

Electronic Thesis and Dissertation Repository

---

7-13-2023 12:00 PM

## Category Learning DLPFC Single-Dissociation by fNIRS

Tim Qiu, *Western University*

Supervisor: Minda, John P., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Neuroscience

© Tim Qiu 2023

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Cognitive Neuroscience Commons](#)

---

### Recommended Citation

Qiu, Tim, "Category Learning DLPFC Single-Dissociation by fNIRS" (2023). *Electronic Thesis and Dissertation Repository*. 9384.

<https://ir.lib.uwo.ca/etd/9384>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

## Abstract

It remains an open question whether modern neuroimaging clearly dissociates the Explicit system that learns by encoding rules, and the Implicit system that learns by encoding information-integration boundaries. Further, there are nearly no applications of fNIRS as a modality in studying category learning. We conduct two behavioural experiments to validate a carefully controlled categorization task intended to dissociate Explicit and Implicit systems. Then we apply fNIRS neuroimaging within-subjects to localize a neuroanatomical dissociation. We localized two effects to R DLPFC (1) a simple single-dissociation of higher activity in RB categorization, and (2) a negative relationship between overall task performance and the magnitude of neural activation. We conclude that explicit and implicit cortical activity are dissociable by neuroimaging, and that fNIRS is a feasible modality to study human categorization.

**Keywords:** Category learning, Neuroimaging, fNIRS, Dissociation, COVIS, Explicit, Implicit

## Summary for Lay Audience

It is already well known that there are at least two ways in which humans learn new categories. The first method applies to categorizing objects such as cups and mugs. How would you distinguish between cups and mugs? It turns out the mind narrows in on one main property, in this case—the handle. After seeing many cups and mugs you might notice that most mugs have a handle. You may come to formulate some kind of a rule: "mugs have handles, cups don't". With that bit of language, you have now categorized most of the cups and mugs you'll ever see. This is known as explicit categorization.

However, sometimes this explicit method may not work for certain types of objects such as cats and dogs. If you look closely, you may realize that it's difficult to narrow in on one 'property' that cleanly divides cats and dogs. Is it the size? The amount of fluff? The whisker length? There is no single clear distinction. However, all of us easily classify cats and dogs. The second way we classify is based on picking up associations between certain objects we see and a label or action that corresponds with it. This is called implicit categorization.

If we learn categories in two different ways, does that mean there are two different circuits of the brain involved? This study aims to answer that question. We do that by designing two tasks, one explicit task, one implicit. Then we use functional near infrared spectroscopy (fNIRS) to measure the blood flowing through the brain at different locations. More blood flow tells us that one part of the brain is working harder. Then we compare blood flow patterns between explicit and implicit.

We find that there was more activity in the right dorsolateral prefrontal cortex (located above your right eyebrow, where the hairline begins) when people are doing explicit categorization. Our findings align with previous work on memory, language, and reasoning. We therefore conclude that there is likely to be distinct underlying brain circuits for explicit and implicit category learning.

## Acknowledgements

First and foremost I would like to acknowledge my partner for their support, through this work and through life.

None of this work would be possible without the professional advice, expertise, and supervision of my mentor, Dr John Paul Minda. Not only have Paul's contribution been invaluable for my work and intellectual development, but he has created a lab environment conducive to growth and learning. For this I am ineffably grateful.

I'd like to give major thanks to my advisory committee and program chair, Dr. Adrian M Owen, Dr. Paul L Gribble, and Dr. Marc F Joannise. Their technical suggestions, critiques, and big picture advice have improved the present document much more than I can manage alone.

Many thanks to my defense committee for their critical evaluation of this present work which has resulted in significantly improved revisions. Thank you Dr. Barbara Fenesi, Dr. Emma G Duerden, and Dr. Adrian M Owen.

Even preceding the present work, I've had the pleasure of collaborating and discussing ideas with Dr. Ana C. Ruiz Prado, Anthony, Chelsea, Dr. Toka (Tianshu) Zhu.

