McLaren et al. (2016, Experiment 2) found that anodal tDCS to the left DLPFC resulted in participants performing numerically worse than cathodal and sham participants when learning two prototype based category structures. However, in contrast to Ambrus et al., anodal participants demonstrated a prototype effect that was not present in sham or cathodal participants.

The effect of tDCS on category learning strategies has also been investigated. Perry and Lupyan (2014) explored if cathodal tDCS affected whether participants relied upon explicit verbal rules to learn categories, or if participants used an II decision boundary. Perry and Lupyan presented participants with categories that could be learnt

¹³ Prototype distortion category learning tasks are also reported to be affected by tDCS; this work was discussed in Chapter 1, and will be considered in more detail in Chapter 6.

This image has been removed by the author of this thesis for copyright reasons.

Figure 5.1. The distribution of stimuli and potential decision boundaries of Perry and Lupyan (2014). Training stimuli are taken from clusters A and B that can be successfully categorized when either applying a diagonal II learning strategy or a unidimensional RB strategy (the dashed lines). Transfer test stimuli are from clusters C, D, E, and F. Figure taken from Perry and Lupyan (2014).

by using either an RB or II decision boundary (see Figure 5.1) somewhat similar to those previously used in COVIS research (e.g., Smith et al., 2014; Ashby et al., 2002). Although the category decision boundary could be learnt successfully by either a unidimensional RB or II strategy, the category structure is somewhat different to that normally used in COVIS related work with the consequence that the training stimuli (clusters A and B in Figure 5.1) can be partitioned successfully using either an RB or an II approach. During 20 minutes (the duration of the study) of 1.75mA cathodal stimulation to Wernicke's area, participants were trained on the stimuli in clusters A and B in Figure 5.1. The effect of cathodal tDCS on the category learning strategy used in training was assessed by transfer stimuli from clusters C, D, E, and F (Figure 5.1). From participants' performance on the transfer trials, the decision boundaries they acquired at training were inferred. Participants' category assignments on transfer trials reflected the decision boundary they generated at training, so if a participant had learnt

using the vertical, unidimensional RB decision boundary they would categorize stimuli from clusters 'C+D' and 'E + F' together, as these clusters in this formation were separated by the same vertical unidimensional line in stimuli space. However, if they had used an II decision boundary they would categorize stimuli from clusters 'E + C' and 'F + D' together, as this cluster formation was separated by the same II style diagonal decision boundary as the training stimuli. Cathodal stimulation to Wernicke's area increased participants' reliance on II learning strategies. This was assumed to represent cathodal stimulation inhibiting the Wernicke area. According to Lupyan et al., this brain region is associated with verbal processes, and therefore cathodal participants were hindered in forming verbal category learning rules necessary when using RB strategies compared to sham participants, so they instead relied on an II learning strategy that did not necessitate such verbal processing.

However, as can be seen in Figure 5.1, the stimuli in the training clusters A and B are much further away from the II decision boundary than from the unidimensional decision boundaries where there is overlap between stimuli in the conditions. As outlined earlier in the thesis, this difference of category separation between the II and RB conditions can potentially impact diverse cognitive processes complicating the inferences that can be made (Newell et al., 2011; Stanton & Nosofsky, 2007). Furthermore, the strategy that Perry and Lupyan (2014) take as providing evidence for an implicit II approach is also entirely consistent with the idea that participants group new stimuli into the category to which they are most similar (i.e., an exemplar based approach, e.g., Nosofsky, 1985; 1986). With these category structures, therefore, it is very difficult to make strong interpretations about the strategy which is being employed and consequently the specific effect that tDCS has.

Taken together, this previous research suggests that different category learning tasks are not affected in the same way by tDCS. While Kincses et al. (2003) and Vercammen et al. (2011) find that performance in the weather prediction task is improved by anodal tDCS, and Perry and Lupyan (2014) report an increase in the engagement of II category learning processes after cathodal tDCS, Ambrus et al. (2011) and McLaren et al. (2016) report that anodal tDCS hinders learning during prototype distortion tasks. To my knowledge there are no tDCS studies that directly compare performance on the traditional RB and II categories used by COVIS (e.g., Filoteo et al., 2010). Examining how the RB and II category structures are modulated by tDCS would therefore help to build a more nuanced picture of the impact that tDCS has on category learning. Investigating this could also inform the research field pertaining to the predictions of COVIS. If performance on II and RB categories is affected differently by tDCS then this could be interpreted as evidence that there are neurally separable systems operational when learning these category structures. Also, if tDCS can be reliably found to increase category learning accuracy in a variety of task types then the method could be practically applied in clinical and educational environments to optimise learning.

In order to investigate whether tDCS will impact RB or II category learning, or indeed, whether stimulation will differentially affect performance in these conditions, Experiment 7 will present participants with the stimuli and decision boundaries of Filoteo et al. (2010) as used in Chapters 2-4. These category structures control for non-essential differences between conditions that could affect learning, such as task difficulty, category separation and the number of stimuli dimensions relevant for learning. Participants learnt the conjunctive RB or II category while undergoing 10 minutes of 1.5mA anodal tDCS to the left DLPFC. Anodal stimulation was chosen for this study as it has previously been found to improve performance on category learning

tasks (e.g., Kincses et al., 2003; Vercammen et al., 2011; but see Ambrus et al., 2011). Anodal stimulation was applied to the left DLPFC (Fp3) in the present study as tDCS to this region has been found to impact category learning (e.g., Ambrus et al., 2011; Kincses et al., 2003; McLaren et al., 2016; Vercammen et al., 2011).

On the basis of the clear neurobiological predictions that COVIS makes, one might expect that stimulation to the left DLPFC, a key region of the explicit system but not the implicit system, should modulate RB learning to a greater extent than II learning. On the other hand, given the results of Chapters 2-3 where no neural differences emerged between the II and RB conditions, and Chapter 4 where no behavioural dissociation was observed between category structures, one might not expect any difference between category structures. Alternatively, single system accounts could reasonably predict that if there is a dissociable pattern of results, this would be in the direction of a greater effect in II learning after tDCS stimulation than RB learning. This would be consistent with the arguments put forward in Chapter 2 that II learning is more neurally demanding than RB learning, and might, therefore, be more sensitive to any effect of tDCS stimulation. Whilst any predictions must inevitably be somewhat tentative given the novel nature of this experiment, it should, regardless of the precise pattern of results, be a valuable addition to our currently impoverished understanding of the influence that tDCS has on category learning.

Method

Participants

80 University of Exeter students (21 male, mean age 20.03 years, SD= 1.91) participated in the study. Participants gave written informed consent and confirmed they passed the safety criteria such as no history of headaches, blackouts, vertigo, ear

trouble, or family history of epilepsy (see Appendix 5.1) before participating. They were remunerated £6 for participation.

Design

There were 2 independent variables in this study: category structure (RB or II) and stimulation type (anodal or sham). This gave a 2 (category structure) X 2 (stimulation type) between-subjects design resulting in 4 conditions: RB with anodal stimulation; RB with sham stimulation; II with anodal stimulation; II with sham stimulation, each containing 20 participants. The dependent variable measured in this experiment was participants' accuracy of responding in each block of 100 trials.

Stimuli

The stimuli were the two-dimensional II and conjunctive RB category structures employed by Filoteo et al. (2010) and used in Chapters 2-4 (see Figure 5.2). As before, the rule necessary for optimal categorization in the RB condition was 'short, upright lines belong in category A, and the rest in category B'. In the RB and II category structure there were 600 stimuli that could be separated into two categories of 300 items. Each stimulus was a black line varying on two dimensions: length and orientation. As in Filoteo et al. (2010), there was 5% overlap between the categories so that the maximum accuracy attainable was 95%.

This image has been removed by the author of this thesis for copyright reasons.

Figure 5.2. The category structures used in Filoteo et al. (2010) and the present study; (a) the conjunctive rule-based condition; (b) the information-integration condition. Solid lines indicate the decision boundary separating category A and category B. Figure taken from Filoteo et al. (2010).

TDCS

The tDCS procedure stimulating the left DLPFC was the same as employed by McLaren et al. (2016). A neuroconn battery driven constant current stimulator (<http://www.neuroconn.de/>) was used to deliver tDCS through two electrodes. These electrodes were inserted into tight sponge pouches, soaked in water and then saline solution. The anodal electrode was placed over the left prefrontal cortex (corresponding to Fp3) and the reference electrode placed above the right eyebrow. The Fp3 was located by measuring half the distance between the inion and nasion. From this point 7cm forward was measured, and then 9cm to the left at a right angle. This area was marked with a hypoallergenic marker. A wet wipe was then used to wipe around this mark and above the right eyebrow.

In the anodal and sham conditions, tDCS was administered at an intensity of 1.5mA, with a fade in and fade out of 5 seconds. The fade in slowly increases stimulation to full intensity over 5 seconds at the start of stimulation, and the fade out

reduces it to nothing over 5 seconds at the end of the task to reduce side effects (Ambrus et al., 2013; Luedtke et al., 2015). In the anodal condition, the stimulation lasted for 10 minutes and began 1.5 minutes before the categorization task. In the sham condition stimulation lasted for 30 seconds, but was still initiated 1.5 minutes before the task began. There was no stimulation concurrent with the task in the sham condition. The use of 30 seconds of stimulation in sham conditions has been found to induce the sensation of tDCS without instigating behavioural effects (Poreisz, Boros, Antal, & Paulus, 2007).

A double blind procedure was employed by setting the neuroconn stimulator to 'study mode'. A double blind procedure was implemented so that the experimenter could not unwittingly influence the participant with expectations about task performance or tDCS symptoms. The neuroconn study mode allows preset conditions to be activated by a random code by an experimenter, who cannot tell what condition has been implemented. A researcher external to this study generated codes for each participant corresponding to the stimulation type, and indicated whether the II or RB E-Prime script should be loaded for that participant.

Procedure

Prior to the study participants entered the laboratory and filled out the appropriate tDCS safety and consent forms. The tDCS was then set up as described above and the participant's code was entered into the stimulator. When the participant was comfortable, the experimenter explained the possible side effects of the tDCS and informed them that when the stimulation began they would have 1.5 minutes to read the 4 instruction screens presented on the computer in front of them. They were told that this would be plenty of time to read all the instructions. When the participant was happy to begin, the experimenter started the stimulation and timed the 1.5 minutes of instructions. Participants were instructed that they had to learn into which of two

categories a series of stimuli belonged using feedback presented after every trial. When 1.5 minutes had passed the experimenter started the categorization task. This task was programmed using E-Prime (Psychological Software Tools, 2002) for the presentation and timing of stimuli and collection of response data.

The trial-by-trial procedure for the RB and II conditions was identical. Each trial began with a black fixation cross presented in the centre of the screen for 250ms. A stimulus then appeared in the middle of the screen until participants responded by pressing the 'z' key on the computer keyboard if they thought the item belonged to category A, or the 'm' key if they thought the item was a member of category B. Feedback ("Correct" or "Incorrect") was then displayed for 500ms and the next trial then immediately began. There were 600 categorization trials presented in 12 blocks of 50 items separated by a self-paced break. The proportion of correct responses was measured for every block of 100 trials (amounting to 6 blocks).

On average, participants finished the 600 learning trials in 20 minutes, with the first 300 trials (blocks 1 to 3) being completed in a mean time of 9 minutes and 9 seconds. For the majority of participants the 8.5 minutes of stimulation coinciding with category learning stopped during block 3 of the task.

Results

In a mixed-design ANOVA, the average accuracy of responding in each of the six trial blocks was entered as the within-subjects factor (6 levels). Category structure and stimulation condition were entered as between-subject factors. The mean accuracy of responding across all blocks can be seen in Figure 5.3. There was a main effect of block on accuracy ($F(5, 380)=31.4, p<.001, \eta^2_p=.29$) and a significant linear trend on this

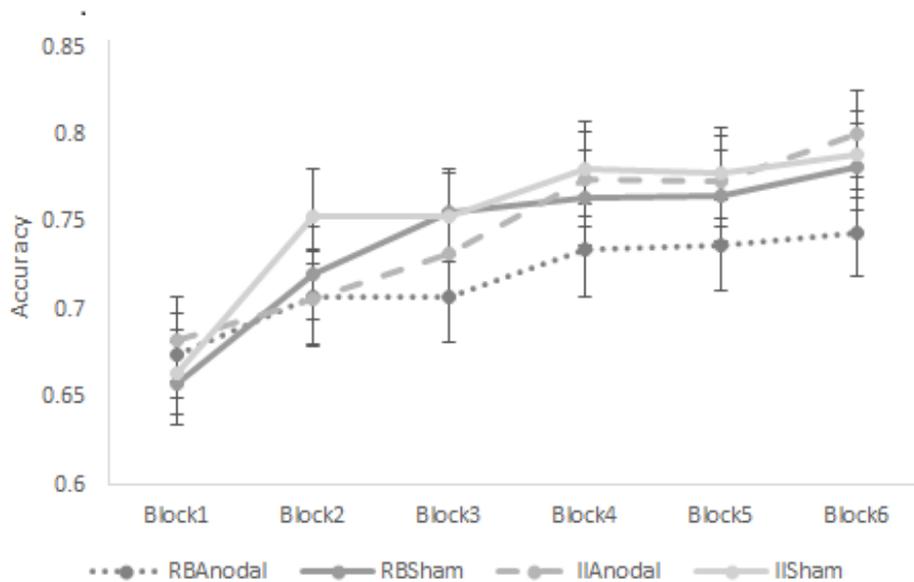


Figure 5.3. Mean accuracy of responding in Experiment 7 across blocks. Error bars represent standard error.

factor ($F(1,76)=72.06, p<.001, \eta^2_p=.49$) demonstrating that participants improved across blocks. However, there was no significant main effect of category structure on response accuracy ($F(1,76)=.82, p=.369, \eta^2_p=.01, BF=0.14^{14}$) and no significant main effect of stimulation condition on performance ($F(1,76)=.5, p=.484, \eta^2_p=.01, BF=0.07$). There were no significant interactions found between block and category structure ($F(5, 380)=.51, p=.768, \eta^2_p=.01$), block and stimulation ($F(5, 380)=1.7, p=.133, \eta^2_p=.02$), or between category structure and stimulation ($F(1,76)=.13, p=.725, \eta^2_p<.01, BF=0.2$). The three way interaction between block, category structure, and stimulation was non-significant ($F(5, 380)=0.96, p=.442, \eta^2_p=.012$).

Performance accuracy during the first half of the study (blocks 1-3) was also examined. This was a planned analysis because block 3 is of significance, as this is the block in which anodal participants ceased receiving tDCS stimulation. This means that

¹⁴ As there is no study that has previously examined the effect of tDCS on RB and II conditions a non-informative uniform prior baseline was used (e.g., Gelman, 2002; Yang & Berger, 1998). These analyses were completed using the Bayes calculator designed by Dienes (http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm).

an analysis of the first three blocks alone gives a measure of learning accuracy during stimulation itself. In a mixed-design ANOVA accuracy in the three blocks was entered as the within-subject factor, and stimulation condition and category structure were entered as fixed factors (see Figure 5.4. for means). There was a main effect of block on accuracy ($F(2, 152)=24.96, p<.001, \eta^2_p=.25$) with participants improving linearly ($F(1,76)=35.42, p<.001, \eta^2_p=.32$). However, there was no significant main effect of category structure on response accuracy ($F(1,76)=.26, p=.613, \eta^2_p=.003, BF=0.08$) and no significant main effect of stimulation on performance ($F(1,76)=.5, p=.486, \eta^2_p=.01, BF=0.07$). However, there was a significant interaction between block and stimulation ($F(2, 152)=4.23, p=.016, \eta^2_p=.05$) which showed a greater increase in performance across the three blocks in the sham participants compared to the anodal participants (see Figure 5.4). There were no significant interactions found between block and category structure ($F(2, 152)=.1, p=.908, \eta^2_p=.001$) or between category structure and stimulation condition ($F(1,76)<.001, p=.991, \eta^2_p<.01, BF=0.1$). The three way interaction between block, stimulation and category structure was non-significant ($F(2, 152) = 1.1, p=.335, \eta^2_p = .01$).

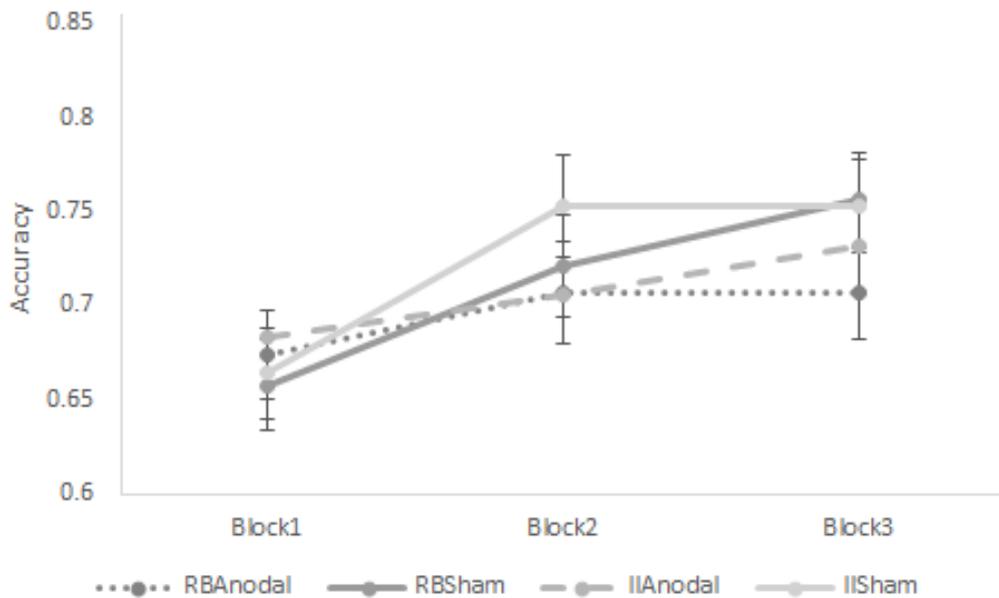


Figure 5.4. Mean accuracy of responding in Experiment 7 across the first three blocks. Error bars represent standard error.

Performance accuracy during the second half of the study (blocks 4-6) was also examined. In a mixed-design ANOVA accuracy in the last three blocks was entered as the within-subject factor, and stimulation condition and category structure were entered as fixed factors (see Figure 5.5. for means). The main effect of block on accuracy was marginally significant ($F(2, 152)=2.39, p = .096, \eta^2_p=.03$). There was no significant main effect of category structure on response accuracy ($F(1,76)=1.39, p=.241, \eta^2_p=.018, BF=0.23$) and no significant main effect of stimulation on performance ($F(1,76)=.41, p=.526, \eta^2_p=.01, BF=0.21$). There were no significant interactions found between block and stimulation ($F(2, 152)= .06, p=.943, \eta^2_p=.001$), block and category structure ($F(2, 152)=.12, p=.887, \eta^2_p=.002$) or between category structure and stimulation condition ($F(1,76) = 0.43, p=.513, \eta^2_p=.01, BF=0.22$). The three way interaction between category structure, stimulation and block was non-significant ($F(2, 152)=.48, p=.619, \eta^2_p=.01$).

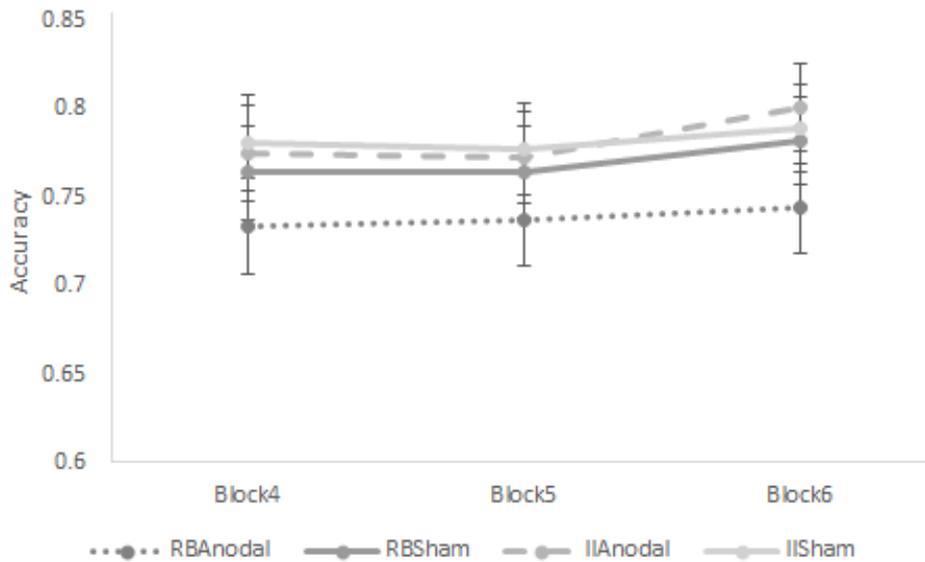


Figure 5.5. Mean accuracy of responding in Experiment 7 across the last three blocks. Error bars represent standard error.

Discussion

Experiment 7 provided the first systematic investigation of the effect of anodal stimulation to the left DLPFC on the learning of RB and II COVIS category structures. Previously, Kincses et al. (2003) and Vercammen et al. (2011) found that prefrontal anodal stimulation enhanced performance on the weather prediction task, while Ambrus et al. (2011) and McLaren et al. (2016; Experiment 2) found that anodal tDCS to the DLPFC hindered learning in a prototype distortion task compared to sham and cathodal participants. Somewhat differently to these previous tDCS studies, the present research found that anodal tDCS to the left DLPFC had no detectable effect on category learning when analysing all 6 blocks of the task. The Bayesian analysis on the non-significant main effect of stimulation suggests that there is strong evidence supporting the null hypothesis (Dienes, 2011) that anodal and sham participants did not differ significantly in accuracy. Therefore, in the category learning research field it seems that the effect

that tDCS to the left DLPFC has on performance is dependent on the category learning task type presented to participants.