I'd like to extend special thanks to Dr. Emily G Nielsen for mentoring me in the early days and showing me the ropes to the field.

Lastly I'd like to recognize the assistance of Shelby Laing, an exceptional undergraduate research assistant who helped with a significant proportion of data collection in the present work.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Summary for Lay Audience</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Category Learning Systems . . . . .	1
1.2 Imaging, Neuropsychology, Comparative Cognition . . . . .	3
1.3 Functional Near-Infrared Spectroscopy . . . . .	6
1.4 Objective . . . . .	8
<b>2 Experiment I</b>	<b>9</b>
2.1 Methods . . . . .	9
2.2 Results . . . . .	11
<b>3 Experiment II</b>	<b>13</b>
3.1 Methods . . . . .	13
3.2 Results . . . . .	14
<b>4 Experiment III</b>	<b>16</b>
4.1 Methods . . . . .	16
4.1.1 Design . . . . .	16
4.1.2 fNIRS Hardware . . . . .	16
4.1.3 fNIRS Analysis . . . . .	18
4.2 Results . . . . .	19
4.2.1 Behavioural . . . . .	19
4.2.2 Subject-level fNIRS . . . . .	21
4.2.3 Group-level fNIRS . . . . .	21
<b>5 Discussion</b>	<b>26</b>
5.1 Neural Dissociation . . . . .	26
5.2 Performance Haemodynamics . . . . .	28

5.3	Multiple Comparisons . . . . .	29
5.4	Feasibility of fNIRS . . . . .	31
5.5	Conclusion . . . . .	32
	<b>Bibliography</b>	<b>33</b>
	<b>Appendix</b>	<b>39</b>
	<b>Curriculum Vitae</b>	<b>39</b>

# List of Figures

2.1	Category structures for the two experimental conditions. Frequency is given in cycles per image, orientation is given in degrees. Each point represents a single sine-wave grating stimuli. Squares denote category A, triangles denote category B. The optimal boundaries are shown as lines separating the two categories. . . . .	10
2.2	Experiment 1 Data. (a) Rule-based (RB) and Information-Integration (II) learning curves. Blue triangle solid line indicates RB category set, red circle dashed line indicates II category set. Error bars denote 95% confidence intervals of the mean. (b) Proportion of participants best fit by RB and II boundary models for respective experimental condition. CR indicates the proportion best fit by 2D general conjunctive classifier, II proportion best fit by 2D general linear classifier, RB proportion best fit by 1D linear classifier. Error bars denote standard error of proportion estimates. . . . .	11
3.1	Learning and imaging protocol for experiments 2 and 3. In Experiment 2 the design was between-subjects so each participant completed 20 minutes of either RB or II conditions. In experiment 3 the design was within-subjects so the same participant completed 20 minutes of both RB and II counterbalanced for a total scanning time of 40 minutes per participant. . . . .	14
3.2	Experiment 2 Data. (a) Rule-based (RB) and Information-Integration (II) learning curves. Blue triangle solid line indicates RB category set, red circle dashed line indicates II category set. Error bars denote 95% confidence intervals of the mean. (b) Proportion of participants best fit by RB and II boundary models for respective experimental condition. II proportion best fit by 2D general linear classifier, RB proportion best fit by 1D linear classifier. Error bars denote standard error of proportion estimates. . . . .	15
4.1	Cortical sensitivity of probe layout. Red points are sources, blue points are detectors, yellow lines joining the points represent the measurement channels between optodes. Note that the true measurement channels take the form of a concave volume passing through the cortex. Sensitivity of the probe layout is displayed as heatmap on the surface of cortex ranging from $0.25 \text{ mm}^{-1}$ (blue) to $1.0 \text{ mm}^{-1}$ (red) . . . . .	17
4.2	Experiment 3 Behavioural Data. (a) RB and II. Blue triangle solid line indicates RB category set, red circle dashed line indicates II category set. Errorbars denote 95% confidence intervals of the mean. (b) Reaction times denoted in milliseconds. Errorbars denote 95% confidence intervals of the mean. . . . .	20

4.3	GRT modeling results (a) Depicts the best fit GRT boundary by experimental condition. Errorbars denote standard error of the proportion. (b) Depicts the relative proportion of all participants best fit respective boundary by cumulative block, all proportions sum to one. . . . .	20
4.4	Projection of fNIRS channel activation for task condition contrasts. Red represents relative activation, blue represents relative deactivation. Images are thresholded displaying only channels with FDR-adjusted p-values < .05 (For full statistics, see Table 4.2) Each disc projection represents the estimated trough of a measurement channel and spanning 20 mm in diameter. 'A' denotes anterior view. . . . .	22
4.5	(a) Channel 11 of R DLPFC exhibited a significant effect of categorization strategy as best fit by GRT modeling. (b) Average estimated HbO <sub>2</sub> haemodynamic response during CATG blocks for participants partitioned as RB or II learners by best fitting GRT models. Dotted lines indicate onset and end of CATG trials. Band indicates standard error of mean haemoglobin concentration. . . . .	24
4.6	(a) Channel 18 of R DLPFC exhibits significant negative relationship between performance and haemodynamic response. (b) Z-scored performance plotted against haemodynamic response. Band denotes standard error of the estimate. . . . .	25
Af1	Model diagnostic plots for the significant channels exhibit group effects. Note that model residuals appear to violate assumption of normality. However with the sample size is >30 (n = 86), the violation has relatively minimal influence on whether OLS is the best linear unbiased estimator. . . . .	40
Af2	Examples of interesting qualitative effects that are initially significant, but sequestered by FDR corrections for multiple comparisons. (a) Estimated HRF on channel 14, L DLPFC, BA9, shows a marked difference between RB and II condition, $t = -2.54$ , $p = .04$ , $p_{FDR} = .18$ , but fails to survive FDR correction. (b) Performance × condition exhibits a visible interaction effect, $t = 2.56$ , $p = .013$ , $p_{FDR} = .17$ , but fails to survive FDR correction. . . . .	40
Af3	Schematic of fNIRS channel non-independence. (a) One source and two detectors, D <sub>1</sub> , D <sub>2</sub> creates two channels of measurements where the sensitivity profile overlaps. Instances of cluster activation may happen at a relatively independent measurement location; A, B; or at a non-independent locations; C. (b) Example of non-independent measurements C failing to pass significance thresholds post-correction for multiple comparisons. Dashed line of $\alpha$ represents the significance threshold of the test-statistic pre and post corrections. Note that cluster C would otherwise pass an uncorrected threshold. . . . .	41



# List of Tables

4.1	Optode Location and ROI Specificity . . . . .	18
4.2	Subject-Level Contrasts. Coordinates denote estimated MNI coordinates based on optode placements according to standard 10-20 EEG positions. BA denote Brodmann's areas. . . . .	23
AtI	Basic Participant Demographics. Age displays means, all other variables display percentages relative to sample size for each respective experiment. . . . .	39

# Chapter 1

## Introduction

Laozi, in the Tao Te Ching writes of the ‘Ten-Thousand Things’, in reference to human experience (Laozi, 1868). In this world, no two objects are exactly congruent, yet myriad objects perceived at various times and places are deftly abstracted to categories by the human mind (Posner & Keele, 1968; Nosofsky, 1986). Without categorization and generalization, each visual object encountered would present as an entirely strange and novel object. Precisely how the human mind acquires or performs categorization is an old question, though remains relevant and active today (Aristotle, 1938).

### 1.1 Category Learning Systems

Convergent lines of evidence from neuropsychological, behavioral, and comparative cognition literature have culminated in an influential theory of category learning: Competition Of Verbal and implicit Systems (COVIS), which posits two behaviourally and anatomically distinct systems (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Maddox, Todd Maddox, & Gregory Ashby, 2004).