However, Experiment 7 also analysed participants' accuracy of responding during trials that were concurrent with stimulation (the first three blocks of the study). There was a significant interaction found whereby participants in the sham conditions improved more across blocks than the anodal participants. This finding suggests that tDCS only impacts learning in the first half of the present study when the stimulation is occurring. This is somewhat surprising given that previous research claims that tDCS effects can last up to 90 minutes after stimulation (Nitsche & Paulus, 2011); however, there are few papers that investigate the longevity of tDCS effects, with very limited work concerning cognitive tasks (for a review see Nitsche et al., 2008) and to my knowledge none in the categorization domain. The procedure used in Nitsche and Paulus measured motor response potential in a finger after anodal tDCS, and it did not require participants to perform a cognitive task. It may be that the effects of tDCS on motor excitability in the fingers are stronger, and longer lasting, however, the results of the present study suggest that the impact of anodal tDCS on more higher process tasks may be weak. However, Figure 5.5 suggests that RB participants receiving anodal stimulation are also less accurate in later learning than all II participants or RB sham participants. While this interaction does not reach significance, this possibility should be investigated further. It may be that increasing the duration or the intensity of tDCS will reveal a significant isolated effect of stimulation to the left DLPFC on RB learning.

The present study reports that learning in RB and II conditions is dampened in blocks that are synonymous with anodal stimulation. This effect seemingly continues into blocks 4 – 6 (see Figure 5.5), although analysis of this does not reach significance. This pattern of results is intriguing given that tDCS research in the cognitive field

assumes that the effect of brain stimulation on higher process tasks last beyond application (e.g., Ambrus et al., 2011, who give the majority of tDCS prior to task performance). Given the past assumptions regarding tDCS, this is a somewhat surprising outcome; therefore, it is important to try and replicate this pattern of results, as well as investigate the potential of a longer term effect of anodal stimulation on RB learning. However, if the present findings are verified in their current form, then it could have widespread implications on how tDCS studies are designed, and the analyses performed on behavioural data sets. If this finding is replicable then future research should ensure that the targeted trial type is synchronised with tDCS, and should focus on analysing tDCS behavioural data in trials where brain stimulation is concurrent with the process being targeted. If this effect of anodal tDCS on performance is robust then it also narrows the applications of the method, especially if the finding generalises to other research fields, as it would only be very specific task designs shown to be enhanced by tDCS (e.g., the weather prediction task) that would be useful to clinical or educational fields.

While some previous tDCS research has found that anodal tDCS to the DLPFC improves category learning (e.g., the weather prediction task, Kincses et al., 2003; Vercammen et al., 2011), others have found category performance impeded from anodal stimulation (e.g., the prototype distortion task, Ambrus et al., 2010; McLaren et al., 2016; Experiment 2). The present study fits with the dampening results reported in prototype distortion tasks as anodal tDCS was found to hinder RB and II performance. However, unlike the prototype distortion tasks, in the present study this effect was only significant during blocks where participants were still receiving stimulation. As previously mentioned, the methods used in these category learning tasks are likely the cause of the differing effects of tDCS, for example there are differences in the stimuli and category structures used as well as the tDCS application. Also, while double

blinding was used in the current study and McLaren et al. (2016) to ensure experimenters expectations did not impact results, other previous studies have only used single blinding, where the experimenter was aware of a participant's stimulation condition (e.g., Ambrus et al., 2011, Kincses et al., 2003; Vercammen et al., 2011).

The present study revealed no significant main effect of category structure on performance and no significant interaction between stimulation condition and category structure in any analysis, suggesting that the effect of tDCS on category learning was not reliably modulated by task type. In the present research, a single system theory could account for such findings. However, there is an non-significant, but visually noticeable, interaction between category structure and stimulation in the last 3 blocks of learning, whereby RB participants perform worse than II participants in anodal conditions. This could offer evidence that converges with COVIS theory, which suggests that RB learning preferentially implicates frontal systems compared with II conditions (e.g., Ashby & Valentin, 2005). Nevertheless, as the condition of tDCS did not statistically differentiate RB and II learning, it is hard to use such results to verify the predictions made by COVIS, and the model should not be championed based on the present study. Again, future research should further investigate the predictions of COVIS by employing tDCS at greater intensities, or for longer periods, to examine whether RB learning will be differentially affected by stronger stimulation of the DLPFC compared to II learning.

Experiment 7 conducted the first investigation of the effect of anodal tDCS on RB and II learning. During anodal stimulation participants demonstrated a smaller increase in performance compared to sham participants, an effect selective only to blocks with concurrent tDCS. Further investigation is now needed to examine if this

effect is replicable, and if RB participants are hindered more by tDCS than II participants in category learning performance proceeding stimulation.

Appendix 1 Chapter 5: Transcranial Stimulation safety-screening questionnaire

NAME OF PARTICIPANT		
Left or right handed?.....	Sex:	M / F
Date of birth.....		
Have you previously had an MRI scan at the University of Exeter?		Yes/No
If so, are you happy for us to access your existing MRI data in this study?		Yes/No

Before receiving Transcranial Stimulation, please read the questions below carefully and provide answers. For a small number of individuals, brain stimulation may carry an increased risk of causing a seizure. The purpose of these questions is to make sure that you are not such a person. You have the right to withdraw from the screening and subsequent scanning if you find the questions unacceptably intrusive. The information you provide will be treated as strictly confidential and will be held in secure conditions. If you are unsure of the answer to any of the questions, please ask the person who gave you this form or the person who will be performing the study.

	Yes	No
Do you suffer from epilepsy, or did you ever have a seizure or convulsion?		
Does anyone of you close relatives have epilepsy?		
Have you ever had severe (i.e., followed by loss of consciousness) head trauma?		
Have you ever had a severe concussion?		
Did you ever suffer from a condition to your brain, such as meningitis?		
Did you ever have a stroke?		
Do you suffer from multiple sclerosis?		
Do you have brain damage, as a result of brain surgery or a disease?		
Did you ever have a surgical procedures to your spinal cord?		
Do you have spinal or ventricular derivations?		
Do you suffer from another neurological condition?		
Do you have metal in the brain/skull (except titanium)? (e.g., retainer, splinters, fragments, surgical clips, etc.)		
Do you have cochlear implants?		

Do you have an implanted neurostimulator? (e.g., DBS, epidural/subdural, VNS)		
Do you have a cardiac pacemaker or intracardiac lines or metal in your body?		
Do you have a medication infusion device?		
Do you have a history of drug abuse or alcoholism?		
Do you have diabetes?		

	Yes	No
Are you pregnant or is there any chance that you might be?		
Do you hold a heavy goods vehicle driving license or bus license?		
Do you take psychiatric or neuroactive medication (e.g. antidepressants)?		
Do you suffer from frequent vertigos or headaches?		
	Yes	No
Is there any congenital deafness in your family?		
Do you have any hearing problems or ringing in your ears?		
Do you have a condition of your cervical vertebrae (e.g. spondylolysis, arthritis or scoliosis)?		
Are you taking any other medications? If yes, please describe below in which occasion(s).		
Have you ever had any adverse effects to TMS or tDCS in the past?		
Have you ever had a fainting spell, syncope or absence? If yes, please describe below in which occasion(s).		
Did you recently have a panic attack?		
Do you experience claustrophobia?		
Have you ever had any adverse effects to MRI?		
Do you have a skin disease (or did you have one in the past)?		

Additional information:

Name and contact details GP:

Transcranial Stimulation Pre-Session Screening

To minimise the risk of Transcranial Stimulation causing an adverse effect, it is important that you answer the following questions accurately before we begin the session.

	Yes	No
In the last 12 hours, have you consumed more than 3 units of alcohol?		
Have you taken recreational drugs in the last 24 hours?		
Did you get a good night's sleep last night, and do you feel alert?		
In the last two hours, have you consumed more than two cups of coffee?		

Chapter 6. Investigating the prototype effect with transcranial Direct Current Stimulation

Chapter 6 focuses on the prototype effect whereby participants respond more accurately to a previously unseen prototype of a trained category compared to previously seen (Palmeri & Nosofsky, 2001) or previously unseen (Posner & Keele, 1968) category exemplars derived from that prototype. While the prototype effect has been extensively studied, much of this evidence is purely behavioural and it is only in recent years that it has received attention from the neuroscience community. One recent, but still little explored, question is how prototype learning can be modulated by tDCS application to the DLPFC. The first study to examine this was conducted by Ambrus et al. (2011) who found that participants performed less accurately in anodal conditions than sham conditions when performing an (A, not A) prototype distortion task; however, anodal tDCS was also found to abolish a prototype effect that was present in sham participants. McLaren et al. (2016, Experiment 2a) also found that 1.5mA of anodal stimulation to the left DLPFC led to decreased accuracy in responding to novel exemplars in a prototype derived category learning task, but with a concurrent enhancement of a prototype effect not present in participants who received sham stimulation. Similarly, in Experiment 2b McLaren et al. found that response accuracy to previously seen exemplars in anodal conditions decreased compared with those receiving cathodal stimulation. Again, though, there was a significant prototype effect in anodal participants that was not present in cathodal participants.

However, as discussed on pages 34-36 in Chapter 1, the conclusions drawn from these two studies may be influenced by the methods used. For example, in Ambrus et al. participants received the majority of tDCS stimulation before the task, which has previously been found to impede performance on motor tasks (Kuo et al., 2008). In McLaren et al., (2016; Experiment 2) there was a potential ceiling effect in the sham

and cathodal conditions due to very accurate performance on both exemplar and prototype trials. This may have masked the presence of a prototype effect in these conditions, as an increase in accuracy to prototype trials would not be apparent when compared with performance in exemplar trials that is nearly perfect (e.g., see Figure 1.4, page 35 in Chapter 1). The aim of the present chapter is to follow up the findings of McLaren et al. to investigate whether employing a more difficult task reduces the accuracy in each stimulation condition and reveals a prototype effect in sham and cathodal participants, as well as in the anodal condition.

The neural effect of tDCS on prototype distortion task performance will also be examined using fMRI. Not only will this allow for the investigation of neural activation during prototype presentation, but will also demonstrate if the left DLPFC and its functionally connected areas are differentially activated after anodal tDCS in comparison with areas engaged after sham tDCS. A relatively restricted number of studies to date have examined the neural correlates of the prototype distortion task independent of tDCS (Aizenstein et al., 2000; Seger et al., 2000; Zeithamova, Maddox, & Schnyer, 2008). One such study, however, was conducted by Zeithamova et al. (2008) who found that prototype distortion tasks implicate a wide neural network including the frontal, parietal and occipital lobes. A surprising gap in this existing literature, though, is that no previous fMRI study has looked at whether neural activation to category prototypes differs from activation to exemplars. A primary aim of Chapter 6, therefore, is to investigate brain regions implicated during prototype responding independently of neural responding to exemplars. This will be the focus of Experiment 10.

Experiment 8

Experiment 8 investigated the effect of anodal and cathodal tDCS to the left DLPFC on the prototype effect during a prototype distortion task¹⁵. All the experiments presented in this chapter use an (A, B) prototype distortion paradigm. Three prototype derived chequerboard categories of the original four categories created by McLaren (1997) were presented to participants to learn in a training phase. Although these visual stimuli were the same as used by McLaren et al. (2016, Experiment 2), three categories were presented instead of two categories to increase the difficulty of the task and reduce performance compared to McLaren et al., eliminating any potential ceiling effects. This would consequently provide a more sensitive indicator of whether there is greater accuracy when responding to prototype trials compared to exemplar trials in either sham or cathodal participants. After training, participants were tested on their categorization accuracy on three types of trial – previously seen exemplars presented in training (old), previously unseen exemplars of each category (new), and the previously unseen prototype from each category. These trials were all presented without feedback. All tDCS was given at the intensity of 1.5mA as in Chapter 5 and McLaren et al. (2016). TDCS lasted for 10 minutes and began just 1.5 minutes before the task so as to negate any potential effects of too much tDCS preceding the task (e.g., Kuo et al., 2008). Stimulation started 1.5minutes prior to the task in the sham condition, but lasted for just 30 seconds.

When considering the results of McLaren et al. (2016; Experiment 2), one could predict that in anodal conditions performance would be worse overall than in sham and cathodal participants on the prototype distortion task, yet a prototype effect would be

¹⁵ Experiment 8 was published as Experiment 1 in McLaren, Carpenter, Civile, McLaren, Ku, Zhao, & Milton (2016)

present in anodal participants but not in sham or cathodal participants. On the other hand, if a ceiling effect in sham and cathodal participants in McLaren et al. is masking the presence of a prototype effect, then results could reveal that sham and cathodal participants also demonstrate a prototype effect in the present study, where accuracy in responding will likely be reduced compared to McLaren et al. (2016; Experiment 2). Conversely again, on the basis of the findings reported by Ambrus et al., one might hypothesise that participants in the anodal condition would have reduced learning overall, and would not demonstrate a prototype effect that is present in sham participants. These hypotheses were first tested in Experiment 8, which used a between-subjects design with participants performing the three category chequerboard prototype distortion task in either an anodal, cathodal, or sham condition.

Method

Participants

50 University of Exeter students (17 males, mean age 21.5 years, $SD = 2.93$) participated in the experiment. Two participants were excluded before analysis due to the computer malfunctioning during the study. Participants gave written informed consent and conformed to all safety criteria before participating (e.g. no history of headaches, concussion, family history of epilepsy; see Appendix 5.1 for complete safety criteria) and they were remunerated £6 for participation.

Stimuli

The stimuli used were a subset of black and white chequerboards from McLaren (1997). These chequerboards were 16x16 squares that contained approximately 50% black and 50% white squares. Three of the four categories (A, B and C) of McLaren (1997) were used; these three categories were derived from three chequerboard

prototypes. The three prototypes used shared 50% of their squares with one another, and each prototype was created to have a relatively visible area of predominately black or white squares in order to distinguish each prototype from the other two (see Figure 6.1). In order to generate the exemplars of these categories the prototypes were distorted - random noise was added to the prototype by changing some of its squares from black to white or vice versa. Each prototype contained 256 squares; to create an exemplar 96 squares of the prototype were randomly selected and re-set to either black or white. Given chance, about half of these squares would be reassigned the same colour as before, so approximately 48 squares changed from a prototype to create an exemplar (see McLaren, 1997, for more details).

The training and test phases were run using E-Prime on a Dell latitude laptop. In both phases, participants were instructed to separate chequerboard stimuli into the three categories (either A, B or C). In the training phase, participants were shown 64 novel exemplars from each category. During the test phase, for each category participants responded to 10 of the exemplars seen during training (old), 10 new category exemplars, and the previously unseen prototype (presented twice; see Figure 6.1 for category prototypes)¹⁶.

¹⁶ A paired sample t-test reveals no significant difference in accuracy of responding to the first prototype presentations compared to the second ($t(47)=0.850$, $p=0.399$, $d = 0.09$, where M_1 = presentation set 1 and M_2 = presentation set 2).

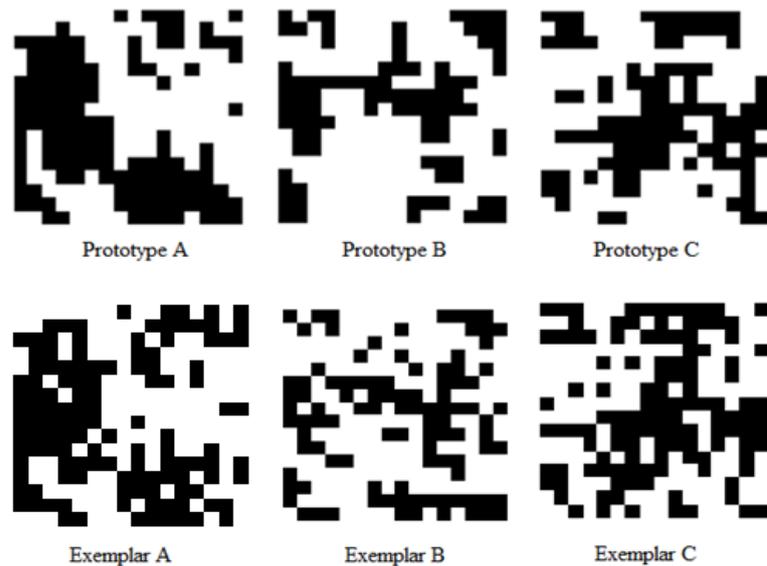


Figure 6.1. The prototypes and typical examples of their derived exemplars for the categories used in Experiments 8, 9 and 10.

TDCS

The tDCS was delivered by two electrodes attached to a Neuroconn battery driven constant current stimulator. Electrodes were inserted into tight fitting 5cm by 7cm synthetic sponge pouches that were dampened with water and saline solution. The first electrode was placed over the left prefrontal cortex (corresponding to Fp3) and the second, reference electrode (the electrode of opposing current that is not placed over the area of intended stimulation; Nitsche et al., 2008) was placed above the right eyebrow. The Fp3 was located by measuring half the distance between the inion and nasion over the centre of the scalp, and from this point moving 7cm forward towards the nasion, and then 9cm to the left. This spot was marked with a hypoallergenic marker, and both areas that would be under the electrodes cleansed with a sensitive wet wipe.

The tDCS was given at an intensity of 1.5mA, with a fade in and fade out of 5 seconds in all conditions. The stimulation was anodal or cathodal depending on the positioning of the two cables inserted into the stimulator. In the anodal and cathodal conditions, the current began 1.5 minutes before the participant started the training

trials. During this time, participants were given the instructions for the task. 10 minutes of stimulation was given to each participant, resulting in 8.5 minutes of stimulation during the training phase. In the anodal sham and cathodal sham conditions stimulation still began 1.5 minutes before the training task; however, stimulation only lasted 30 seconds so it had stopped by the time the training task had begun. This procedure was the same as used by McLaren et al. (2016; Experiment 2).

The double blind procedure. To ensure that the experimenter could not inadvertently influence the performance of participants due to expectation, a double blind procedure was implemented using two experimenters. Experimenter 1 ran the study and gave instructions to participants, whereas Experimenter 2 set up the tDCS based on conditions provided by a third party. After the participant had filled out the relevant safety information, the tDCS montage (electrode positioning) was set up by Experimenter 1 with the assistance of Experimenter 2. Experimenter 1 then left the room as the stimulation parameters and current type were set up by Experimenter 2. After this was completed Experimenter 1 re-entered the room, and when the participant was happy to begin, the stimulation was started by Experimenter 2 who then left the room. Experimenter 1 proceeded to explain the instructions for the task, and 1.5 minutes into stimulation the training phase began.

Design and procedure

Participants were randomly allocated to one of the three between-subject stimulation conditions. The anodal and cathodal stimulation conditions each contained 16 participants. The sham conditions also contained 16 participants, of which 8 received anodal sham stimulation and 8 received cathodal sham stimulation.

When the tDCS stimulation began, participants read through 3 instruction screens that informed them that they would see trials of single chequerboard stimuli that

they had to categorize as belonging to category A, B or C using the keys ‘C’, ‘V’, or ‘B’ respectively. They were told that at first they would be guessing, but from the feedback received they should be able to learn which chequerboards belonged in which category. This instruction phase lasted 1.5 minutes. Subsequently, the training phase began where participants were shown 192 novel chequerboard exemplars (64 from each category) in three blocks of 64 randomised trials. Each trial consisted of a fixation cross in the centre of the screen which lasted for 500ms, followed by the chequerboard for 3 seconds. The participant made their response in this time but the stimulus remained on the screen for the full 3 seconds. Corrective feedback (‘correct’, ‘incorrect’ or ‘time out!!!’) was then presented for 750ms. Following the procedure of McLaren et al. (2016; Experiment 2) the next trial began immediately after the feedback ceased. After completing the 3 training blocks participants were told that the tDCS was switched off and the head set was removed. Participants then performed the test phase which was a single block consisting of 66 trials of old and new exemplars and the prototype from each category presented in a randomised order. The procedure was the same as in the training phase, except no feedback was given and the next trial began immediately after stimulus presentation.

Results

Figure 6.2 demonstrates the mean response accuracy in all stimulation conditions for the training phase and all trial types at test. A one-way ANOVA was completed with accuracy of responding (proportion of correct responses) in the training trials entered as the dependent factor. Stimulation condition (anodal, cathodal, or sham) was entered as the fixed factor. There was no significant effect of stimulation condition on accuracy of responding during training trials ($F(2,45)=.41$, $p=.669$, $\eta^2_p=.018$; Figure 6.2.a). A mixed-design ANOVA was completed to analyse whether stimulation affected performance on the three test phase trial types (old, new, and prototype). Stimulation condition was entered as the between-subjects factor and performance in

old, new and prototype trials entered as a repeated measures variable. There was no significant main effect of stimulation on response accuracy during the test phase ($F(2,45)=1.06$, $p=.354$, $\eta^2_p=.045$; see Figure 6.2.b). However, there was a significant interaction between stimulation and test trial type ($F(4,90) = 2.72$, $p = .035$, $\eta^2_p=.108$) and a significant main effect of trial type, ($F(2,90)=6.38$, $p=.003$, $\eta^2_p=.124$), with prototype trials more accurately responded to than old and exemplar trials (Figure 6.2.b).

To investigate the prototype effect, the average accuracy of responding to exemplar trials during test was subtracted from average accuracy of responding to prototype trials for each participant. A positive score indicated that the participant was performing better in prototype trials than exemplar trials, and a negative score the opposite. This calculation was performed in three ways first comparing performance in prototype trials to average accuracy in responding to old exemplars, and then comparing prototype performance to average accuracy in responding to new exemplars. Finally, performance in prototype trials was also compared to the average performance in both new and old exemplar trials combined. There were two reasons for including three baseline measures (old, new, and average exemplar trials): first, all three baseline measures were used in McLaren et al. (2016) which this study aimed to closely replicate. Secondly, Experiments 8 and 9 were designed to inform a MRI study that would be the first exploration of the neural correlates of the prototype effect. There is no previous research to suggest whether prototype trials are neurally responded to differently compared to old or new trials. Therefore, Experiment 10 will use all three baselines to complete the first investigation of brain activation differences to exemplars and prototype presentations.

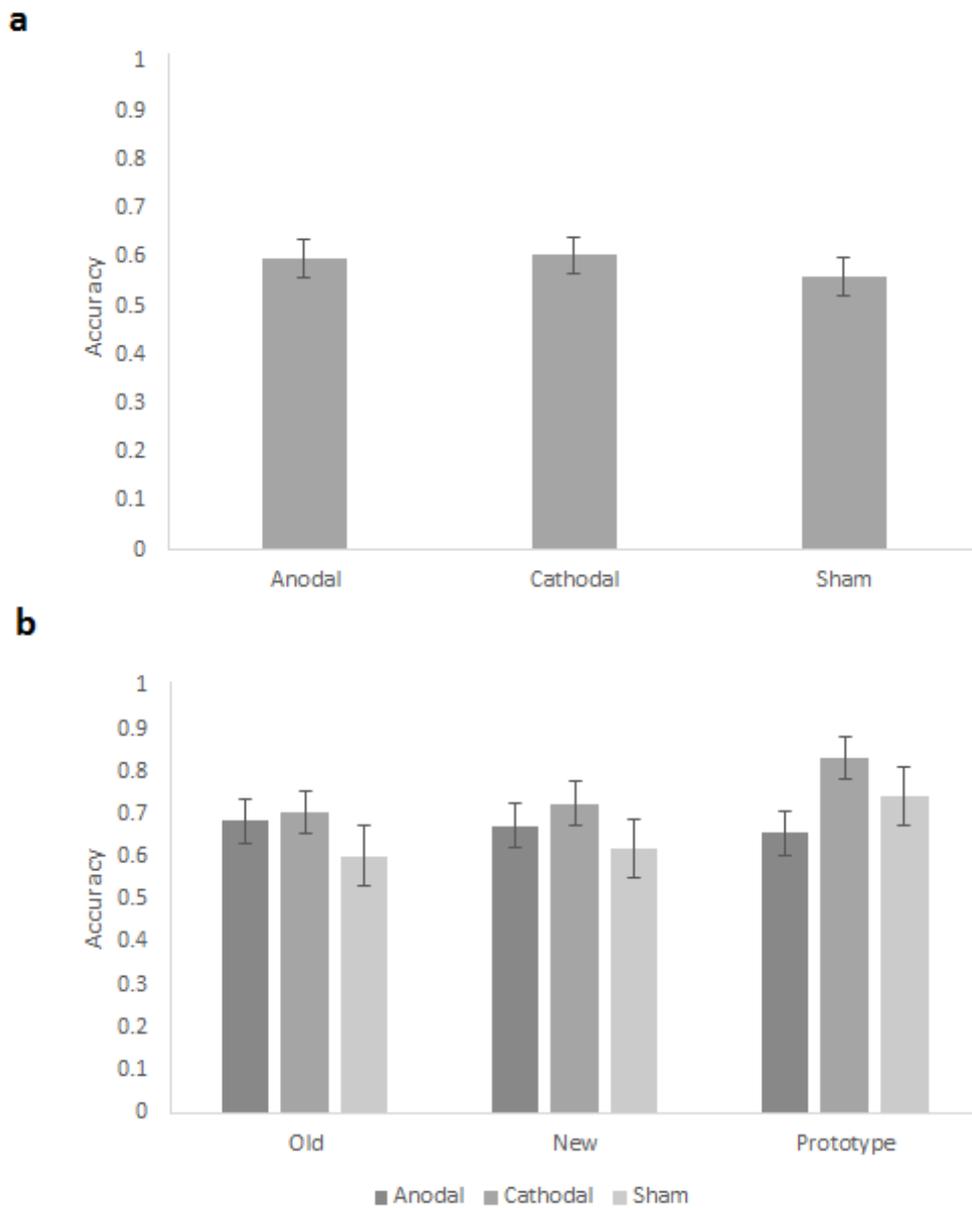


Figure 6.2. Proportion of correct categorization responses in the anodal, cathodal, and sham conditions in Experiment 8; (a) the accuracy of responding during the training phase; (b) the accuracy of responding during old, new and prototype trials during the test phase. Error bars show standard error.

Figure 6.3.a shows the mean difference in accuracy of responding on prototype trials compared with old trials (i.e. prototype performance – old performance). The calculated prototype effect was entered as a DV into a one-way ANOVA with the three stimulation conditions entered as the fixed factor. There was a main effect of stimulation ($F(2,45)=4.11$, $p=.023$, $\eta^2_p=.154$). Planned comparisons indicate that participants in the anodal condition had a significantly smaller prototype effect than participants in the cathodal condition ($F(1,45)=5.84$, $p=.020$, $\eta^2_p=.115$) and the sham condition ($F(1,45)=6.48$, $p=.014$, $\eta^2_p=.126$). However, the prototype effects of cathodal and sham participants did not differ significantly from one another ($F(1,45)=.017$, $p=.897$, $\eta^2_p<.001$, $BF=0.32$).

In order to assess if the prototype effects found in each condition were significantly different from zero, the mean difference in accuracy between prototype and old trials for each participant was compared to 0 (which represents no difference in response accuracy between the prototype and old stimuli) in a one-sample t-test. This was conducted separately for each stimulation condition. The prototype effect was significantly bigger than zero in cathodal participants ($t(15)=3.33$, $p=.005$, $d = 1.15^{17}$) and sham participants ($t(15)=2.73$, $p=.016$, $d = 0.99$). However, the prototype effect was not significant in the anodal condition ($t(15)=-.57$, $p=.578$, $BF=0.16$, $d = 0.22$) and was numerically in the opposite direction.

¹⁷ In this chapter, Cohen's d for the prototype effect was calculated with M_1 = prototype effect in the cathodal, anodal and sham conditions, $M_2 = 0$.

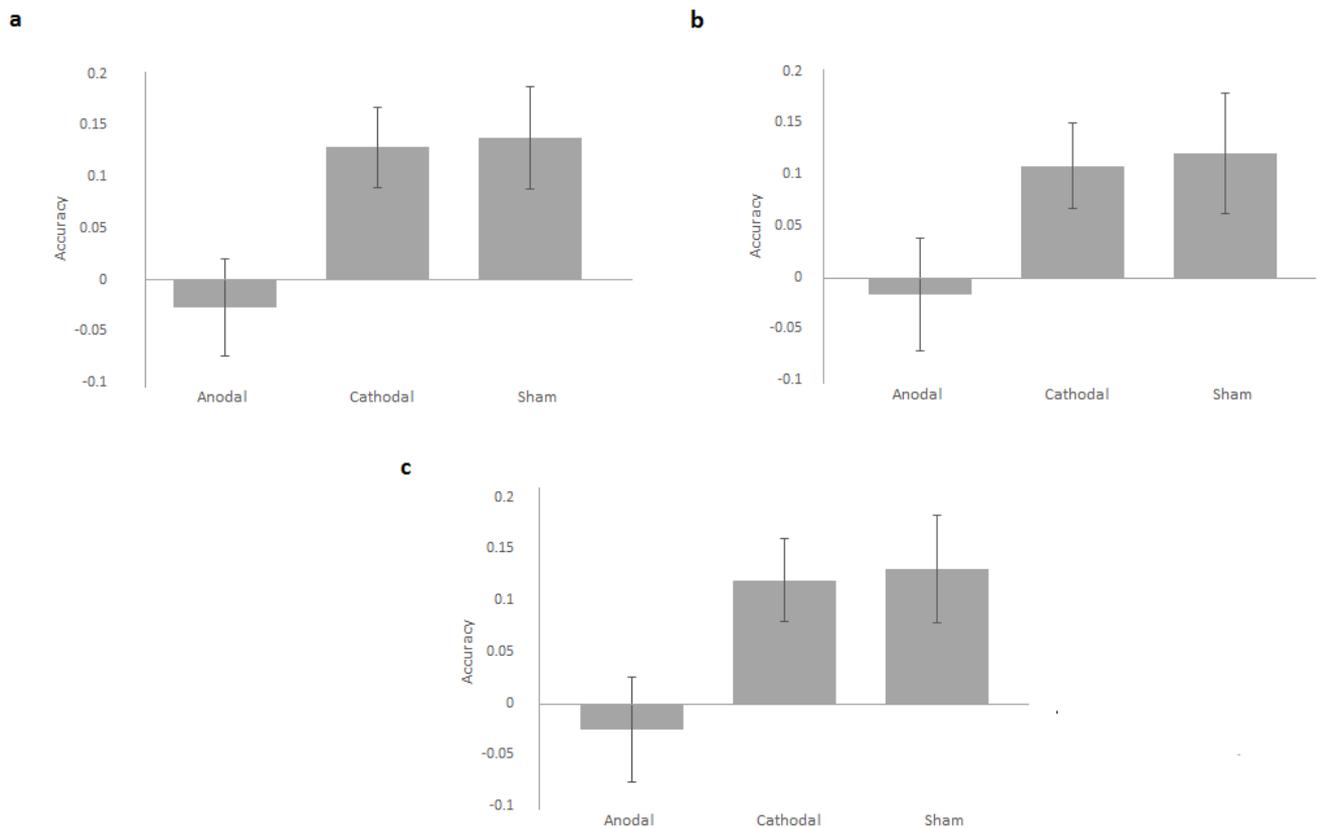


Figure 6.3. The prototype effects present in Experiment 8; (a) calculated by subtracting accuracy in old trials from prototype trials; (b) calculated by subtracting accuracy in new trials from prototype trials; (c) calculated by subtracting accuracy in old and new trials from prototype trials. Positive bars represent more accurate performance in prototype trials compared to exemplar trials. Error bars represent standard error.

An alternative measure of the prototype effect was calculated by subtracting performance on new exemplars from performance on prototype items (i.e., prototype performance – new exemplar performance). Figure 6.3.b shows these mean differences. The prototype effect mean differences were again entered as a DV into a one-way ANOVA with the three conditions of stimulation entered as the fixed factor. Unlike the analysis of the prototype effect calculated using old trials, there was no significant effect of stimulation ($F(2,45)=2.11$, $p=.133$, $\eta^2_p=.86$). However, when performing planned contrasts on this analysis, participants in the anodal condition trended towards having a smaller prototype effect than participants in the cathodal condition ($F(1,45)=2.86$, $p=.098$, $BF=0.13$, $\eta^2_p=.06$) and the sham condition ($F(1,45)=3.46$,

$p=.070$, $BF=0.12$, $\eta^2_p=.07$). The prototype effects of cathodal and sham participants did not differ significantly from one another ($F(1,45)=.029$, $p=.867$, $BF=1.01$, $\eta^2_p=.001$).

One-sample t-tests revealed that the prototype effect was significantly bigger than zero in the cathodal condition ($t(15)=2.6$, $p=.020$, $d = 0.92$) and approaching significantly bigger than zero in sham participants ($t(15)=2.04$, $p=.060$, $BF= 1.39$, $d = 0.71$). However, the prototype effect was not significantly bigger than zero in the anodal condition ($t(15)=.31$, $p=.762$, $BF=0.2$, $d = 0.13$).

A final measure of the prototype effect was calculated by averaging the accuracy of responding to old and new trials for each participant to give an overall measure of exemplar accuracy. This was then subtracted from average prototype accuracy to calculate the prototype effect (see Figure 6.3.c). The prototype effect was entered as a DV into a one-way ANOVA with the three conditions of stimulation entered as a fixed factor. A significant main effect of stimulation was found ($F(2,45)=3.3$, $p=.046$, $\eta^2_p=.128$). When performing planned contrasts on this analysis, participants in the anodal condition displayed a smaller prototype effect than participants in the cathodal condition ($F(1,45)=4.57$, $p=.038$, $\eta^2_p=.092$) and the sham condition ($F(1,45)=5.31$, $p=.026$, $\eta^2_p=.11$). The prototype effects of cathodal and sham participants did not differ from one another ($F(1,45)=.028$, $p=.869$, $BF=0.99$, $\eta^2_p=.001$).

One-sample t-tests indicated that the prototype effect was significantly bigger than zero in the cathodal condition ($t(15)=2.99$, $p=.009$, $d = 1.06$) and in sham participants ($t(15)=2.5$, $p=.025$, $d = 0.88$), but not in the anodal condition ($t(15) = .51$, $p=.620$, $BF=0.17$, $d = 0.21$).

Discussion

Experiment 8 found that participants in the cathodal and sham conditions consistently demonstrated a significant prototype effect. However, this prototype effect was not present in participants who had received anodal tDCS. These findings are contradictory to McLaren et al. (2016; Experiment 2) who found a prototype effect present in participants who had received anodal tDCS to the left DLPFC that was not present in cathodal or sham participants. The addition of a third category to learn in Experiment 8 made the task more difficult in order to account for the potential ceiling effect in the sham and cathodal conditions of McLaren et al. (2016; Experiment 2) where near perfect accuracy on the exemplar trials could have prevented detection of superior performance on the prototype trials. Increasing the difficulty of the task in the present study could have allowed for prototype effects in control conditions to be better observed, which may explain why the present study finds the effect in sham and cathodal participants when McLaren et al. did not. However, this argument cannot explain why performance in anodal conditions differed between Experiment 8 and McLaren et al. (2016). This result needs to be further investigated to see if it is replicable.

The present results are broadly similar to the findings of Ambrus et al. (2011) who also report an abolition of a prototype effect after tDCS to the left DLPFC that *is* present in control participants. However, while Ambrus et al. found the prototype effect and overall learning diminished after anodal stimulation, Experiment 8 reports that the prototype effect can be abolished with tDCS, while overall learning across all other trial types at test remains unaffected. It should be noted, though, that methodological differences between Ambrus et al. and Experiment 8, such as the timing and intensity of stimulation, the number of stimuli presented, and the stimuli themselves, make direct

comparisons between the two studies suboptimal, and could account for the discrepancy.

The results of the present study and Ambrus et al. offer evidence that stimulation of the left DLPFC influences the prototype effect. However, these experiments cannot conclude whether the DLPFC is more or less engaged in anodal participants as a result of stimulation, which could potentially provide insight into whether anodal tDCS is inhibiting or exciting activation in this area. It is also possible that anodal stimulation may influence regions functionally connected to the left DLPFC, which may be contributing to the behavioural effect found. This possibility led to the formulation of Experiment 10, an fMRI study which aimed to investigate the neural areas more (or less) activated during prototype distortion learning after anodal tDCS compared with sham conditions. In order to do this, however, I needed to modify the procedure of Experiment 8 to make it suitable for an fMRI protocol. Experiment 9 aimed to replicate the results of Experiment 8 and examine whether the effect persists with the revised fMRI-compatible procedure.

Experiment 9

Experiment 8 found that anodal tDCS to the left DLPFC abolished a prototype effect that was present in sham and cathodal participants. The first aim of Experiment 9 was to replicate the abolition of the prototype effect in anodal participants found in Experiment 8 in order to confirm that the results did indeed contradict the behavioural pattern of McLaren et al. (2016). However, while Experiment 8 allowed for inference regarding the behavioural impact of tDCS on the prototype effect, conclusions about the neural areas representing the prototype effect engaged in anodal participants compared

to sham participants could not be made from the behavioural data alone. Therefore, Experiment 10 was designed which aimed to examine the potential neural effects of tDCS during a prototype distortion task through a combined tDCS and fMRI design. While fMRI has previously been used to explore the neural effects of tDCS on non-cognitive functions such as resting state neural connectivity (Peña-Gómez, et al., 2013) and on neural activation during motor tasks (e.g., Baudewig, Nitsche, Paulus, & Frahm, 2001; Kim et al., 2012), there is very little research that examines the effect of tDCS on cognitive tasks using fMRI. To my knowledge there has been no tDCS and neuroimaging research studying the neural effect of anodal tDCS on category learning. Such a study would allow a greater insight into the neural areas implicated by anodal stimulation of the left DLPFC in a prototype distortion task.

However, in order to investigate the neural effects of tDCS on the prototype distortion task, the procedure implemented in Experiment 8 had to be adapted for fMRI experimentation. Firstly, for safety reasons, all tDCS had to be completed before the participants entered the MRI scanner and began the training task. Although there is evidence to suggest that the effects of tDCS last up to 90 minutes (Nitsche & Paulus, 2011) which would encompass the entirety of the fMRI session, completing all 10 minutes of stimulation prior to training could impact the results found in comparison to Experiment 8 (e.g., Kuo et al., 2008), particularly as there is potential that different stimulation timings may have driven the discrepancy in results between Ambrus et al. (2011) and McLaren et al. (2016; Experiment 2). Chapter 5 also highlights how performance concurrent with tDCS stimulation can be differentially affected compared to performance that is not concurrent with stimulation. Experiment 9 aimed to ensure that the pattern of results found in the three category prototype distortion task of Experiment 8 was not changed by this modification to stimulation timing.

The test phase from Experiment 8 was also modified. Currently, no fMRI study directly examines neural activity during prototype presentation, nor the neural correlates of the prototype effect, as the prototype stimulus is not presented enough times for an analysis to reach appropriate power (e.g., Zeithamova et al., 2008). Experiment 10 aimed to rectify this, and examine the brain regions activated during prototype presentation. In order to build up a reliable neural signal of prototype classification, the category prototypes need to be presented more often than in Experiment 8 (and in previous related imaging studies, e.g., Aizenstein et al., 2000; Reber et al., 1998; Zeithamova et al; 2008). Each prototype was now presented 8 times during test. To examine whether repeatedly presenting the prototype could affect responding towards it, an old and a new exemplar from each category were also repeatedly presented 8 times along with 8 non-repeated old and new exemplars. The second aim of Experiment 9 was, therefore, to ensure that the influence of anodal tDCS on the prototype effect remained when repeatedly presenting the prototype at test. In Experiments 9 and 10, only anodal and anodal sham conditions were tested, as in Experiment 8 there were no significant differences in performance between cathodal and sham participants. In order to utilize a matched comparison in terms of current direction the anodal sham condition was used as the control.

Method

Participants

32 University of Exeter students (8 males, mean age 19.6 years, SD = 1.41) participated in the study. As before, participants gave written informed consent and confirmed they adhered to all safety criteria before participating, and they were remunerated £6 for participation.

Stimuli

The stimuli used were the same three categories of chequerboard stimuli as Experiment 8, with the identical training phase. In the test phase there were 120 trials separated into two blocks of 60. The test phase consisted of 40 trials for each category consisting of 8 presentations of the prototype, 8 repeated presentations of an old stimulus, 8 repeated presentations of a new stimulus, and a further 8 old and 8 new non-repeated stimuli. Category responses were made as before using the “C”, “V” and “B” keys on a laptop keyboard.

TDCS

The tDCS was delivered in the same montage with the same intensity, duration, fade in and fade out as Experiment 8. However, in Experiment 9 the full 10 minutes of stimulation was given before the training task began. In this study only an anodal and a sham anodal condition were used. As before, the sham stimulation lasted for 30 seconds. The same double blind procedure was used as in Experiment 8.

Design and procedure

Participants were randomly allocated to one of the between-subject conditions: anodal or sham stimulation. After the full 10 minutes of stimulation (or sham stimulation) participants were told that the tDCS had finished. The tDCS headset was then removed, and participants read through the instructions and the training phase began. Here a fixation cross was presented for 250ms, followed by the stimulus which remained on the screen for 3 seconds during which time the participant made their response. Feedback was presented for 500ms. There was a varied ITI between each trial lasting between 500-3500ms (the ‘jitter’ required for the fMRI protocol in Experiment 10).

The training trials were presented in 4 blocks of 48 trials with an 8 second break between blocks. After the training phase participants then performed the test phase which consisted of two blocks of 60 trials. The test procedure was the same as in the training phase. However, after the 3 second stimulus presentation, instead of feedback presentation there was a blank screen presented for an ITI of 500-3500ms before the next trial began.

Results

Repeated trial accuracy compared with non-repeated trials

In order to determine if repeating a stimulus 8 times during the test phase influenced performance, accuracy in responding to non-repeated and repeated exemplars was compared. In a 2x2 ANOVA response accuracy was entered as the dependent variable. Trial repetition (repeated or non-repeated) and trial type (old or new) were entered as the fixed factors. There was little numerical difference in response accuracy between these four conditions (the mean proportion of correct responses was .63 in repeated old trials (SD=.25), .62 in old non-repeated trials (SD=.22), .57 in new repeated trials (SD=.26), and .61 (SD=.25 in new non-repeated trials). There was no significant main effect of trial repetition ($F(1,124)=.11$, $p=.74$, $\eta^2_p=.001$), or trial type ($F(1,124)=.58$, $p=.45$, $\eta^2_p=.005$), and no interaction between trial type and repetition ($F(1,124)=.24$, $p=.628$, $\eta^2_p=.002$)¹⁸.

Analysis of training and test accuracy

A one-way ANOVA revealed that there was no significant effect of stimulation on accuracy during training (anodal mean accuracy = 0.62, SD = 0.16; sham mean

¹⁸ Average performance to repeated exemplars (old and new) and prototype was compared in a paired samples t-test. Participants responded to the prototypes more accurately than repeated exemplars ($t(31)=2.90$, $p=.007$, $d = 0.26$, where M_1 =prototype, and M_2 = exemplar trials).

accuracy = 0.53, SD = 0.18; $F(1, 30)=2.02$, $p=.166$, $\eta^2_p=.063$). A mixed-design ANOVA with accuracy of responding to old, new and prototype test trials entered as the within-subjects factor, and stimulation condition entered as the between-subjects factor revealed no significant main effect of stimulation ($F(1, 30)=.59$, $p=.448$, $\eta^2_p=.019$; Figure 6.4). Performance differed significantly between the three trial types, ($F(2, 60)=8.55$, $p=.001$, $\eta^2_p=.222$) with prototype items demonstrating the highest accuracy. However, the interaction between stimulation condition and trial type was not significant ($F(2, 60)=1.79$, $p=.175$, $\eta^2_p=.056$).

The prototype effect was calculated in the same way as Experiment 8, again using comparisons of prototype trial accuracy with the accuracy of old trials and new trials, and the average accuracy to all exemplars. Figure 6.5.a shows the mean difference in accuracy of responding to prototype trials compared with old trials at test. Again, positive bars reflect the prototype effect whereby participants perform better on prototype trials compared with old trials. The prototype effect was significantly bigger than zero in the sham condition ($t(15)=3.46$, $p=.003$, $d = 1.29$) but not in the anodal condition ($t(15)=.42$, $p=.682$, $BF=0.4$, $d = 0.14$). There was a trend towards the prototype effect in the anodal condition being smaller than that of sham participants

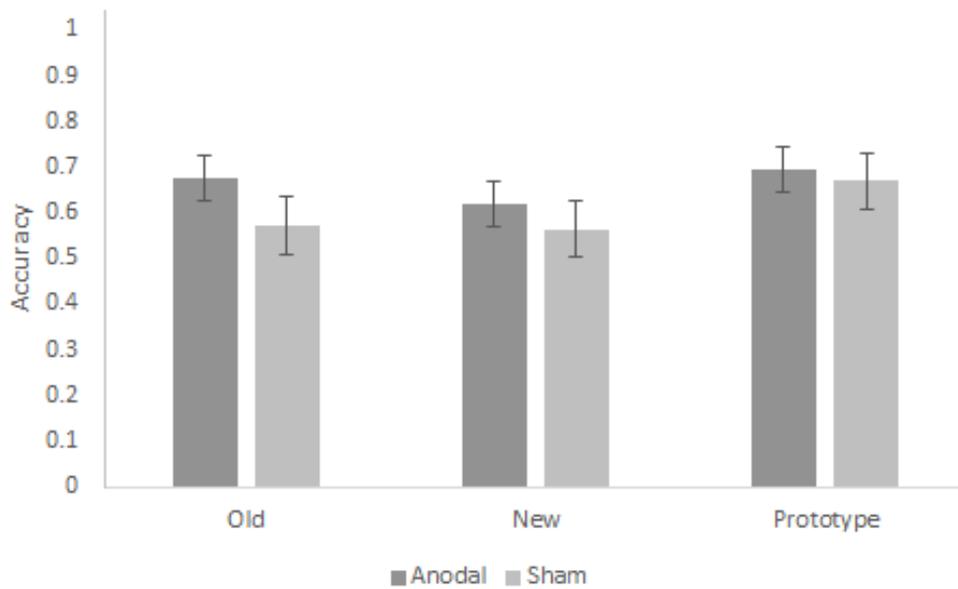


Figure 6.4. Proportion of correct categorization responses during old, new and prototype trials during the test phase of Experiment 9 in the anodal and sham anodal conditions. Error bars represent standard error.