The explicit system tests and stores rules which are snippets of language that segregate stimuli into two or more categories, each accompanied by a distinct language label, (Ashby et al., 1998) For example: Small timepieces are watches, large timepieces are clocks. This rule can be formally represented in a coordinate space where the axes correspond to relevant feature dimensions of the stimuli. In this case, size is the relevant dimension. Generally, timepieces exceeding the size of one’s wrist (diameter  $> 5$  cm) is considered a ‘clock’, while smaller articles (diameter  $< 5$  cm) are considered ‘watches’. In this way, the language rule can be formally described as a linear decision boundary where (diameter = 5 cm). Objects falling on either side of this boundary would be subsequently categorized as a “watch” or “clock” (Ashby, 2014; Hélie, Turner, Crossley, Ell, & Ashby, 2017). There will certainly be exceptions to the rule, such as a large novelty watch which may exceed the 5 cm decision boundary but is still deemed a watch. However, the general purpose of category learning is not to achieve perfect segregation but rather to reduce the complexity of perceptual information in the surrounding environment (Posner & Keele, 1968; Nosofsky, 1986). In this regard, exceptions to the rule are tolerated, provided that the rule is sufficiently useful in compressing the information complexity of the whole object set (all possible timepieces). The effectiveness of such explicit categorization is self-evident in the present example, the concept of ‘all possible timepieces’ is rarely discussed in common parlance, the word ‘timepiece’ itself is uncommonly used in English (Corpus of Contemporary American English frequency ranking: #27065), whereas ‘watches’ (COCA: #2375) and ‘clocks’ (COCA:

#2725) are universally understood and frequently mentioned words among English speakers (Davies, 2010). During the rule learning process, candidate rules are stored and manipulated in working memory (WM) and the process is anatomically instantiated as reverberating circuits primarily involving the prefrontal cortex, anterior cingulate, head of the caudate, and hippocampus (Ashby et al., 1998; Minda & Miles, 2010). The explicit system can be tested by creating Rule-Based (RB) categories, stimuli which are segregated by a verbalizable boundary, often unidimensional and linear (Maddox et al., 2004; Ashby, 2014).

The implicit system functions by reward-mediated procedural learning and it is capable of arbitrarily mapping a set of stimuli representations to a set of motor responses (Ashby et al., 1998). Many object sets do not lend themselves to clean categorization mediated by a verbalizable rule. Consider that cats and dogs are easily discriminated, yet lack a prominent and obvious discriminating feature that is easily described by language. In these cases, multiple features such as the height, width, length, shape, color, texture must be integrated to discern the object set into coherent categories (Ashby et al., 1998). Formally, the distinction between two implicit categories can also be represented as a decision boundary dividing a coordinate space into two or more regions (Ashby, 2014; Hélie et al., 2017). However, implicit boundaries qualitatively differs from rule-based boundaries. Implicit boundaries must integrate two or more relevant feature dimensions, in a 2D stimuli space, implicit boundaries must describe a curve that spans both  $x$  and  $y$  stimuli dimensions. In 3D space, implicit boundaries must describe a plane or surface spanning all of  $x$ ,  $y$ , and  $z$  stimuli dimensions. Further, implicit boundaries can be both linear or nonlinear, whereas rule-based boundaries are constrained to be linear only (Ashby & Maddox, 2011). Implicit learning relies on feed-forward architecture, beginning in extrastriate visual cortex, proceeding sequentially through striatum, globus pallidus, thalamus, and terminating in the premotor cortex (Ashby et al., 1998; Minda & Miles, 2010). The implicit system can be tested by creating Information-Integration (II) categories, stimuli which are segregated by non-verbalizable boundaries that are multidimensional, and may be linear or nonlinear (Ashby, 2014). Henceforth, the explicit/implicit terminology will strictly refer to the two category learning systems as posited by COVIS, and the Rule-Based/Information-Integration (RB/II) terminology will strictly refer to the designed category structure of the stimuli.