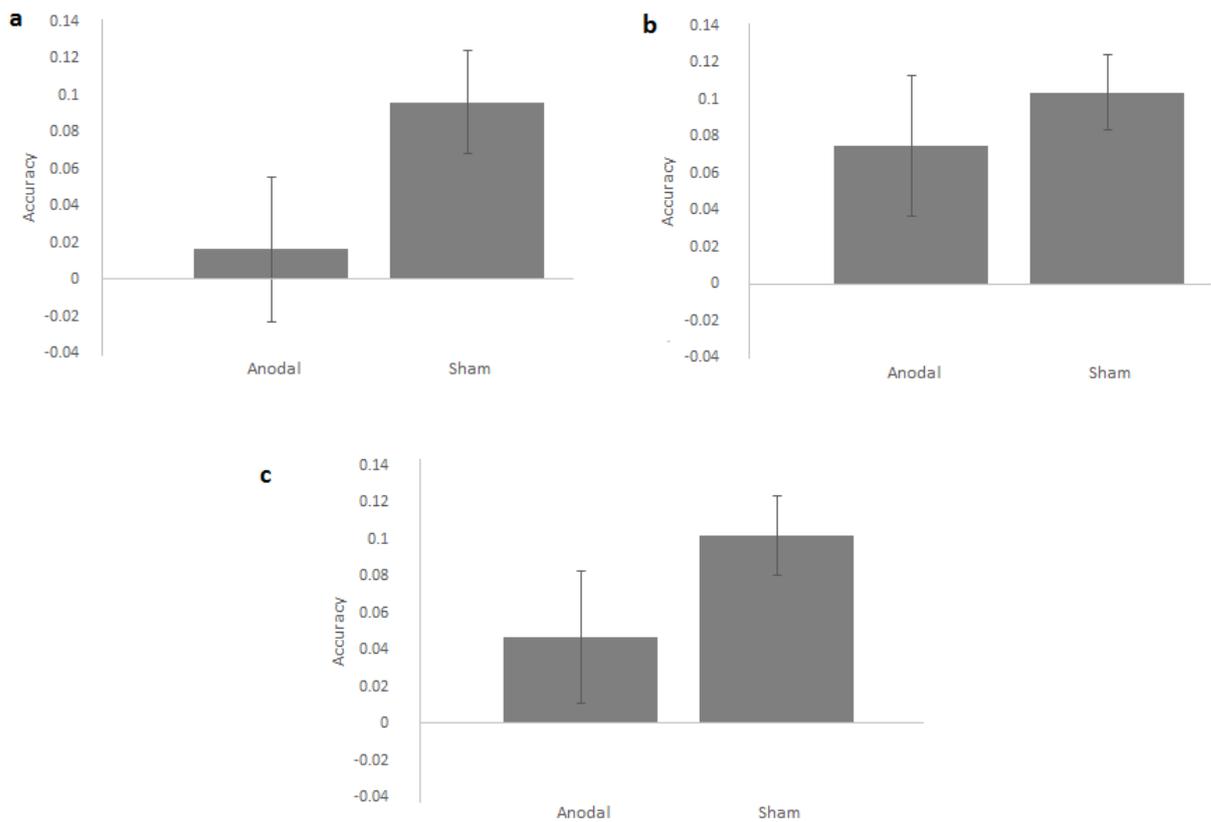


Figure 6.5. The prototype effects present in Experiment 9; (a) calculated by subtracting accuracy in old trials from prototype trials; (b) calculated by subtracting accuracy in new trials from prototype trials; (c) calculated by subtracting accuracy in old and new trials from prototype trials. Positive bars represent more accurate performance in prototype trials compared to exemplar trials. Error bars represent standard error.

($t(30)=1.67$, $p=.053$ (one-tailed), $BF=0.14$, $d = 0.58^{19}$). A one-tailed analysis was planned in advance as the analogous comparison for the same analysis in Experiment 8 yielded a significant result in this direction.

Figure 6.5.b shows the mean difference in accuracy of responding to prototype trials at test compared with new trials. One-sample t-tests revealed the prototype effect was significantly bigger than zero in the sham condition ($t(15)=5.11$, $p<.001$, $d = 1.77$) with a non-significant trend in the anodal condition ($t(15)=1.97$, $p=.068$, $BF=2.78$, $d = 0.75$). The prototype effects in the anodal and sham conditions did not differ significantly ($t(30)=.67$, $p=.511$, $BF=0.06$, $d = 0.17$).

Figure 6.5.c shows the mean difference in prototype accuracy compared with average exemplar accuracy at test. The prototype effect was significantly bigger than zero in the sham condition ($t(15)=4.68$, $p<.001$, $d = 1.77$) but not in the anodal condition ($t(15)=1.32$, $p=.207$, $BF=0.76$, $d = 0.51$). However, the size of the prototype effects in the anodal and sham conditions did not differ significantly ($t(30)=1.32$, $p=.197$, $BF=0.09$, $d = 0.44$).

Combined analysis of Experiment 8 and Experiment 9

The above analyses were repeated combining the data from the anodal and anodal sham participants of Experiments 8 and 9. While such cross-experiment comparisons should, of course, be taken with caution, they provide some insight as to whether the procedural differences had any significant impact on the pattern of responding. This approach is also likely to increase the power of the analysis due to the increased sample size.

¹⁹ In this chapter, comparisons of the prototype effects in the anodal and sham participant were calculated with $M_1 = \text{Sham}$, $M_2 = \text{Anodal}$.

A two-way ANOVA was completed with training accuracy entered as the DV. The common conditions of stimulation (anodal or sham anodal) between Experiments 8 and 9, and the experiment number (Experiment 8 or 9) were entered as fixed factors. There was again no significant main effect of stimulation condition ($F(1, 52)=.93$, $p=.339$, $\eta^2_p=.018$; Figure 6.6.a) and no significant main effect of experiment number ($F(1, 52)=.24$, $p=.624$, $\eta^2_p=.005$) on training trials. The interaction between stimulation and experiment was not significant ($F(1, 52)=.85$, $p=.360$, $\eta^2_p=.016$; Figure 6.7.a). A mixed-design ANOVA was completed with response accuracy in old, new and prototype test trials entered as the within-subjects factor, and stimulation condition and experiment number entered as the between-subject factors. There was no significant main effect of stimulation ($F(1, 52)=.05$, $p=.831$, $\eta^2_p=.001$; Figure 6.6.b) and no significant main effect of experiment ($F(1, 52)=1.86$, $p=.178$, $\eta^2_p=.035$) on test trial performance. There was a significant difference in responding between the different trial types at test ($F(2, 104)=4.24$, $p=.017$, $\eta^2_p=.075$), with prototypes responded to more accurately. The interaction between trial type and stimulation was approaching significance ($F(2, 104)=2.79$, $p=.066$, $\eta^2_p=.051$). There was no interaction between stimulation and experiment number ($F(1, 52)=1.53$, $p=.222$, $\eta^2_p=.029$; Figure 6.7.b), or trial type and experiment number ($F(2, 104)=1.01$, $p=.369$, $\eta^2_p=.019$).

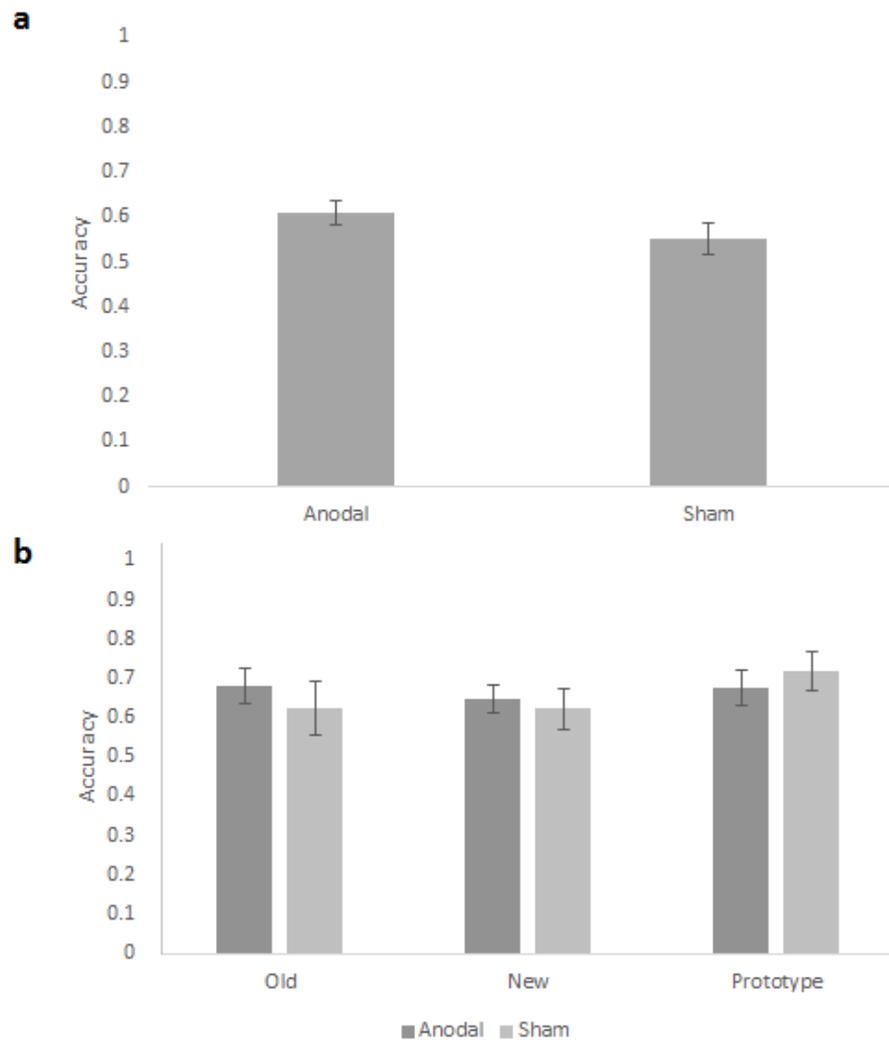


Figure 6.6. Proportion of correct categorization responses in the anodal and sham anodal conditions of Experiments 8 and 9 combined; (a) the accuracy of responding during the training phase; (b) the accuracy of responding during old, new and prototype trials during the test phase. Error bars represent standard error.

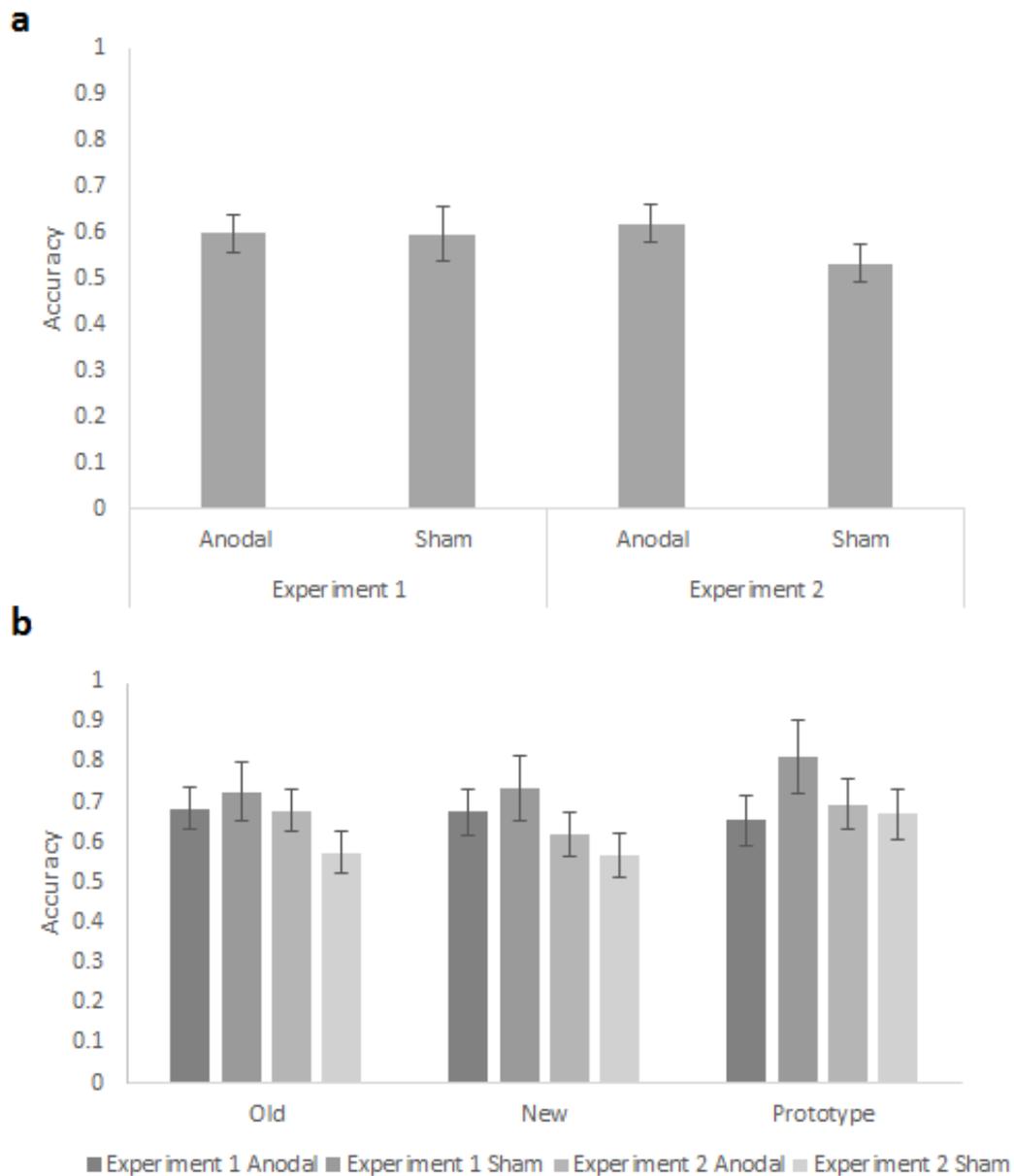


Figure 6.7. Proportion of correct categorization responses in the anodal and sham anodal conditions of Experiments 8 and 9; (a) the accuracy of responding during the training phase; (b) the accuracy of responding during old, new and prototype trials during the test phase. Error bars represent standard error.

Figure 6.8 shows the mean difference in the average accuracy of responding to prototypes compared with performance on old trials, new trials, and the average of old and new trials at test across both experiments. When comparing prototype trial performance to old trials the prototype effect was significantly bigger than zero in the sham condition ($t(23)=3.24$, $p=.004$, $d = 0.91$), but not in the anodal condition ($t(31)=-.18$, $p=.862$, $BF=0.4$, $d = 0.08$) and the main effect of stimulation was significant, as the prototype effect in sham participants was significantly larger than that in the anodal participants ($F(1, 52)=4.58$, $p=.037$, $\eta^2_p=.081$; Figure 6.8.a). The main effect of experiment on the prototype effect was not significant ($F(1, 52)=.35$, $p=.559$, $\eta^2_p=.007$) and there was no interaction between stimulation condition and experiment ($F(1, 52)=.14$, $p=.715$, $\eta^2_p=.003$).

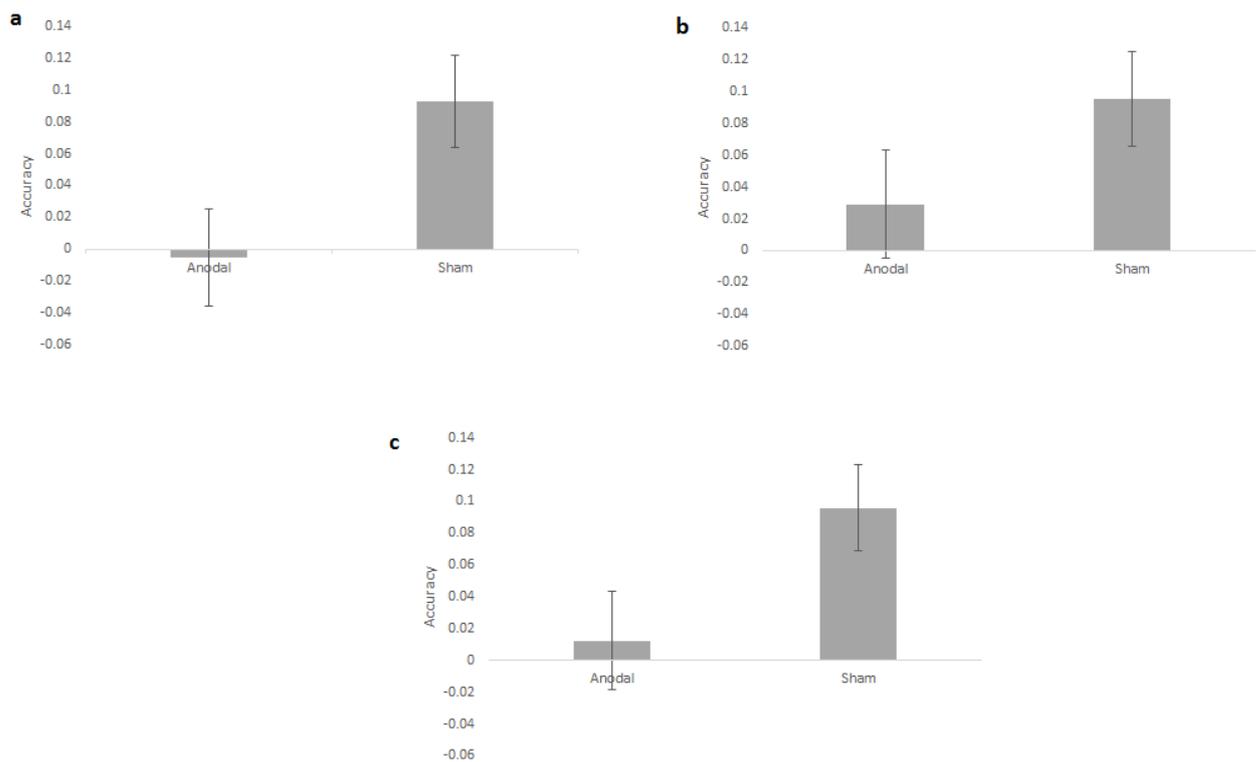


Figure 6.8. The prototype effects present in Experiments 8 and 9; (a) calculated by subtracting accuracy in old trials from prototype trials; (b) calculated by subtracting accuracy in new trials from prototype trials; (c) calculated by subtracting accuracy in old and new trials from prototype trials. Positive bars represent more accurate performance in prototype trials compared to exemplar trials. Error bars represent standard error.

When comparing prototype trial performance to new trials the prototype effect was significantly bigger than zero in the sham condition ($t(23)=3.21$, $p=.004$, $d = 0.94$) but not in the anodal condition ($t(31)=.86$, $p=.397$, $BF=0.26$, $d = 0.22$). The prototype effect was not significantly different in anodal and sham participants ($F(1, 52)=1.66$, $p=.204$, $\eta^2_p=.031$; Figure 6.8.b). The main effect of experiment was not significant ($F(1, 52)=1.46$, $p=.232$, $\eta^2_p=.027$), with no interaction present between stimulation condition and experiment ($F(1, 52)=.48$, $p=.492$, $\eta^2_p=.009$).

Comparing performance in prototype trials to performance in both old and new trials also revealed a prototype effect significantly bigger than zero in the sham condition ($t(47)=4.61$, $p<.001$, $d = 0.91$) but not in the anodal condition ($t(63)=.52$, $p=.602$, $BF=0.08$, $d = 0.16$). There was a trend towards a main effect of stimulation, with a larger prototype effect in sham participants than in anodal participants ($F(1, 52)=3.4$, $p=.071$, $\eta^2_p=.061$; Figure 6.8.c). The main effect of experiment was not significant ($F(1, 52)=1.03$, $p=.314$, $\eta^2_p=.019$) and there was no interaction between stimulation and experiment ($F(1, 52)=.37$, $p=.545$, $\eta^2_p=.007$).

Discussion

The results of Experiment 9 found the same basic behavioural pattern as Experiment 8 - anodal tDCS abolished a prototype effect that was present in sham participants.

However, these effects were somewhat less pronounced than in Experiment 8, although statistically the differences between experiments were not significant indicating that the procedural changes had no impact on the basic pattern of results. Unlike Experiment 8, the prototype effects found in sham participants were only marginally larger than in anodal participants during old trials, and were not significantly larger than those found in anodal conditions during new trials. To account for the numerically weaker effects in Experiment 9 compared with Experiment 8, Experiment 10 increased the sample size

from 16 to 24 participants per condition to improve the sensitivity of the experiment. It was also found that repeatedly presenting exemplar trials did not affect accuracy of responding compared to non-repeated trials. Experiment 10 included repeated old and new trials in the test phase so that prototype trials were not the only repeated stimuli presented.

Experiment 10

As discussed in Chapter 1, previous neuroimaging research has found a wide network of neural regions implicated in learning prototype distortion tasks. For example, the occipital lobes are often implicated in these tasks (e.g., Aizenstein et al., 2000; Reber et al., 1998, Seger et al., 2000). Areas of the parietal lobes and temporal lobes, including the MTL, are also recruited during (A, B) prototype distortion learning (e.g., Seger et al., 2000; Zeithamova et al., 2008). Of particular relevance given the site of stimulation in my experiments, Seger et al. (2000) found the right DLPFC activated during (A, B) learning. They also found that more accurate learners engaged the left DLPFC more than less accurate learners.

Experiments 8 and 9 demonstrate that anodal tDCS to the left DLPFC influences the behavioural presence of the prototype effect. However, the spatial resolution of tDCS is limited in part due to the relatively large electrodes used, which likely could stimulate adjacent neural areas inadvertently. The positive influence of non-stimulated brain areas functionally connected to the targeted region also limits spatial resolution (Nitsche et al., 2007). Therefore, Experiment 10 aimed to further investigate the whole brain neural processes that underlie the impact of tDCS on the prototype effect. This experiment also examined the neural response during prototype presentation. As previously discussed, the brain regions activated during prototype presentation have not

yet been investigated in the literature; this study could consequently allow inferences as to whether prototypes are differently processed in the brain compared to previously seen exemplars derived from such prototypes.

The predictions for Experiment 10 are resultant from past fMRI work that used prototype distortion tasks. It is expected that overall learning in Experiment 10 will replicate previous fMRI studies that report activation in the frontal, parietal and the occipital lobes (e.g., Seger et al., 2000; Zeithamova et al., 2008). Nosofsky et al., (2012) also considered the possibility that prototype extraction relies on brain regions such as the MTL for the calculation of the summed similarity of all exemplars seen. It is therefore possible that the MTL will be more active in prototype trials compared with exemplar trials, as the prototype stimuli will have the greatest summed similarity of all the chequerboards shown at test.

Method

Participants

Participants were 48 University of Exeter students (19 males, mean age 20.53 years, $SD = 5.47$). As in Experiment 9, participants gave written informed consent and confirmed they adhered to all safety criteria for tDCS, as well as MRI safety criteria (e.g., no metallic implants) before participating. Participants were remunerated £7 for participation.

Stimuli

The training and test chequerboard stimuli were identical to those of Experiment 9. Responses to stimuli were now made on a hand held key pad with three buttons in the MRI scanner. Category A corresponded to the far left button, B the middle button and C the far right button on the keypad.

TDCS

The tDCS was delivered in the same montage at the same intensity, duration, fade in and fade out as Experiments 8 and 9. Anodal participants received 10 minutes of stimulation in a room next to the MRI scanner. Anodal sham participants unknowingly had only 30 seconds of anodal stimulation and then a further 9.5 minutes of no stimulation before entering the MRI scanner. The same double blind procedure was used as in Experiments 8 and 9.

FMRI data acquisition

Each participant completed one scanning session in a 1.5-T Phillips Gyroscan magnet, equipped with a Sense coil. As in Chapters 2 and 3 a T2*-weighted echo planar sequence (TR = 3000ms, TE = 45ms, flip angle = 90°, 36 transverse slices, 3.5 x 2.5 x 2.5mm) was used. Once in the MRI scanner, soft cushions were placed around the participant's head to minimize head movements. Participants completed three runs. The training trials were presented in the first two runs, each of which contained 180 scans. The test trials were presented in the final run of 200 scans. Each run began with five 'dummy' scans before the first trial started. After the two training phases and one test phase, functional scans were completed using a standard volumetric anatomical MRI using a 3-D T1-weighted pulse sequence (TR = 25ms, TE = 4.1ms, flip angle = 30°, 160 axial slices, 1.6 x 0.9 x 0.9mm).

Design and procedure

As in Experiment 9, the design was between-subjects with participants receiving either anodal (n=24) or anodal sham (n=24) stimulation. After the 10 minutes of stimulation (or sham stimulation) was complete, the tDCS headset was removed and the participant entered the MRI scanner where they read the task instructions on a screen

viewed through a mirror on the head coil during the initial MRI survey scan. The same training and test phases as used in Experiment 9 were then presented. Each trial began with a 500-3500ms ITI, a 250ms fixation cross was then presented on the screen which was replaced by a stimulus that remained on the screen for 3 seconds during which time participants made their response. In the training phase feedback was then presented for 500ms before the next trial began; however in the test phase no feedback was presented. In the training phase, each run consisted of two 48 trial blocks separated by an 8 second break. There were 2 blocks of 60 category learning trials in the test phase again separated by an 8 second break.

FMRI data analysis

As in Chapters 2 and 3, data analysis was performed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>). Functional images were corrected for acquisition order, realigned to the mean image, and motion artifacts were corrected for through reslicing. The resliced images were then coregistered with the structural T1 volume. The structural volumes were spatially normalised. The spatial transformation was applied to the T2* volumes which were then spatially smoothed using a Gaussian Kernel of 8mm full-width half maximum. A high-pass filter (128s) was applied to account for low frequency drifts.

Random effect whole brain analyses were conducted using the general linear model with a combined statistical threshold of $p < .001$ (uncorrected) and a threshold of 21 contiguous voxels, which together produce an overall corrected threshold of $p < .05$, according to AlphaSim as implemented in the REST toolbox (Version 1.8; Song et al., 2011). There were two models built: a training model and a test model. Correct trials, incorrect trials, and timeouts were all included as separate regressors for each of the two runs in the training model. In the test model correct trials, incorrect trials and timeouts

were all included as separate regressors for old repeated trials, old non-repeated trials, new repeated trials, new non-repeated trials and prototype trials. As proposed by Grinband et al., (2008) the RT to each stimulus was included as a 'duration' in the model. For both the training and test analyses, the BOLD response was modelled using a canonical hemodynamic response function (HRF) together with temporal and dispersion derivatives, and the six head movement parameters were included as covariates.

The analyses presented first compared correct categorization trials in anodal and sham groups separately to an implicit baseline (the intervals in between the events described above). Additionally, prototype trials were also compared to old and new trials. As in Chapters 2 and 3 conjunction analyses were performed examining areas of common activation in sham and anodal participants. These conjunction analyses used a logical 'and' function through the minimum statistic to the conjunction null hypothesis (MS/CN; Nichols et al., 2005) technique implemented in SPM8. Both of these contrasts were completed with a combined threshold of $p < .001$ (uncorrected) and a cluster threshold of 21. A priori ROI analyses comprising the left DLPFC (the region stimulated by tDCS) were conducted using thresholds of $p = 0.005$, and 10 contiguous voxels. This mask included BA's 6,8,9,10, and 11, as the regions targeted during stimulation. Normalised MNI space coordinates were transformed to Talairach space (<http://imaging.mrcctu.cam.ac.uk/imaging/MniTalairach>) to establish activation sites as defined in the atlas of Talairach and Tournoux (1988).

Results

Behavioural analysis

Figure 6.9 shows the mean accuracy of responding during the training phase (Figure 6.9.a) and during old, new and prototype trials at test (Figure 6.9.b). A one-way

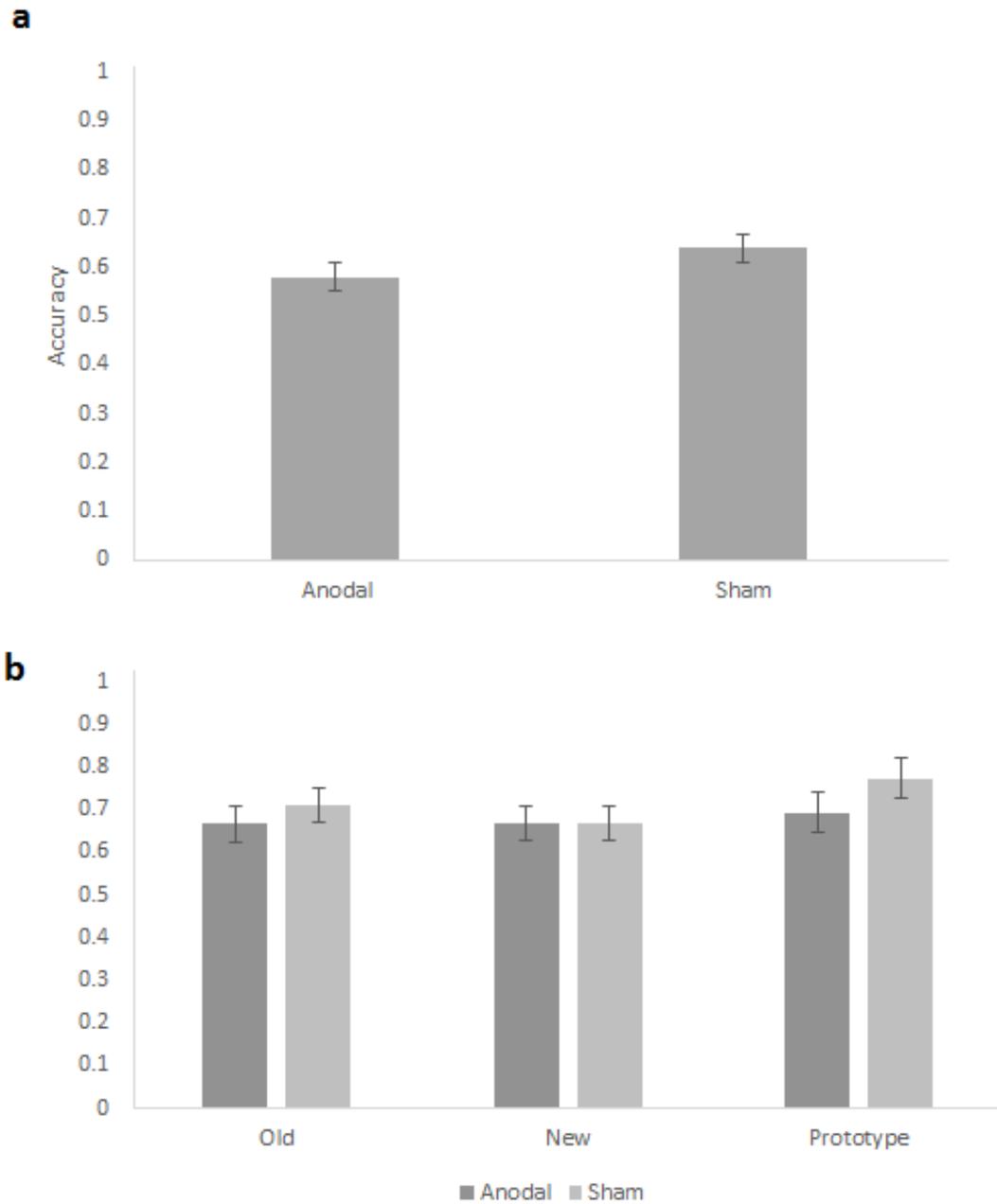


Figure 6.9. Proportion of correct categorization responses in the anodal and sham anodal conditions in Experiment 10; (a) the accuracy of responding during the training phase; (b) the accuracy of responding during old, new and prototype trials during the test phase. Error bars represent standard error.

ANOVA revealed no significant main effect of stimulation condition on training trials ($F(1, 48)=2.13, p=.152, BF=0.58, \eta^2_p=.044$). A mixed-design ANOVA as completed for Experiment 9 revealed no significant main effect of stimulation condition on performance in the test trials ($F(1, 46)=.501, p=.448, \eta^2_p=.011$). There was a significant difference in performance between test trial types ($F(2, 92)=7.42, p=.001, \eta^2_p=.139$) with more accurate responding in prototype trials (Figure 6.9.b). The interaction between trial type and stimulation was non-significant ($F(2, 92)=2.54, p=.085, \eta^2_p=.052$).

As in Experiment 9, in order to determine if repeating trials at test affected response accuracy, a 2x2 ANOVA was completed with accuracy in exemplar trials entered as the dependent variable, and trial repetition (repeated or non-repeated) and trial type (old or new) entered as the independent variables. Mirroring Experiment 9, there was little numerical difference between these conditions (the mean proportion of correct responses was .71 in repeated old trials ($SD=.23$), .67 in old non-repeated trials ($SD=.2$), .67 in new repeated trials ($SD=.2$), and .67 ($SD=.21$) in new non-repeated trials). There was no significant main effect of trial repetition ($F(1,188)=.64, p=.426, \eta^2_p=.003$) or trial type ($F(1,188)=.43, p=.516, \eta^2_p=.002$) on responding, and no interaction between repetition and trial type ($F(1,188)=.41, p=.522, \eta^2_p=.002$)²⁰.

The prototype effect was calculated in the same way as in Experiments 8 and 9. Figure 6.10.a shows the mean difference in accuracy of responding to prototype trials compared to old trials. The prototype effect was significantly bigger than zero in the sham condition ($t(23)=2.37, p=.026, d = 0.65$) but not in the anodal condition ($t(23)=.99, p=.334, BF=0.41, d = 0.3$), and the prototype effects found in the anodal and

²⁰ Average performance in repeated exemplar trials (old and new) and prototype trials were compared in a paired samples t-test. Participants responded to the prototypes more accurately than repeated exemplars ($t(47)=2.02, p=.049, d = 0.19$).

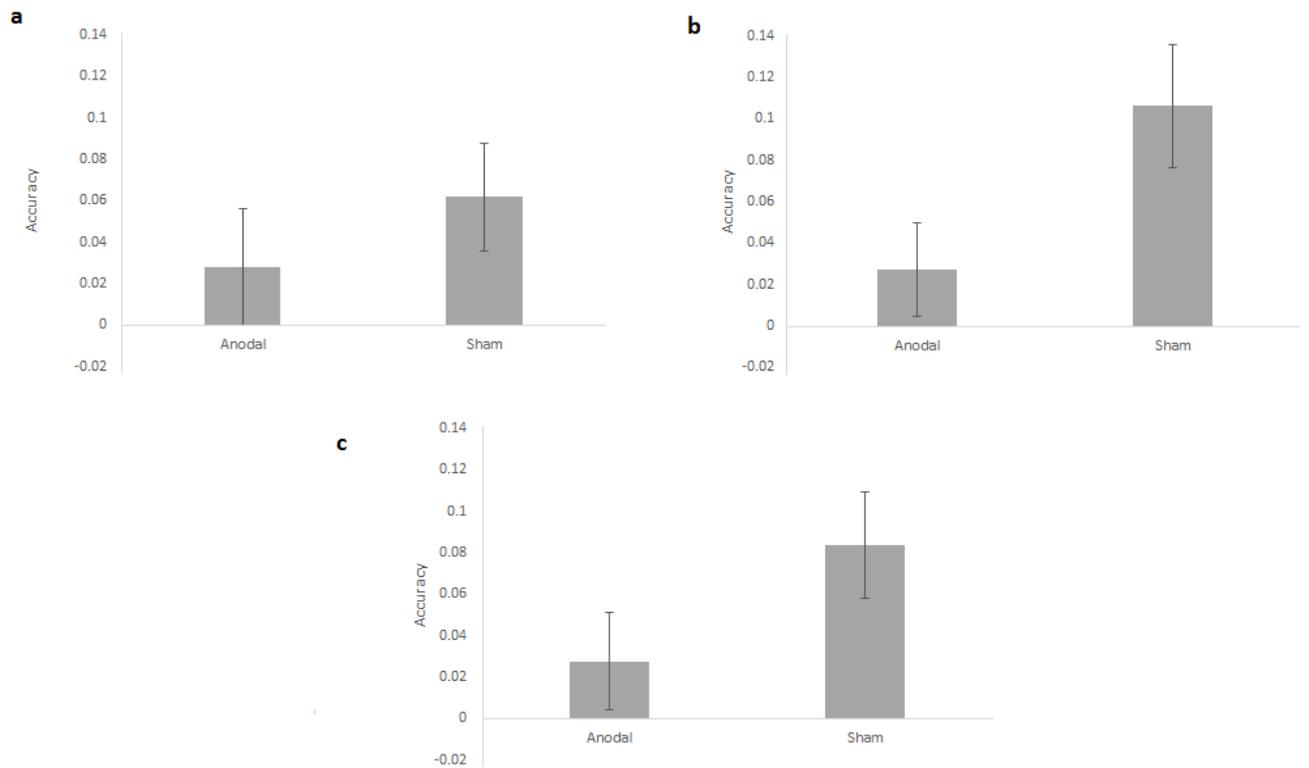


Figure 6.10. The prototype effect in Experiment 10; (a) calculated by subtracting accuracy in old trials from prototype trials; (b) calculated by subtracting accuracy in new trials from prototype trials; (c) calculated by subtracting accuracy in old and new trials from prototype trials. Positive bars represent more accurate performance in prototype trials compared to exemplar trials. Error bars represent standard error.

sham conditions did not differ significantly from each other ($F(1,46)=.77, p=.384, \eta^2_p=0.02, BF=0.01$).

Figure 6.10.b shows the mean difference in accuracy of responding to prototype trials compared with new trials at test. The prototype effect was significantly bigger than zero in the sham condition ($t(23)=3.58, p=.002, d = 1.04$), but not in the anodal condition ($t(23)=1.21, p=.239, BF=0.23, d = 0.39$). The prototype effect in the sham condition was significantly larger than in the anodal condition ($F(1,46)=4.53, p=.039, \eta^2_p=0.06$).

Figure 6.10.c shows the mean difference in the average accuracy of responding to prototype trials compared to all exemplars seen. The prototype effect was significantly bigger than zero in the sham condition ($t(23)=3.3, p=.003, d = 0.87$) but

not in the anodal condition ($t(23)=1.2$, $p=.241$, $BF=0.3$, $d = 0.35$). Although the prototype effects found in the anodal and sham conditions did not differ significantly ($F(1,46)=2.63$, $p=.112$, $\eta^2_p=0.05$, $BF=0.07$), in a one-tailed analysis this result approaches significance ($p=.056$).

Imaging analysis – Training phase

Whole brain activation across both runs of the training phase was first analysed for participants in the anodal and sham conditions separately. Correct categorization compared to the implicit baseline in the anodal condition led to extensive activation in the bilateral inferior and middle occipital gyrus (see Figure 6.11.a). There was also activation in the left inferior and superior parietal lobes, left postcentral gyrus, right parietal lobe, and the left inferior (BA 9 and 47) and middle (BA 9) frontal gyrus. Similar neural activation was observed during the training phase in sham participants (see Figure 6.11.b). The overlap in activation between anodal and sham participants was confirmed in a conjunction analysis looking at the common areas recruited across the anodal and sham contrasts (both with thresholds of $p<.001$ and 21 contiguous voxels; Figure 6.11.c). There was extensive activity in the occipital lobes and activation in the bilateral inferior and superior parietal lobes as well as the left precuneus.

Next, activation in the anodal and sham conditions was directly contrasted. To reveal areas that were more active in anodal participants a subtraction analysis was completed ((correct anodal trials - implicit baseline) - (correct sham trials - implicit baseline)). The ROI analysis found no areas of the DLPFC more activated in anodal participants compared with sham participants. In a whole brain analysis the left putamen (cluster size 44) and right insula (BA 13; cluster size 25) were more engaged in anodal participants compared to sham participants. When analysing areas more active in sham participants compared to anodal participants ((correct sham trials- implicit baseline) –

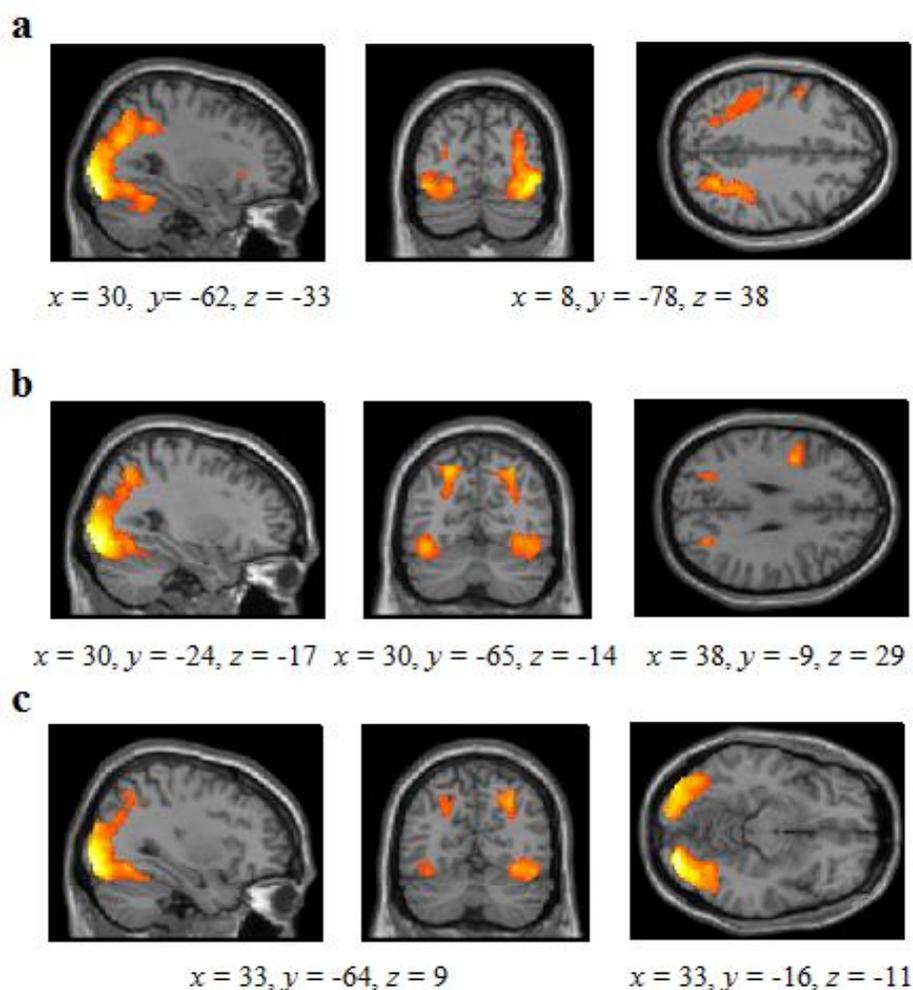


Figure 6.11. Whole brain analysis of the training phase only, thresholded at $p < .001$ and 21 contiguous voxels; (a) areas activated in the anodal condition; (b) areas activated in the sham condition; (c) common areas of activation in the anodal and sham conditions. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

(correct anodal trials – implicit baseline)) the ROI analysis revealed no areas of the DLPFC more engaged in sham participants than anodal participants. However, the left insula (BA 13; cluster size 35) was more activated in sham participants compared to anodal participants in the whole brain analysis.

Imaging analysis – The prototype effect

To examine the prototype effect, a whole brain analysis comparing areas active in the prototype trials to activation during repeated old trials (correct prototype

activation – correct old activation) was completed for anodal and sham participants separately. Neural processing of familiar stimuli differs from that of novel stimuli, for example, the substantia nigra and ventral tegmental area are more implicated when presented with novel visual stimuli than familiar stimuli (Bunzeck & Düzel, 2006). Therefore, non-repeated exemplar trials were excluded from these analyses to control for any novelty effects when comparing the repeated prototype to non-repeated exemplars. There were no areas more active during prototype trials compared with old trials in either condition, or in the sham condition compared with the anodal condition. However, the right occipital lobe (BA 18; peak coordinate $x=18, y=-72, z=5$; 32 contiguous voxels) was more engaged in anodal participants than sham participants. The DLPFC ROI analysis did not yield activation in any of these investigations.