COVIS ambitiously attempts to stitch together the computational, algorithmic, and implementation levels into a cohesive and holistic description of human category learning (Marr, 1982; Ashby et al., 1998). For simplicity, the organization of computational, algorithmic, and implementation levels may be colloquially approximated as, the ‘behaviour’, ‘information’, and ‘physiology’, respectively (Marr, 1982). Minor refinements notwithstanding, the computational (‘behaviour’) and algorithmic (‘information’) levels of COVIS stands largely intact under critical tests (Maddox et al., 2004; Minda & Miles, 2010; Ashby & Maddox, 2011). Ambiguity remains at the level of anatomical implementation—the ‘physiology’ (Nomura et al., 2007; Carpenter, Wills, Benattayallah, & Milton, 2016; Soto, Waldschmidt, Hélie, & Ashby, 2013). Neuroimaging advances of technology and technique present novel critical tests to COVIS at the level of anatomy. There is currently an active effort to apply modern neuroimaging methods to probe the question of whether real-time functional activity of explicit and implicit systems are dissociable. Thus far, conclusive neuroimaging evidence in favour or against COVIS remains elusive. This is the central problem to which the present study is applied.

## 1.2 Imaging, Neuropsychology, Comparative Cognition

There are two fMRI studies of direct relevance to the present work, Nomura and colleagues measured brain activity of participants learning RB and II categories, and were the first to reported a direct significant contrast of explicit and implicit conditions, localizing explicit activation to the left medial temporal lobe, and implicit activation to the right body of the caudate (Nomura et al., 2007). However this study was followed by a replication attempt from an independent group, nearly a decade later, which failed to find a dissociation (Carpenter et al., 2016). In particular, Carpenter's replication discussed several critiques of the original study by Nomura. The overarching critique was that the original study by Nomura did not employ sufficiently stringent controls of ostensibly irrelevant or minor variables. These are; (1) the standard distance between categories which may confound functional activation due to representational differences, (2) the number of relevant stimuli dimensions which may cause selective attentional confounds, (3) the categorization performance which may cause confounds due to cognitive effort, (4) using incorrect trials as a subject-level baseline for contrasts which is convenient, but washes out the differential contributions of incorrect feedback to the learning mechanism of explicit and implicit systems. In sum, Carpenter and colleagues posit that the differential neural activity reported by Nomura and colleagues may be partially or even totally explained by the aforementioned unaccounted confounds (Nomura et al., 2007; Carpenter et al., 2016). Tightening experimental design, Carpenter and colleagues conclusively reported no significant differences in brain activity between explicit and implicit learning; they replicated the event-related design of Nomura's study, analyzed their own block-based design, and attempted to partition participants by General Recognition Theory (GRT) model-based results, but failed to find a meaningful effect in all approaches to support the assumptions of COVIS. In fact, Carpenter and colleagues found a small, antithetical effect of higher implicit related activation in the right parahippocampal gyrus when comparing later runs against earlier runs (Carpenter et al., 2016).

These two studies are representative of the present state of the neuroimaging literature in addressing the question of dissociable category learning systems (Nomura et al., 2007; Carpenter et al., 2016). Studies range widely in their design, intervention, stimuli, imaging modalities, and analytical approaches. Unsurprisingly, reports are lukewarm in their support for COVIS theory and there are anatomical inconsistencies across studies that remains to be elucidated (Milton & Pothos, 2011; Nomura et al., 2007; Gureckis, James, & Nosofsky, 2011; Morgan, Zeithamova, Luu, & Tucker, 2020; Wu, Fu, & Rose, 2020; Soto et al., 2013; Reber, Gitelman, Parrish, & Mesulam, 2003; Aizenstein et al., 2000; Helie, Roeder, & Ashby, 2010; Carpenter et al., 2016).