The above analyses were repeated using new trials as a comparison for prototype items. In anodal participants, there were no brain areas more implicated in prototype trials than in new trials. Yet, in sham participants the left caudate body (peak coordinate $x=-18, y=-18, z=27$; 31 contiguous voxels) was recruited more in prototype trials compared to new trials. No areas were more active for this comparison (prototype – new trials) in sham participants than in anodal participants, or in anodal participants compared with sham participants. The DLPFC ROI did not reveal differences in any of these analyses.

Neural areas more active during prototype presentation than in both new and old trials combined were also investigated. This was the highest powered analysis due to the inclusion of all the exemplar trials in the same investigation. In anodal participants the bilateral middle temporal lobes, the left superior occipital gyrus, the right inferior frontal gyrus (BA 47), and the right fusiform gyrus were more active in prototype trials than in exemplar trials (Figure 6.12). However, the DLPFC ROI did not reveal any

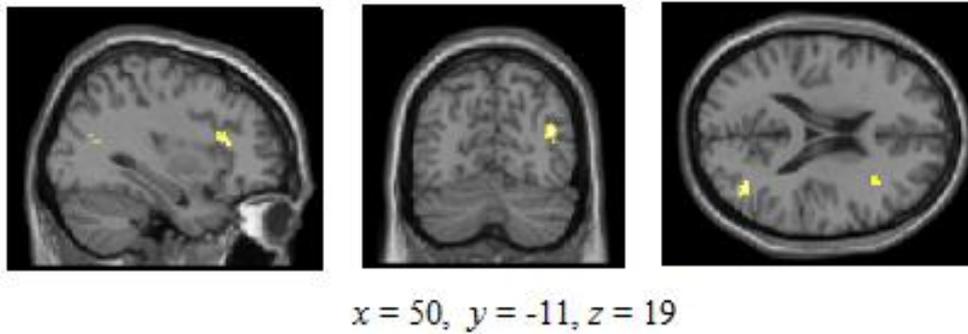


Figure 6.12. Analysis of brain areas more active in prototype trials compared with all repeated exemplar trials in anodal participants thresholded at $p < .001$ and 21 contiguous voxels. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

differences in activation between these trial types. In sham participants there were areas of the left middle occipital gyrus and left middle temporal gyrus activated, but the DLPFC ROI did not find any areas above threshold.

When contrasting areas more implicated in anodal participants than sham participants during prototype trials compared with exemplar trials (see Table 6.1) the engaged areas included the left inferior frontal gyrus (BA 47), the right middle frontal gyrus (BA 47), the left posterior cingulate, the right parahippocampal gyrus (BA 28), middle temporal gyrus, precuneus and the bilateral parietal lobes. The DLPFC ROI analysis revealed that the left BA 11 was significantly more activated (peak coordinate $x = -28, y = 39, z = -5$; 35 contiguous voxels) in anodal participants than in sham participants on this contrast. However, there were no areas more activated in sham participants compared with anodal participants.

Imaging analysis by trial type in the test phase

Analysis of previously unseen prototype trials. In a novel investigation of the neural areas engaged during prototype presentation, whole brain activation during correct prototype trials was compared to the implicit baseline for participants in the anodal and sham conditions separately. Correct categorization in the anodal condition

Table 6.1

Whole Brain Analysis of Areas More Activated in Correct Prototype Trials Compared with Exemplar Trials: A Comparison of Areas More Engaged in Anodal Participants Compared with Sham Participants on This Measure

Region	Cluster size	BA	Talairach Coordinates			z-score
			x	y	z	
Left Inferior Frontal Gyrus	90	47	-40	21	-1	4.21
Left Posterior Cingulate	217	29	-10	-46	19	4.20
Left Lingual Gyrus		18	-10	-54	3	3.87
Left Cingulate Gyrus		31	-14	-43	26	3.42
Left Insula	31	13	-30	-20	21	4.20
Left Superior Occipital Gyrus	21	19	-34	-80	32	4.00
Left Middle Temporal gyrus	24	-	-46	-43	-1	3.93
Right Middle Frontal Gyrus	25	47	38	37	-7	3.80
Right Middle Temporal Gyrus	24	19	36	-61	16	3.79
Right Lingual Gyrus	29	18	18	-70	4	3.78
Right Precuneus	57	31	16	-63	20	3.72
Right Parahippocampal Gyrus	27	28	18	-15	-20	3.65

Note that indented rows indicate voxels in the same cluster as the non-indented row above

led to activation in the bilateral inferior and middle occipital gyrus as well as in the bilateral superior and left inferior parietal lobes. There was also activation in the bilateral middle (BA 39 and left BA 37), and inferior (BA 37) temporal lobes during correct responding to prototype trials in anodal participants (see Figure 6.13.a).

A similar pattern of recruitment was found in sham participants (see Figure 6.13.b) with activation again present in the bilateral middle and left inferior occipital gyrus, the right superior parietal lobe, the bilateral inferior temporal gyrus, the bilateral precuneus, and the left fusiform gyrus. However, there were no areas more activated in anodal participants when compared with sham participants, or in sham participants compared with anodal participants when responding to prototype trials. The ROI analysis revealed no activation in any of these contrasts.

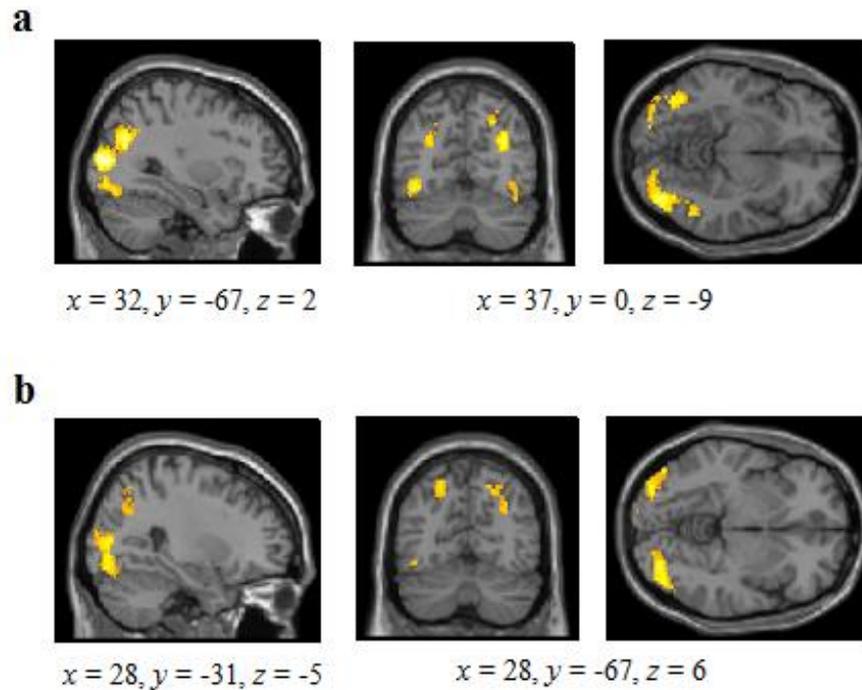


Figure 6.13. Whole brain analysis of correct prototype trials thresholded at $p < .001$ and 21 contiguous voxels; (a) areas activated in the anodal condition; (b) areas activated in the sham condition. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

Analysis of previously seen exemplar trials (old). Whole brain activation in correct old trials in the test phase (repeated and non-repeated) was compared to an implicit baseline for participants in the anodal and sham conditions separately. Correct categorization in the anodal condition (see Figure 6.14.a) led to extensive activation in the bilateral occipital lobes, as well as in the right fusiform gyrus, the left postcentral gyrus, and left precuneus. This analysis was repeated for sham participants (see Figure 6.14.b). The bilateral middle and right superior occipital gyrus was activated along with the bilateral fusiform gyrus and left postcentral gyrus.

Activation in anodal and sham participants during old trials was directly compared. The DLPFC ROI revealed no areas were recruited more in sham participants than in anodal participants. However, in the whole brain analysis the left superior frontal lobe and precentral gyrus were more active in sham participants compared with

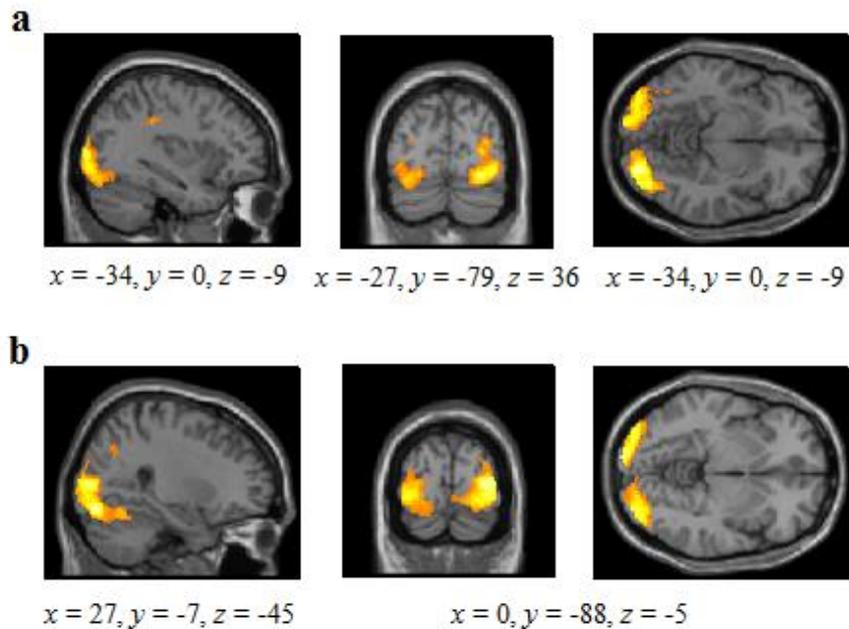


Figure 6.14. Whole brain analysis of correct old trials in the test phase only, thresholded at $p < .001$ and 21 contiguous voxels; (a) areas activated in the anodal condition; (b) areas activated in the sham condition. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

anodal participants (see Table 6.2). When reversing this analysis and subtracting activation in the sham participants from the anodal participants, there were no areas more engaged in anodal participants compared with sham participants, and the DLPFC ROI analysis did not reveal any regions of activity.

Analysis of previously unseen trials (new). To analyse areas engaged during correct new trials (repeated and non-repeated), a whole brain analysis compared activation to an implicit baseline for participants in the anodal and sham conditions separately. Correct categorization in the anodal condition (see Figure 6.15.a) again revealed activation in the bilateral middle and inferior occipital gyrus, the left inferior parietal lobe, the right superior parietal lobe, and the left postcentral gyrus. This analysis was also completed for sham participants (see Figure 6.15.b). Again there was activation in the bilateral middle and left inferior occipital gyrus, the left inferior parietal lobes, the right superior parietal lobule, and the left postcentral gyrus.

Table 6.2.

Whole Brain Analysis of Areas More Activated in the Sham Condition Compared to the Anodal Condition in Correct Old Trials During the Test Phase

Region	Cluster size	BA	Talairach Coordinates			z-score
			x	y	z	
Left Superior Frontal Gyrus	76	10	-14	56	-3	4.72
Left Precentral Gyrus	33	6	-38	-11	60	3.77

When directly comparing activation between anodal and sham participants, the left hippocampus/parahippocampal gyrus (BA 37, cluster size 25) was more engaged in anodal participants compared to sham participants, and no areas were active above threshold in the DLPFC ROI analysis. However, when analysing neural areas more active in sham participants compared to anodal participants, the ROI analysis found that areas of left precentral gyrus (BA 6; peak coordinate $x=-14, y=-9, z=59$; 75 contiguous voxels) were present. In the corresponding whole brain analysis areas of the right inferior and middle frontal gyrus (BA's 47 and 11, 127 contiguous voxels) and the left

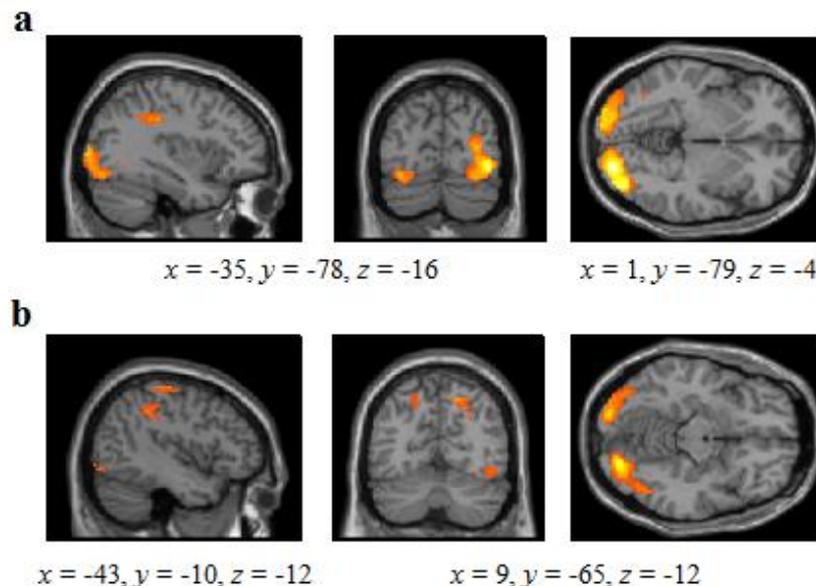


Figure 6.15. Whole brain analysis of correct new trials, thresholded at $p < .001$ and 21 contiguous voxels; (a) areas activated in the anodal condition; (b) areas activated in the sham condition. The coordinates indicate the origin for the image displayed. Lighter colours indicate higher z-scores.

precentral gyrus (BA 6, 42 contiguous voxels) were more active in sham participants compared with anodal participants, corresponding with the neural activation found in the analysis of old trials.

Discussion

In Experiment 10 anodal stimulation once again abolished a behavioural prototype effect that was present in sham participants. Also, there was no significant effect of stimulation type on overall performance on the training or test trials. This offers a further replication of Experiments 8 and 9 and is consistent with Ambrus et al. (2011) who found tDCS inhibited prototype effects. Yet, Ambrus et al. also reported that anodal stimulation lowers overall performance on the prototype task. As discussed, the procedural differences between Ambrus et al. and the present study (e.g., the number of categories to be learnt, the type of stimuli) could account for this discrepancy. This finding is also contrary to McLaren et al. (2016) who found anodal stimulation lowered performance on new (Experiment 2a) and old (Experiment 2b) trials at test. Again, this difference is likely attributable to the fact that participants in McLaren et al. (2016, Experiment 2) only learnt two categories, with participants performing numerically better on the exemplar trials than those in the present chapter, which may have masked a potential prototype effect. When calculating the neural correlates of the prototype effect using all exemplar trials (prototype trials – old and new trials), the bilateral frontal lobes (the left inferior and right middle frontal lobes), the left DLPFC (in a ROI analysis), and the right middle temporal lobe, including the parahippocampal gyrus, were more activated in anodal participants than in sham participants. Yet, in no analysis of the prototype effect was there evidence of brain areas more implicated in sham participants compared with anodal participants.

The occipital and parietal lobes were recruited by anodal and sham participants during the training phase and all tests trials. These lobes were also active more during prototype presentation compared with exemplar presentation in anodal and sham participants. This supports previous neuroimaging research that has implicated these areas in prototype distortion tasks (e.g., Seger et al., 2000; Zeithamova et al., 2008).

As well as being active in the training phase in both anodal and sham conditions, the frontal lobes were more engaged by sham participants compared with anodal participants during old trials (left superior frontal lobe, BA 10; and precentral gyrus, BA 6) and new trials (left precentral gyrus, BA 6; right inferior frontal lobe, BA 47; and middle frontal lobe, BA 11). This frontal lobe activation is distinct from that found in analyses of the prototype effect where it is anodal participants rather than sham participants that engage the left inferior and right middle frontal lobes (both BA 47) more. Conversely, in responding to new stimuli only the left hippocampus and parahippocampal gyrus were more engaged in anodal participants when compared with sham participants. However, analyses of the anodal participants reveal no neural areas more active compared with sham participants on old or prototype trials at test. The theoretical implications of these results will now be considered in the general discussion.

General Discussion

The experiments presented in this chapter examine how tDCS to the left DLPFC influences the behavioural prototype effect, as well as the neural correlates that underlie this effect. While overall learning was equal in both anodal and sham participants, the experiments presented in this chapter demonstrate the abolition of a prototype effect in anodal participants that is consistently present in sham participants. These findings support the results reported by Ambrus et al. (2011) who found the same effect after

anodal tDCS to the DLPFC. This is a novel contribution to the research field as it demonstrates that the prototype effect can be suppressed by anodal stimulation in (A, B) tasks as well as in the (A, not A) task used by Ambrus et al. (2011). Zeithamova et al. (2008) suggested that these two tasks implicate different neural areas, so finding that tDCS reduces the prototype effect in both studies suggests that the core neural areas underlying behavioural performance in prototype distortion tasks may be similar. Future fMRI research could explore this possibility. The current experiments also suggest that the abolition of the prototype effect in Ambrus et al. is not due to the majority of stimulation occurring prior to task performance. Experiment 8 gave the majority of tDCS during task performance, whereas stimulation was completed prior to any training in Experiments 9 and 10, yet the overall pattern of behavioural results remained broadly unchanged.

The finding that the prototype effect present in sham participants is abolished in anodal participants contradicts the results of McLaren et al. (2016, Experiment 2), who found that anodal stimulation to the left DLPFC in a two category prototype distortion task induced a prototype effect that was not present in sham or cathodal participants. As previously discussed, a potential ceiling effect in McLaren et al. (2016, Experiment 2) could explain why prototype effects were not found in their sham and cathodal conditions. The lowering of overall performance in the current chapter by introducing a third prototype derived category could account for the presence of the prototype effect in sham and cathodal conditions. However, the increased complexity of the task does not explain why a prototype effect is observed under anodal stimulation in McLaren et al. but not in the present experiment.

Yet, while seemingly contradictory, the results of the present chapter appear compatible with the findings reported in McLaren et al.'s Experiment 2 (2016). The

tDCS procedures and visual stimuli in McLaren et al. and Chapter 6 were the same, so the difference in the behavioural patterns found is likely due to the addition of an extra category in the present work. The McLaren, Kaye, and Mackintosh (1989; MKM) model suggests that when participants are repeatedly presented with visual stimuli, perceptual learning enhances their ability to learn to discriminate between the stimuli shown, meaning they can more easily spot differences between stimuli that look very similar, and are less likely to confuse them. Perceptually similar stimuli contain many common elements (or features), for example, a certain cluster of black squares present in a chequerboard stimulus. These common features between the experienced stimuli become predictable and less salient with repeated presentation, a process often referred to as latent inhibition. However, features that are unique to a stimulus, and therefore novel to the participant at presentation, become more salient. This makes discrimination between stimuli easier as the unique features between them become more apparent.

In the three category learning task there will be heavy feature overlap between conditions as each of the three categories share some features with one or both of the other categories. When prototype stimuli are experienced after training, there will consequently be few salient features separating them from each other, as the common items of all categories have been inhibited, and are not salient. This means there are fewer salient features between the prototypes that participants must attend to, making the prototypes easy to distinguish from each other. As the prototypes are easier to discriminate following perceptual learning, they are less likely to be incorrectly categorized, as there will be no salient features that overlap between the categories. Similarly, in the three category task, the exemplars have a higher chance of being incorrectly categorized than the prototypes, as there are more previously unseen and therefore salient visual features in the exemplars to pay attention to, so they are less

discriminable from each other than the prototypes. This, therefore, leads to the propensity for a prototype effect.

However, in the two category prototype distortion task, there are fewer common features in the prototypes, as there is one less category that has features that could overlap with stimuli from opposing categories. There are, therefore, more salient, novel features in the two category prototypes, meaning they are less distinct from each other than in the three category problem. This means that the amount of latent inhibition in the two category task will be less than in the three category task, which reduces participants ability to discriminate easily between the category prototypes. In other words, because not only the exemplars but also the prototypes have a large number of unique features, this leaves little propensity for the presence of a prototype effect.

Anodal stimulation, however, is found to reverse the behavioural prototype effect pattern found in the two and three category learning task. One possibility is that that anodal stimulation to the DLPFC disrupts latent inhibition which consequently reduces perceptual learning. According to the MKM account (McLaren et al., 1989) the relative salience of the units is driven by an error correcting process (the modulator illustrated in Figure 6.16.a) and it is this mechanism that provokes latent inhibition following exposure to the stimuli. If anodal tDCS disrupts the modulatory input, though, the MKM model, broadly speaking, becomes analogous to the McClelland and Rumelhart (1985) model (shown in Figure 6.16.b). According to this account, it is the common features (or elements) of a stimulus that are the most salient and the unique features that are the less salient (in other words the opposite pattern to that predicted by the MKM model). This greater saliency of common features therefore increases generalisation between items.

This image has been removed by the author of this thesis for copyright reasons.

Figure 6.16. (a) The modulator in this diagram represents prediction error (McLaren et al., 1989), and how it increases the salience of features in a single category (each ring represents a span of visual features). The temperature diagram shows that more predictable stimulus features become less active (darker shading) which leads to latent inhibition and increased salience of unique features in stimuli; (b) Prediction error is removed which reverses this effect (McClelland & Rumelhart, 1985). Common features between stimuli now become more salient, increasing learning by associative processes (lighter shading). Figure taken from McLaren et al. (2016).

How though does this purported increase in generalisation and reduction in perceptual learning accommodate the discrepant results in the two and three category tasks? One clue to this may be in the results of the control conditions. In the sham condition of the three category procedure I found a significant prototype effect, while in McLaren et al.'s two category procedure sham stimulation did not evoke a prototype effect. It appears plausible that this reflects the relative dominance of the aforementioned generalisation and perceptual learning processes. In the three category task there is likely to be increased levels of generalisation as each category can have common features with two rather than just one other category which would be expected to make category discrimination more difficult. However, and critically, as mentioned above, it would also be accompanied by high levels of perceptual learning too and it is the strength of this perceptual learning that drives the prototype effect. In contrast, in

the two category task, the generalisation process may have more influence than the more modest levels of perceptual learning which means that the prototypes will not have a discriminatory advantage over the exemplars.

One possibility then is that anodal tDCS reduces perceptual learning and increases generalisation both within and between-categories. Within-category generalisation should enhance any prototype effect but between-category generalisation would be expected to diminish it. A key factor may therefore be that the amount of between-category generalisation is likely to be significantly smaller in the two category task than the three category task (where there is an extra category for overlap to occur). This means that in the two category task, within-category generalisation is likely to be greater than between-category generalisation leading to a prototype effect, whilst for the three category task between-category generalisation is likely to dominate, leading to the attenuation, or even elimination, of the prototype effect. For a more detailed explanation of this theory and supporting simulations, please see McLaren et al. (2016).

This explanation can also be linked to the most notable result of Experiment 10 - the greater activation in the left DLPFC (and the left inferior frontal gyrus) during prototype categorization compared to all exemplar trials in anodal participants compared with sham participants. The greater activation in these regions could plausibly reflect the inhibition or disruption of the modulator in the MKM model that is assumed to drive the perceptual learning. It would clearly be of value in future work to use transcranial magnetic stimulation (TMS), which has better spatial resolution than tDCS, to examine if disruption of these regions does indeed produce a similar pattern to the results I observed in this chapter.

In anodal participants the right parahippocampal gyrus and middle temporal gyrus were also more active than in sham participants during prototype presentation

compared with all exemplar trials at test. Nosofsky et al., (2012) suggested that the MTL could be implicated in calculating the summed similarity of all seen exemplars and extracting the prototypical visual features experienced. McLaren et al. (2016) suggest that tDCS to the left DLPFC reduces latent inhibition leading participants to generalise features of visual stimuli they experience. Increased MTL engagement could represent an extraction of common (prototypical) features between stimuli – increasing the salience of mutual features. Not only could this explain the reduced prototype effect as discussed previously, but also offers an explanation of the greater MTL activation in anodal participants compared with sham participants. According to this account, one might predict that there would be significant effective connectivity between the MTL and the DLPFC that is modulated by anodal tDCS. This prediction could be examined in future work using dynamic causal modelling (Friston, Harrison, & Penny, 2003).

In the past, research has implicated the occipital and parietal lobes in (A, B) prototype distortion tasks (e.g., Seger et al., 2000; Zeithamova et al., 2008); however, prior to Experiment 10 the research field had not investigated neural activation correlated purely with prototype presentation. Experiment 10 aimed to rectify this, and the parietal lobes were found to be more active in prototype trials than at baseline, allowing us to extend the findings of past work (e.g., Zeithamova et al., 2008) to responding specifically to prototypical stimuli. The parietal lobe activation found in prototypes trials can also be considered somewhat related to past work which has found this region specifically implicated in stimulus generalisation (Seger, Braunlich, Wehe, & Liu, 2015). Accordingly, given that the prototype trials require generalisation from previously experienced exemplars, this process appears a plausible candidate for the parietal lobes activation observed.

Another region that fMRI studies suggest is commonly active in prototype distortion tasks is the occipital lobe. For example, Seger et al. (2000) suggested that during prototype distortion learning, the occipital lobes perform a visual feature analysis of each perceived stimulus to judge category membership. The present chapter extends this research, as it can be more directly inferred that the occipital lobes are more active in response to prototype trials than exemplar trials. Considering this suggestion, it could be that activation in the occipital lobes is stronger for prototype trials as these stimuli would share the maximum common features with previously seen exemplars, meaning the visual feature analysis in the occipital lobes would recognise more of the previously experienced features than in exemplar trials.

The results presented in Chapter 6 may offer support for prototype theories of category representation (e.g., Minda & Smith, 2001; 2002; Smith, et al., 2008) over exemplar theories (e.g., Nosofsky, 1985; 1986). This work suggests that anodal stimulation selectively inhibits the prototype effect and not overall performance on exemplar trials, and that prototypes are dealt with differently in the brain compared to exemplar stimuli. This advocates that prototypes are processed qualitatively differently compared with category exemplars.

In conclusion, the present chapter consistently finds that anodal stimulation reduces a prototype effect that is present in sham participants. These findings resonate with and extend the work of Ambrus et al. (2011) as the reduction in performance is selective to prototype trial responses. The present chapter implemented a novel tDCS and fMRI design to examine neural activation during prototype presentation, an investigation previously overlooked in the literature. When considering the neural effect of anodal tDCS on the prototype effect, it was found that the MTL and the DLPFC were more activated in anodal participants compared with sham participants. This pattern of

results is compatible with the idea that anodal stimulation to the left DLPFC disrupts the modulator component of the MKM model, which attenuates perceptual learning. This consequently reduces the distinctiveness of the prototypes relative to exemplar items leading to the loss of the prototype effect.

Chapter 7. General Discussion

The present thesis was separated into two strands. Strand one (Chapters 2 – 5) investigated the neurobiological underpinnings of category learning with reference to the separable explicit and implicit learning systems proposed by COVIS (Ashby et al., 1998), as well as critically exploring several behavioural dissociations taken to support the COVIS model. Experiment 1 (Chapter 2) investigated if non-essential differences between category structures predicted to preferentially engage either the explicit or implicit learning system (RB and II tasks respectively) could account for the neural dissociations during decision making processes previously reported in the literature that support COVIS (Nomura et al., 2007). Chapter 3 extended the results of Chapter 2 by examining the neural correlates of feedback processing in RB and II learning. Experiment 2 first assessed if delaying feedback in the procedure of Experiment 1, necessary for fMRI analysis of the feedback processing stage, would differentially affect learning in the II or the RB condition. Experiment 3 then implemented this procedure in an fMRI study to examine if neural dissociations were apparent during feedback processing in RB and II learning. Behavioural predictions of COVIS were examined in Chapter 4. Experiment 4 was designed to replicate the dissociation reported by Smith et al. (2014), where II learning was impaired by deferring feedback whereas RB learning was not. Experiments 5 and 6 aimed to generalise this effect to a procedure that used conditions equated on task difficulty and the number of stimulus dimensions relevant for categorization. Finally, Experiment 7 in Chapter 5 was the first investigation of the effect of tDCS on RB and II category learning performance.

The second strand of this thesis (Chapter 6) investigated the effect of tDCS on the presence of the prototype effect in a prototype distortion task. This strand built on

previous research from McLaren et al. (2016; Experiment 2) which demonstrated anodal tDCS to the left DLPFC induced a prototype effect in participants that was not present in sham and cathodal participants. However, there was a potential ceiling effect present in McLaren et al.'s work, so Experiments 8 and 9 investigated whether this finding generalised to a more difficult categorization task that included an extra prototype derived category. Experiment 10 implemented a joint tDCS and fMRI method never previously used to investigate the neural correlates of the prototype effect or the brain regions engaged in prototype presentation. The main findings of each chapter will now be discussed and the implications of each experiment, as well as potential future research, will be considered.

Chapter 2 (Experiment 1)

Previous neuroimaging research into the neural coordinates of RB and II learning found evidence of dissociable neural systems, where RB learning implicates regions of the frontal lobes and the MTL, and II learning relies more on the substantia nigra and the caudate body and tail (e.g., Nomura et al., 2007; Nomura & Reber, 2008). However, the unidimensional RB condition used by Nomura et al. included non-essential differences to the II condition, as there was only one stimulus dimension relevant for learning compared with the two that required attention in the II condition. Also, the RB stimuli were closer to the decision boundary than in the II structure. When using conjunctive RB and II category structures matched for difficulty, the number of relevant dimensions for learning, and category separation, there was extensive neural overlap between the two learning conditions, including brain regions associated with the explicit system of COVIS such as the MTL, the frontal lobes and the caudate head, as well as areas associated with the implicit system including the caudate body and tail, the putamen and the substantia nigra. Critically, the results indicated that the MTL, an area associated

with the explicit system, was more active in II participants than RB participants, a dissociation that is maintained when examining only the second half of training, and when including only participants who used the optimal decision boundary to learn according to the GRT modelling analyses.

COVIS could account for the extensive activation overlap in the present study by assuming that it represents both the explicit and the implicit systems being simultaneously engaged during category learning trials, with the non-dominant system not involved in the decision (e.g., Ashby & Crossley, 2010). However, such an explanation cannot account for the key finding of Experiment 1 that the MTL, an area implicated in the explicit system of COVIS, is more active in II participants compared with RB participants. If both systems were simultaneously active in RB and II trials, then this explanation would not anticipate the MTL to be more active in II conditions, as it is part of the suboptimal system for that category structure. This result is also contrary to previous research that implicates the MTL more in RB learning compared with II learning (Nomura et al., 2007).

It seems likely that it is the difference in the RB tasks employed in the present work and Nomura et al. (2007) which is driving the discrepancy in results. While Nomura et al. used a unidimensional RB structure, I used a multidimensional conjunctive structure. Previous behavioural research has suggested that controlling for extraneous variables like the number of dimensions relevant for learning - which I was able to do by using a conjunctive RB task - offer alternative explanations for the dissociations supporting COVIS in the literature, such as both RB and II learning being the result of a single explicit system with the easier to learn RB task (such as the one used by Nomura et al.) being less susceptible to certain behavioural manipulations (e.g.,

Edmunds et al., 2015). The present study also extends these concerns to neuroimaging dissociations.

A single system could potentially account for the greater MTL engagement in II conditions compared with RB conditions due to the increased demand on cognitive processes when performing the more complex II task to the same standard as the RB task. This result can be similarly accounted for by a dual process model that relies on two explicit systems of category learning such as ATRIUM (Erickson & Kruschke, 1998). Here, for example, while the RB category decision boundary could be optimally learnt by a rule-based system, performance in the II condition could be supplemented by exemplar storage. Davis et al. (2012a) similarly provide a viewpoint of the data that fits with Experiment 1, but not the assumptions of COVIS. In light of Davis et al., one could predict that the II condition requires stimuli to be stored in numerous clusters. As there are more exceptional stimuli to a rule in II conditions, participants must continually generate new clusters in which to store stimuli, a processes required much less in RB conditions whereby there are less exceptions. This could account for the increased MTL activation in the II condition if this area were implicated in generating novel clusters as well as retrieving information from multiple clusters.

Chapter 3 (Experiments 2 and 3)

The design of Experiment 1 did not enable an examination of what brain areas are specific to stimulus responding and what regions are involved in feedback processing. This appears an important question to answer given that COVIS predicts that there are large differences between feedback processes in the explicit and implicit systems. One possibility therefore is that there is overlap in the decision processes of RB and II learning but dissociable activation at the feedback stage. Chapter 3 investigated this

possibility. To do this, alterations to the method of Experiment 1 were necessary in order to temporally disentangle the decision and feedback neural responses.

First, Experiment 2 investigated whether delaying feedback differentially impaired II performance compared to RB performance. This feedback manipulation was a necessity for fMRI analysis of activation during feedback processing. Previous research suggests that delaying feedback hinders performance in participants learning II category structures in comparison with II participants receiving immediate feedback, an effect not present in RB participants (e.g., Maddox et al., 2003; Maddox & Ing, 2005). In Experiment 2, delaying feedback presentation by up to 1500ms did not differentially impair performance in conjunctive RB and II category learning conditions. However, the feedback delay period used in Experiment 2 was at least 1 second shorter than that used by Maddox et al. and 3.5 seconds shorter than that of Maddox and Ing, meaning that the present results cannot be directly compared to these previous experiments. Yet, with this caveat in mind, the findings of Experiment 2 are not consistent with what COVIS would predict, offering support for the conclusions of Experiment 1 that controlling for potential methodological confounds between RB and II conditions leads to no dissociation in performance.

This finding could support conclusions in line with Dunn et al. (2011) who suggest that methodological artefacts in studies such as Maddox and Ing (2005) can selectively impair performance in II participants. As previously discussed, the use of a less complex unidimensional RB structure as a comparison for II performance in Maddox et al. (2003) may have reduced the impact delaying feedback could have on RB participants because their learning system is less taxed, as only one stimulus dimension must be learnt compared with the multidimensional II condition, making RB learning less susceptible to disruption by suboptimal feedback conditions (e.g., Edmunds et al.,

2015). While Maddox and Ing (2005) controlled for the number of relevant stimulus dimensions in their study, the mask presented during the delay between stimulus and feedback presentation was very similar to the categorization stimuli, which could have disadvantaged II participants who likely had to engage more cognitive processes compared to RB participants, making them more susceptible to interference from the visually similar mask during learning (Dunn et al., 2011). While converging with Dunn et al., Experiment 2 also lays the ground work for Experiment 3. As delaying feedback by up to 1500ms did not selectively disadvantage II learning compared to RB learning, I considered that implementing such a delay in an fMRI procedure to investigate feedback processing would not change the learning systems used in the RB and II conditions of Experiment 1.

In an fMRI study (Experiment 3) there were extensive areas of activation overlap between RB and II participants during feedback processing, including regions implicated in the explicit (e.g., the MTL, the caudate head) and the implicit (e.g., the putamen and the caudate body) systems proposed by COVIS. Interestingly, in RB participants the MTL was more activated in learning session 1 than session 2, whereas in II participants the MTL was more engaged in session 2 than session 1. However, as the interaction between category structure and learning session was not significant, these results should be interpreted cautiously.

Finding extensive neural activation overlap between RB and II participants confirms and extends the results of Experiment 1 that RB and II participants not only use similar neural pathways during category decision making (Experiment 1) but also during feedback processing. As in Experiment 1, the MTL was more active in II participants than RB participants, a pattern not in line with COVIS. This is further evidence for my inference that the MTL activation found in Nomura et al. (2007)

resulted from the use of a unidimensional RB category that had a perceptually less discriminable decision boundary than the II condition, and extends this conclusion to activation during feedback processing.

While the spatial resolution of fMRI is good, it is not possible to establish whether all of the diverse areas shown to be activated are necessary for the learning of the II and RB structures. For example, it is possible that the MTL activation was not essential for category learning. While this is plausible, it is still difficult, though, to explain from this perspective why the MTL activates more for the II condition than the RB condition; in contrast, this difference is readily compatible with the idea that II learning requires greater memory demands to compensate for the absence of an easily verbalizable rule. Nevertheless, to test this explanation for the present result and that of Experiment 1, it would be valuable in the future to investigate patients with MTL lesions to see if they, as would be predicted from this theory, perform worse in acquiring II categories than RB categories. Similarly, II participants in Experiments 1 and 3 may learn by storing more individual exemplars in memory compared with RB participants, who may instead store verbal rules. If II participants did spend more time memorising the individual exemplars than RB participants who could more easily rely on an abstract rule, then when presented with previously seen stimuli, recognition of old exemplars in the II condition should be better than in the RB condition. This appears a fruitful avenue of future research.

Chapter 4 (Experiments 4, 5 and 6)

Smith et al. (2014) found that II participants were impaired by deferring feedback to the end of a six trial block, whereas RB participants were unimpeded. Smith et al. concluded that this finding offered some of the strongest evidence for COVIS as the

results could not be accommodated by a single system theory. Experiment 4 replicated this finding as II participants receiving deferred feedback performed less accurately than II participants receiving immediate feedback after every trial, whilst there was no significant effect of feedback deferral on RB learning. However, Experiment 4 also found that II participants performed worse in general compared to RB participants regardless of the feedback condition. I considered that this difference in overall learning between unidimensional RB and II participants suggested a difficulty confound between the two category structures, with unidimensional conditions being generally easier to acquire than the multidimensional II category structure.

The difficulty confound between unidimensional and II category structures implied by Experiment 4 (as well as Chapters 2 and 3, and previous research, e.g., Edmunds et al., 2015) suggests that II and conjunctive RB structures, which both have two relevant dimensions, place greater cognitive demands on processes such as WM compared to unidimensional structures, (e.g., Milton et al., 2008; Wills et al., 2015). Therefore, unidimensional categories may be more resilient to manipulations that make feedback suboptimal, such as deferring feedback, than II and conjunctive conditions where the learning process is already taxed in comparison. Similarly, unidimensional structures can also be accurately learnt in the absence of feedback, meaning that unidimensional RB participants will likely be able to deal with the non-specific feedback of deferred conditions, whereby the feedback is not presented in the same order as trial presentation (Ashby et al., 1999). This may, again, make unidimensional categories less vulnerable to suboptimal feedback.

These arguments are supported by the findings of the remaining two experiments in Chapter 4. Experiment 5 found that when matching the number of dimensions relevant for category learning in the RB and II conditions, the dissociation

presented by Smith et al. (2014) vanished. Instead, when conjunctive RB categories were used as a comparison for II learning, both RB and II participants were impaired by deferring feedback. This was further confirmed by Experiment 6 that also found conjunctive RB participants hindered in the deferred feedback condition compared to the immediate feedback condition when doubling the number of participants. These findings support the conclusion that a discrepancy in difficulty between unidimensional RB and II conditions is responsible for the dissociation reported by Smith et al. (2014) and Experiment 4.

In the future, studies could be conducted that support the inferences made by Chapter 4. While it seems likely that the switch from a unidimensional to a conjunctive RB category is causing the contrasting results between Experiment 4, and Experiments 5 and 6, the unidimensional category structure used in Experiment 4 also employed different visual stimuli to those used in the conjunctive category (i.e. dot patterns versus single lines). Whilst it is unclear why such a manipulation would cause the difference in results between the experiments, and certainly it is not something predicted by COVIS, it is possible that this difference in stimuli may have influenced the present results. In future research, the dot pattern stimuli could be separated into two categories by a conjunctive RB decision boundary as opposed to a unidimensional decision boundary (e.g., see Figure 4.4). This could then be used as a comparison for performance in the II condition of Smith et al. (2014) that is matched on the number of relevant stimuli dimensions and relative difficulty. According to the account that I have advocated throughout this thesis, there should be no dissociation between the II and conjunctive RB conditions and learning of both category structures should be impaired by deferred feedback.

Chapter 5 (Experiment 7)

Experiment 7 investigated whether learning accuracy in conjunctive RB and II categories (Filoteo et al., 2010) was affected by anodal tDCS to the left DLPFC. The findings of Experiment 7 indicate that while overall RB and II learning was unaffected by anodal tDCS, there was a significant interaction between stimulation condition and performance on blocks performed *during* tDCS, where performance increased more across the first three blocks in the sham participants than in the anodal participants. This research is somewhat contrary to past work using the weather prediction category learning task where anodal stimulation enhances learning (Kincses et al., 2003; Vercammen et al., 2011). These discrepancies in results are likely due to the very different task types implemented, which vary on the number of stimuli presented, stimuli type and probability of category membership. They also underscore the fact that there is currently a paucity of relevant work from which to draw strong inferences in the effect that tDCS has on category learning. Nevertheless, the present results indicate that the effect of anodal tDCS on category learning may be modulated by the task employed.

Statistically, the present findings suggest that learning in RB and II conditions was only affected by tDCS during stimulation. However, it should also be considered that the experiments of Chapter 6 report an effect of tDCS after stimulation has ceased, suggesting that the durability of tDCS on learning may be task dependent. Findings of behavioural effects after tDCS in Chapter 6 suggest that the potential interaction visible in Figure 5.5 may be the result of anodal tDCS. Future research must investigate this possibility before commenting on how transient the effects of tDCS are. Increasing the duration and intensity of tDCS, or doubling the number of participants in the study, could be used to study this possibility.

The qualitative differences in the strength of the tDCS effect between Chapters 5 and 6 suggest that before any novel learning task is presented to individuals in training environments, pilot studies in laboratory settings should be completed to ensure whether anodal stimulation increases or decreases task performance, and decide the optimal duration of tDCS required during the task.

An interesting avenue of research to pursue would be the effect of anodal stimulation on unidimensional RB compared to II learning. Chapter 4 would predict that dissociations between RB and II performance are more likely when non-essential differences between conditions are ignored. If it is found that unidimensional RB and II performance is differentiated by tDCS, there is further evidence that using unidimensional RB conditions increases the likelihood of finding a neural dissociation between traditional COVIS tasks, calling into question previous behavioural and neurobiological research that have inferred separable systems on the basis of unidimensional and multidimensional category comparisons (e.g., Nomura et al., 2007; Maddox et al., 2003).

Chapter 6 (Experiments 8, 9 and 10)

Strand two of this thesis examined the prototype effect using both fMRI and tDCS. Previous tDCS research is contradictory in that some find anodal stimulation to the left DLPFC impairs a prototype effect that is present in sham and cathodal conditions (Ambrus et al., 2011), whereas others find the opposing pattern of results, whereby anodal stimulation generates a prototype effect that is not present in sham or cathodal participants (McLaren et al., 2016; Experiment 2). Chapter 6 used a more complex, three category prototype distortion task to investigate this phenomenon and to account for potential ceiling effects in McLaren et al. (2016). Chapter 6 also undertook a novel

investigation into the neural areas correlated with the prototype effect, an area of study previously untapped in the research field.

Chapter 6 found that participants receiving sham or cathodal stimulation during a prototype distortion task responded more accurately to prototype trials compared with previously seen or unseen exemplars after training on category stimuli derived from the prototypes (the prototype effect). Conversely, in anodal conditions, this prototype effect was not found. This is a contrasting result to previous work (McLaren et al., 2016, Experiment 2) that found anodal stimulation induced a prototype effect that was not found in cathodal and sham conditions.

This result supports and extends previous research by Ambrus et al. (2011) who found anodal stimulation abolished a prototype effect that was present in sham and cathodal participants. However, Ambrus et al. also found that anodal stimulation reduced overall performance on the task, whereas the present Chapter suggests the impairment in anodal participants is selective to the prototype effect. This discrepancy could be due to the different prototype distortion tasks presented. While Chapter 6 applied a three category (A, B) task, Ambrus et al. used an (A, not A) task where the participant only learnt to distinguish a single category. Zeithamova et al. (2008) found the neural regions engaged by these two prototype distortion tasks were diverse, and likely reliant on different brain networks. This could explain the differing results in the present experiments and Ambrus et al. (2011).

The present pattern of results is seemingly robust as they were found in the behavioural analyses of the three experiments. However, the findings could be further investigated in the future using TMS. TMS allows for more precise stimulation of a targeted neural area, negating the criticism of Nitsche et al. (2007) that tDCS electrodes have a large surface area, and therefore may also stimulate adjacent brain regions as well as the target region. If the behavioural pattern of results presented in Chapter 6 is

replicated, then it can be inferred that the targeted DLPFC in the tDCS methodology is responsible for the behavioural effect, rather than a simultaneously active neural region inadvertently targeted.

Experiment 10 used a novel tDCS and fMRI method to investigate the brain regions associated with the prototype effect. The DLPFC and the MTL were more active in anodal participants than sham participants in contrasts comparing neural engagement during prototype trials to activation during both old and new exemplar trials. This is a significant finding as the behavioural pattern of results from the three experiments suggest that this is the only measure on which the participants in the two stimulation conditions differ.