A series of three studies employed an instruction-based intervention of explicit and implicit learning conditions applied to dot-distortion stimuli (Aizenstein et al., 2000; Reber et al., 2003; Gureckis et al., 2011; Posner & Keele, 1968). All three report a qualitative difference in the pattern of brain activity between explicit and implicit category learning but fail to find a quantitative difference—Aizenstein's report did not directly contrast explicit and implicit conditions, and both Reber, Gureckis did not observe a statistically significant direct contrast between explicit and implicit conditions (Aizenstein et al., 2000; Reber et al., 2003; Gureckis et al., 2011). Further, it can be argued that instruction-based intervention is subject to confounds due to participant compliance. Without specifically designing distinct RB and II categories, one must rely solely on participant compliance to preserve integrity of experimental conditions. Regardless, these studies report a consistent qualitative overlap of explicit activation across studies, predominantly in subregions of prefrontal cortex (PFC) such as: inferior frontal gyri, medial frontal gyri, dorsolateral prefrontal cortex (DLPFC), frontal eye fields, and an-

terior prefrontal cortex which corresponds to Brodmann's areas (BA) 8, 9, 10, 44, 46. Qualitative implicit activation patterns predominantly overlapped in the supplementary motor area (SMA) corresponding to BA 6 (Aizenstein et al., 2000; Gureckis et al., 2011; Reber et al., 2003).

A multi-voxel pattern analysis approach examined category learning automaticity, which generally includes over 10,000 trials per participant, and distributed learning over spaced sessions (Soto et al., 2013). The results of the initial sessions are relevant to the present investigation, which discriminated ventrolateral prefrontal cortex (VLPFC) uniquely associating with explicit learning, while SMA, premotor cortex (PMd), and primary motor cortex (M1) are uniquely associated with implicit activity. A canonical general linear model (GLM) based analysis on the same dataset corroborated the distinction of VLPFC and SMA, corresponding to BA 44, 45, 47 and BA 6, respectively (Helie et al., 2010). However, another study reported mixed activation of general PFC gyri and SMA in both explicit and implicit conditions when analyzed by canonical GLM on a small pilot sample (Morgan et al., 2020). They also applied MVPA to find a unique association of inferior frontal gyri to explicit activity, but the choice of stimuli were somewhat idiosyncratic: American football defensive-line formations. Football formations certainly have increased ecological validity compared to traditional stimuli such as sine-wave gratings, but they also introduce many additional confounding effects at higher levels of cognition (prior knowledge, cultural idiosyncrasies, complex stimuli dimension interactions) which may be difficult to account for (Morgan et al., 2020). Another study using novel stimuli sets also reported unclear dissociation in support of COVIS theory (Milton & Pothos, 2011).

Finally, there exists one fNIRS-based study, which primarily investigated the interaction of audio and visual categorization effects, and happened to discover an effect of visual rule-based category learning in DLPFC (Wu et al., 2020).

Thus, the present state of the neuroimaging literature with respect to a double dissociation of COVIS systems is ambiguous. However, these studies were initially motivated by a large variety of previous work in comparative cognition and neuropsychology which report functional double-dissociations between explicit and implicit systems.

In these studies, deficits of PFC are frequently targeted to dissociate between explicit and implicit category learning (Ashby & Ell, 2001; Maddox et al., 2004). Lesions or diseases affecting PFC correspond to functional impairment during explicit categorization tasks (Bozoki, Grossman, & Smith, 2006; Robinson, Heaton, Lehman, & Stilson, 1980; Virag et al., 2015; Maddox & Filoteo, 2001; Brown & Marsden, 1988). The Wisconsin Card Sorting Task is a classic explicit category learning task in which patients with prefrontal cortical insult show impairment relative to demographically matched controls (Robinson et al., 1980; Virag et al., 2015; Janowsky, Shimamura, Kritchevsky, & Squire, 1989). Patients suffering a variety of frontal lobe lesions such as tumours, hemorrhage, clots, and infarcts consistently demonstrate impaired explicit learning performance relative to matched controls (Robinson et al., 1980). These specific lesion studies on the Wisconsin Card Sorting Task are frequently and well replicated (Ashby & Ell, 2001; Janowsky et al., 1989). Memory studies corroborate these findings; a lesion study of explicit and implicit memory which implemented category learning of stimuli found that while the explicit memory task was significantly impaired in patients, a corresponding implicit memory task showed no impairment relative to healthy controls (Gershberg, 1997). The prefrontal cortex is also disproportionately sensitive to neurotoxic effects of chronic alcoholism (Lewohl et al., 2000). Consequently, chronic alcoholics have impaired executive function and explicit category learning performance, but retain normal performance in implicit category learning tasks compared to healthy controls (Virag et al., 2015; Janowsky et al., 1989). Similarly, early stage Alzheimer's disease preferentially affects prefrontal cortex and medial temporal lobe; patients suffer-