The behavioural pattern of results found throughout this chapter and these neural findings could fit with a theory proposed by McLaren et al. (2016). McLaren et al. consider that latent inhibition increases discrimination of prototype stimuli compared to exemplars in the three category task, leading to a prototype effect in sham participants. However, anodal tDCS to the left DLPFC is proposed to disrupt latent inhibition. This increases generalisation between categories and reduces the ability to discriminate prototype stimuli, negating the presence of a prototype effect (e.g., as in McClelland & Rumelhart, 1985).

Previous research has suggested that the MTL is involved in extracting the summed similarity from experienced stimuli, and is implicated in the generalisation process (Nosofsky et al., 2012; Seger et al., 2015). This accounts for the finding that this area is more implicated in anodal participants than in sham participants on measures of the prototype effect, and fits with the above explanation that anodal participants use generalisation to learn the categories.

In the future, it would be interesting to assess the two category prototype distortion task of McLaren et al. (2016) in a joint tDCS and fMRI study. This would

allow for a comparison of the engaged neural areas with those found in Experiment 10. Neurally, it could be predicted that tDCS will affect the brain in the same way in the two and three category task. This would support the proposal from McLaren et al. (2016) as it would offer evidence that tDCS affects the brain in two and three category tasks in the same way, but the category structures are causing different behavioural effects. TMS can also be used to enhance the conclusions drawn from Experiment 10. TMS reduces activation in a targeted area whereas anodal tDCS stimulates the neurons under an electrode. It would be an interesting avenue of research to see the effect of directly reducing activity in the DLPFC as opposed to stimulating it. If a prototype effect is now present, there is further evidence for the conclusions of Chapter 6 and McLaren et al. (2016) that it is an increase of DLPFC activation in response to anodal tDCS that is inhibiting perceptual learning, and consequently attenuating the prototype effect.

Concluding remarks

The aim of the present thesis was to use a broad range of methods to understand the neural underpinnings of category learning behaviour. I consider that this thesis offers novel contributions to the research field. Using a variety of techniques, I provide new insight into the behavioural processes of category learning in RB, II and prototype distortion tasks and the neural systems that underlie these phenomena. The present work argues that there is little evidence for the predictions of COVIS, as the proposed implicit system was not apparent when controlling for confounding variables between RB and II conditions. To account for the findings presented throughout Strand 1, COVIS should include the MTL as a key part of the neurobiological system that learns II conditions. According to the account I have put forward, the MTL is involved in the storing of specific exemplars in declarative memory. Ashby and Rosenthal (in press) have begun

to consider that the implicit system is capable of storing individual exemplars; however, this process is still considered inaccessible to conscious thought. From a processing perspective, the involvement of the MTL is likely to implicate explicit memory, so whilst COVIS may be correct in theorising multiple learning systems (and my data can be explained in these terms) it seems more parsimonious to suggest that both of these systems are likely to be explicit in nature.

This thesis argues that the neural and behavioural results of category learning studies are contingent on the category structures presented to participants, and that conclusions should be drawn in light of the structures learnt (for example whether an II structure is compared to a unidimensional or multidimensional RB learning task). This is apparent in Strand 2, whereby the behavioural impact of tDCS is moderated by task difficulty (e.g., McLaren et al., 2016). Overall, this work enhances our knowledge of the brain regions involved in learning RB and II tasks and in responding to category prototypes, as well as describing novel ways in which to investigate category learning.

References

- Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., & Carter, C. S. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience*, *12*, 977–87.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Ambrus, G. G., Al-moyed, H., Chaieb, L., Sarp, L., Antal, A., & Paulus, W. (2013). The fade-in-short stimulation-fade out approach to sham tDCS - reliable at 1mA for naive and experience subjects , nut not investigators. *Brain Stimulation*, *5*, 499–504.
- Ambrus, G. G., Zimmer, M., Kincses, Z. T., Harza, I., Kovács, G., Paulus, W., & Antal, A. (2011). The enhancement of cortical excitability over the DLPFC before and during training impairs categorization in the prototype distortion task. *Neuropsychologia*, *49*, 1974–1980.
- Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology*, *92*, 1144–1152.
- Ashby, F. G. (1992). Multidimensional models of categorization. In Ashby, F. G. (Ed.). *Multidimensional models of perception and cognition*. Mahwah, NJ: Erlbaum.
- Ashby, F. G., Alfonso-Reese, L. A, Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., & Crossley, M. J. (2010). The neurobiology of categorization. In M. Denis, P. C. Quinn, & S. E Lee, *The Making of Human Concepts* (pp. 75-98). New York: Oxford University Press.

- Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, *31*, 1114–1125.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.
- Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 50–71.
- Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*, 666–677.
- Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *9*, 83–89.
- Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*, 1178–1199.
- Ashby, F. G., & Rosenthal, L. (in press). A neural interpretation of exemplar theory. *Psychological Review*
- Ashby, F. G., & Valentin, V. V. (2005). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen, & C. Lefebvre, (Ed.), *Handbook of Categorization in Cognitive Science* (pp. 547–572). New York: Elsevier.
- Baudewig, J., Nitsche, M. A., Paulus, W., & Frahm, J. (2001). Regional Modulation of BOLD MRI Responses to Human Sensorimotor Activation by Transcranial Direct. *Magnetic Resonance in Medicine*, *201*, 196–201.
- Bindman, L. J., Lippold, O. C. J., & Redfean, J. W. T. (1962). Long-lasting changes in

- the level of the electrical activity of the cerebral cortex produced by polarizing currents. *Nature*, 196, 584–585.
- Bourne, L. E. (1970). Knowing and using concepts. *Psychological Review*, 77, 546-556.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124, 1–38.
- Bunzeck, N., & Düzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/VTA. *Neuron*, 51, 369–379.
- Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition : Extension and limits The prototype effect in face recognition : Extension and limits. *Memory & Cognition*, 27, 139–151.
- Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. (2016). A Comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human brain mapping*, 37, 3557-3574.
- Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience*, 19, 249–265.
- Clark, V. P., Coffman, B. A., Mayer, A. R., Weisend, M. P., Lane, T. D., Calhoun, V. D., Raybourn, E. M., Garcia, C. M., & Wassermann. (2012). TDCS guided using fMRI significantly accelerates learning to identify concealed objects. *NeuroImage*, 59, 117–128.
- Coffman, B. A., Clark, V. P., & Parasuraman, R. (2013). Battery powered thought: Enhancement of attention, learning, and memory in healthy adults using transcranial direct current stimulation. *NeuroImage*, 85, 895–908.
- Coffman, B. A., Trumbo, M. C., & Clark, V. P. (2012). Enhancement of object detection with transcranial direct current stimulation is associated with increased

- attention. *BMC Neuroscience*, *13*, 108.
- Colreavy, E., & Lewandowsky, S. (2008). Strategy development and learning differences in supervised and unsupervised categorization. *Memory & Cognition*, *36*, 762–775.
- Davis, T., Love, B. C., & Preston, A. R. (2012a). Learning the Exception to the Rule: Model-Based fMRI Reveals Specialized Representations for Surprising Category Members. *Cerebral Cortex*, *22*, 260–273.
- Davis, T., Love, B. C., & Preston, A. R. (2012b). Striatal and hippocampal entropy and recognition signals in category learning: simultaneous processes revealed by model-based fMRI. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 821–39.
- Davis, T., Xue, G., Love, B. C., Preston, A. R., & Poldrack, R. A. (2014). Global neural pattern similarity as a common basis for categorization and recognition memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*, 7472–84.
- Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*, 274–290.
- Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2014). Identifying strategy use in category learning tasks: A use for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 933-949.
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*, 172–179.
- Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 840–

- Edmunds, C. E. R., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *The Quarterly Journal of Experimental Psychology*, *68*, 37–41.
- Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M. H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage*, *36*, 511–521.
- Ell, S. W., Marchant, N. L., & Ivry, R. B. (2006). Focal putamen lesions impair learning in rule-based, but not information-integration categorization tasks. *Neuropsychologia*, *44*, 1737–1751.
- Ell, S. W., Weinstein, A., & Ivry, R. B. (2010). Rule-based categorization deficits in focal basal ganglia lesion and Parkinson's disease patients. *Neuropsychologia*, *48*, 2974–2986.
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107–104.
- Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology*, *19*, 212–22.
- Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the Frontal Lobes: The Effects of Engaging Executive Functions on Perceptual Category Learning. *Psychological Science*, *21*, 415–423.
- Flöel, A., Suttorp, W., Kohl, O., Kürten, J., Lohmann, H., Breitenstein, C., & Knecht, S. (2012). Non-invasive brain stimulation improves object-location learning in the elderly. *Neurobiol. Aging. Neurobiology of Aging*, *38*, 1682–1689.
- Fregni, F., Boggio, P. S., Nitsche, M., Berman, F., Antal, A., Feredoes, E., Marcolin, M. A., Rigonatti, S. P., Silva, M. T., Paulus, W., & Pascual-Leone, A. (2005).

- Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research*, 166, 23–30.
- Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19, 1273–1302.
- Functional Imaging Laboratory. (2009). SPM8 [Computer software]. Retrieved from <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Gelman, A. (2002). Prior distribution. In A. H. El-shaarawi & W. W. Piegorsch (Ed.), *Encyclopedia of Environmetrics* (Vol. 3, pp. 1634–1637). Chichester: John Wiley & Sons, Ltd.
- Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). NeuroImage Detection of time-varying signals in event-related fMRI designs. *NeuroImage*, 43, 509–520.
- Heindel, W. C., Festa, E. K., Ott, B. R., Landy, K. M., & Salmon, D. P. (2013). Prototype learning and dissociable categorization systems in Alzheimer’s disease. *Neuropsychologia*, 51, 1699–1708.
- Helie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics*, 72, 1013–1031.
- Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and categorization of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, 101, 116–122.
- Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford: Oxford University Press.

- Kim, C. R., Kim, D. Y., Kim, L. S., Chun, M. H., Kim, S. J., & Park, C. H. (2012). Modulation of cortical activity after anodal transcranial direct current stimulation of the lower limb motor cortex: a functional MRI study. *Brain Stimulation*, *5*, 462–467.
- Kincses, Z. T., Antal, A., Nitsche, M. A., & Paulus, W. (2003). Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the current stimulation of the prefrontal cortex in the human. *Neuropsychologia*, *42*, 113–117.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.
- Koenig, P., Smith, E. E., Glosser, G., DeVita, C., Moore, P., McMillan, C., Gee, J., & Grossman, M. (2005). The neural basis for novel semantic categorization. *Neuroimage*, *24*, 369–383.
- Kuo, M. F., Unger, M., Liebetanz, D., Lang, N., Tergau, F., Paulus, W., & Nitsche, M. A. (2008). Limited impact of homeostatic plasticity on motor learning in humans. *Neuropsychologia*, *46*, 2122–2128.
- Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. (2006). Insight and strategy in multiple cue learning. *Journal of Experimental Psychology: General*, *44*, 162–183.
- Lopez-Paniagua, D., & Seger, C. A. (2011). Interactions within and between corticostriatal loops during component processes of category learning. *Journal of Cognitive Neuroscience*, *23*, 3068–3083.
- Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic bulletin & review*, *9*, 829–835.
- Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective & Behavioral Neuroscience*, *7*, 90–108.
- Luedtke, K., Rushton, A., Wright, C., Jürgens, T., Polzer, A., Mueller, G., & May, A.

- (2015). Effectiveness of transcranial direct current stimulation preceding cognitive behavioural management for chronic low back pain: sham controlled double blinded randomised controlled trial. *British Medical Journal*, *350*, h1640.
- Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition*, *124*, 36–49.
- Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*, 309–332.
- Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 650–662.
- Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*, 582–591.
- Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 227–245.
- Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 100–107.
- Maldjian, J. A., Laurienti, P. J., Burdette, J. B., & Kraft, R. A. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, *19*, 1233–1239.

- Marshall, L., Mölle, M., Hallschmid, M., & Born, J. (2004). Transcranial direct current stimulation during sleep improves declarative memory. *Journal of Neuroscience*, *24*, 9985–9992.
- Matuski, K. (2014). grt: General recognition theory. r package version 0.2: Retrieved from <http://CRAN.R-project.org/package=grt>.
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*(2), 159-188.
- 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., Carpenter, K., Civile, C., McLaren, R., Ku, Y., Zhao, D., & Milton, F. (2016). Assessing the neural basis of uncertainty in perceptual category learning through varying levels of distortion. In J. B. Trobalon & V.D. Chamizo (Ed.), *Associative learning and cognition: Homage to Professor N. J. Mackintosh* (pp. 37-68). Barcelona: Universitat de Barcelona.
- 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* (pp. 102–130). Oxford: Oxford university press.
- Metcalfe, J., & Fisher, R. P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition*, *14*, 164–173.
- Milton, F., Butler, C. R., Benattayallah, A., & Zeman, A. Z. J. (2012). The neural basis of autobiographical memory deficits in transient epileptic amnesia. *Neuropsychologia*, *50*, 3528–3541.
- Milton, F., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity

- sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 676–692.
- Milton, F., Muhlert, N., Butler, C. R., Benattayallah, A., & Zeman, A. Z. J. (2011). The neural correlates of everyday recognition memory. *Brain and Cognition*, *76*, 369–381.
- Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of COVIS. *European Journal of Neuroscience*, *34*, 1326–1336.
- Milton, F., Wills, A. J., & Hodgson, T. L. (2009). The neural basis of overall similarity and single- dimension sorting. *NeuroImage*, *46*, 319–326.
- Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 775-799.
- Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 275-292.
- Minda, J. P. & Smith, J. D. (2011). Prototype models of categorization: basic formulation, predictions, and limitations. In E. M. Pothos & A. J. Wills, *Formal approaches in cateorization* (pp. 40-64). Cambridge: Cambridge University Press.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wagner, T. D. (2000). Control of cognitive processes: Attention and performance XVIII. *Cognitive Psychology*, *41*, 49–100.
- Monsell, S., & Driver, J. (2000). Banishing the control homunculus. In S. Monsell & J. Driver (Ed.), *Control of cognitive processes: Attention and performance XVIII* (pp. 3–32). Cambridge: MIT Press.
- Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: a state-trace analysis. *Memory & Cognition*, *38*, 563–81.

- Newell, B. R., Dunn, J. C., & Kalish, M. (2011). 6 Systems of Category Learning: Fact or Fantasy?. *Psychology of Learning and Motivation-Advances in Research and Theory*, 54, 167-215.
- Newell, B. R., Moore, C. P., Wills, A. J., & Milton, F. (2013). Reinstating the frontal lobes? Having more time to think improves implicit perceptual categorization: a comment on Filoteo, Lauritzen, and Maddox (2010). *Psychological Science*, 24, 386–389.
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, 25, 653–660.
- Nitsche, M. A., Cohen, L., Wassermann, E. M., Priori, A., Lang, N., Antal, A., Hummel, F., Boggio, P. S., Fregni, F., & Pascual-Leone, A. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation*, 3, 206–23.
- Nitsche, M. A., Doemkes, S., Karakoese, T., Antal, A., Liebetanz, D., Lang, N., Tergau, F., & Paulus, W. (2007). Shaping the effects of transcranial direct current stimulation of the human motor cortex. *Journal of Neurophysiology*, 97, 3109–3117.
- Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology*, 527, 633–639.
- Nitsche, M. A., & Paulus, W. (2011). Transcranial direct current stimulation—update 2011. *Restorative Neurology and Neuroscience*, 29, 463–492.
- Nomura, E., Maddox, W., Filoteo, J., Ing, A., Gitelman, D., Parrish, T., Mesulam, M. M., & Reber, P. (2007). Neural Correlates of Rule-Based and Information-Integration Visual Category Learning. *Cerebral Cortex*, 17, 37–43.
- Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral*

Reviews, 32, 279–291.

- Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Attention, Perception, & Psychophysics*, 38, 415-432.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, 115, 39-57.
- Nosofsky, R. M., Denton, S. E., Zaki, S. R., Murphy-Knudsen, A. F., & Unverzagt, F. W. (2012). Studies of implicit prototype extraction in patients with mild cognitive impairment and early Alzheimer’s disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 860–880.
- Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of “ multiple-system ” phenomena in perceptual categorization, 7, 375–402.
- Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, 9, 169-174-180.
- Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: implicit learning or cognitive complexity? *Memory & Cognition*, 33, 1256–1271.
- Olejnik, S. & Algina, J. (2000). Measures of Effect Size for Comparative Studies: Applications, Interpretations and Limitations. *Contemporary Educational Psychology*, 25, 241-286.
- Palmeri, T. J., & Flanery, M. A. (1999). Learning About Categories in the Absence of Training: Profound Amnesia and the Relationship Between Perceptual Categorization and Recognition Memory. *Psychological Science*, 10, 526–530.
- Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The*

- Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 54, 197–235.
- Peña-Gómez, C., Sala-Lonch, R., Junqué, C., Clemente, I. C., Vidal, D., Bargalló, N., Falcón, C., Valls-Solé, J., Pascual-leone, Á., & Bartrés-Faz, D. (2013). Modulation of large-scale brain networks by transcranial direct current stimulation evidenced by resting-state functional MRI. *Brain Stimulation*, 5, 252–263.
- Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, 135, 66–72.
- Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, 2, 67–70.
- Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews*, 32, 197–205.
- Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during cognitive skill learning. *Neuropsychology*, 13, 564–574.
- Poreisz, C., Boros, K., Antal, A., & Paulus, W. (2007). Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. *Brain research bulletin*, 72, 208-214.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.
- Pothos, E. M., & Bailey, T. M. (2009). Predicting category intuitiveness with the rational model , the simplicity model , and the generalised context model predicting category intuitiveness with the rational model , the simplicity model , and the generalised context model. *Journal of Experimental Psychology: Learning , Memory , and Cognition*, 35, 1062–1080.
- Pothos, E. M., Chater, N., & Hines, P. (2011). *The simplicity model of unsupervised*

- categorization*. In E. M. Pothos & A. J. Wills (eds.), *Formal approaches in categorization*, Cambridge: Cambridge University Press.
- Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification : A simplicity approach. *Cognition*, *107*, 581–602.
- Psychology Software Tools, Inc. (2002). E-Prime [Computer software]. Retrieved from <http://www.pstnet.com>.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing [Computer software]. Retrieved from <http://www.R-project.org/>.
- Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson’s disease. *Behavioral Neuroscience*, *113*, 235–242.
- Reber, P. J., Stark, C. E., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 747–750.
- Richler, J. J., & Palmeri, T. J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 75–94.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.
- Schnyer, D. M., Maddox, W. T., Ell, S., Davis, S., Pacheco, J., & Verfaellie, M. (2009). Prefrontal contributions to rule-based and information-integration category learning. *Neuropsychologia*, *47*, 2995–3006.
- Scoville, W. B., & Milner, B. (2000). Loss of recent memory after bilateral hippocampal lesions. 1957. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *12*, 103–113.
- Seger, C. A., Braunlich, K., Wehe, H. S., & Liu, Z. (2015). Generalisation in category

- learning: The roles of representational and decisional uncertainty. *Journal of Neuroscience*, *35*, 8802–8812.
- Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *The Journal of Neuroscience*, *25*, 2941–2951.
- Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex*, *16*, 1546–1555.
- Seger, C. A., Dennison, C. S., Lopez-paniagua, D., Peterson, E. J., & Roark, A. A. (2011). Dissociating hippocampal and basal ganglia contributions to category learning using stimulus novelty and subjective judgments, *55*, 1739–1753.
- Seger, C. A., & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 149–161.
- Seger, C. A., Poldrack, R. A., Prabhakaran, V., Zhao, M., Glover, G. H., & Gabrieli, J. D. (2000). Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia*, *38*(9), 1316–1324.
- Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review*, *18*, 414–421.
- Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred Feedback Sharply Dissociates Implicit and Explicit Category Learning. *Psychological Science*, *25*, 447–457.
- Smith, J. D., Redford, J. S., & Haas, S. M. (2008). Prototype abstraction by monkeys (Macaca mulatta). *Journal of Experimental Psychology: General*, *137*, 390–401.
- Song, X. M., Dong, Z. Y., Long, X. Y., Si, S. F., Zuo, X. N., Zhu, C. Z., He, Y., Yan, C. G., & Zang, Y. F. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS One* *6*, e25031.
- Soto, F. A., Waldschmidt, J. G., Helie, S., & Ashby, F. G. (2013). Brain activity across

- the development of automatic categorization: A comparison of categorization tasks using multi-voxel pattern analysis. *NeuroImage*, 71, 284–297.
- Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279–306.
- Stanton, R. D., & Nosofsky, R. M. (2007). Feedback interference and dissociations of classification: evidence against the multiple-learning-systems hypothesis. *Memory & Cognition*, 35, 1747–1758.
- Stanton, R. D., & Nosofsky, R. M. (2013). Category number impacts rule-based and information-integration category learning: A reassessment of evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1174–1191.
- Stark, C. E., & Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 12760–12766.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*. New York: Thieme Medical Publishers
- Taleb, N. N. (2010). *The Bed of Procrustes: Philosophical and Practical Aphorisms*. Great Britain: Allen Lane.
- Vercammen, A., Rushby, J. A., Loo, C., Short, B., Weickert, C. S., & Weickert, T. W. (2011). Transcranial direct current stimulation influences probabilistic association learning in schizophrenia. *Schizophrenia Research*, 131, 198–205.
- Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *NeuroImage*, 56, 1791–1802.
- Wills, A. J., Inkster, A. B., & Milton, F. (2015). Combination or Differentiation? Two

- theories of processing order in classification. *Cognitive Psychology*, *80*, 1–33.
- Worthy, D. A., Markman, A. B., & Maddox, T. W. (2013). Feedback and stimulus-offset timing effects in perceptual category learning. *Brain and Cognition*, *81*, 283–293.
- Xie, Z., Maddox, W. T., McGeary, J. E., & Chandrasekaran, B. (2015). The C957T polymorphism in the dopamine receptor D2. *The Journal of Neurophysiology*, *113*, 3281–3290.
- Yang, R., & Berger, J. O. (1998). *A Catalog of Noninformative Priors, ISDS discussion paper* (pp. 97-42), Institute of Statistics and Decision Sciences, Duke University.
- Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unverzagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: evidence for single-system models. *Journal of the International Neuropsychological Society*, *9*, 394–406.
- Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*, 387–398.
- Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable Prototype Learning Systems: Evidence from Brain Imaging and Behavior. *Journal of Neuroscience*, *28*, 13194–13201.
- Zink, C. F., Pagnoni, G., Chappelow, J., Martin-Skurski, M., & Berns, G. S. (2006). Human striatal activation reflects degree of stimulus saliency. *NeuroImage*, *29*, 977–983.