ing from Alzheimer's disease have similar implicit category learning performance and representation profiles compared to controls (Bozoki et al., 2006). These neuropsychological dissociations are corroborated by comparative cognition data. Monkeys—which have simpler prefrontal cortex development relative to humans, are incapable of abstracting a category rule whereas human counterparts have no difficulty on the same category sets (Shepard, Hovland, & Jenkins, 1961; Smith, Minda, & Washburn, 2004). Conversely, on implicit-type category learning tasks, the difference between monkey and human performance is less drastic (Smith et al., 2004). Comparisons have also been made within humans across age groups. Children under 5 have not yet reached sufficient prefrontal development to abstract and manipulate language-based rules (Minda, Desroches, & Church, 2008). On the other hand, older adults have reduced executive function—specifically in working memory capacity (Bopp & Verhaeghen, 2005). Across a range of 6 classic—Shepard, Hovland, Jenkins—category tasks, children under 5 implicitly learn all tasks, even for tasks in which adults can trivially learn the optimal rule (Shepard et al., 1961; Minda et al., 2008; Miles & Minda, 2009). In the opposite direction, comparisons of older and young adults show that older adults are still capable of learning rules, but when the rule exceeds a certain complexity, their performance deteriorates compared to young adult controls (Rabi & Minda, 2016). Taken together, these studies show that PFC is critical for successful RB category learning, whereas PFC insult does not necessarily deteriorate implicit category learning (Maddox & Filoteo, 2001).

In contrast, basal ganglia diseases, specifically those posing strong insult to the striatum and the caudate body tend to impair implicit category learning while leaving explicit category learning relatively intact (Cools, van den Bercken, Horstink, van Spaendonck, & Berger, 1984; Maddox & Filoteo, 2001; Knowlton, Paulsen, Swenson, & Butters, 1996). Note that in many cases explicit category learning may also be impaired, but the magnitude of impairment is limited whereas implicit category learning impairment tends to be catastrophic—patients suffering from basal ganglia disorders are frequently unable to categorize above chance accuracy (Cools et al., 1984; Brown & Marsden, 1988). Parkinson's disease, in the early stages primarily disrupts the basal ganglia but not the prefrontal cortex; patient studies demonstrate implicit categorization impairment, but little to no impairment for both explicit category learning and executive function tests such as Stroop or working memory tasks. (Cools et al., 1984; Maddox & Filoteo, 2001; Brown & Marsden, 1988). Similarly, Huntington's disease primarily targets the striatum and specifically the caudate body, a critical node of the implicit system (Ashby et al., 1998; Maddox et al., 2004). As a consequence, Huntington's patients display marked decreases in probabilistic (implicit) categorization tasks and a comparatively minor decrease in tasks require verbal function which implicates the PFC (Knowlton et al., 1996). The implicit system as defined by COVIS theory specifies the heavy reliance on basal ganglia-thalamocortical circuits (Ashby et al., 1998; Alexander, DeLong, & Strick, 1986). Dense projections originating from the putamen pass through the ventrolateral nuclei of the thalamus and terminate in the PMd and SMA (Dum & Strick, 2005; Matelli & Luppino, 1996; Alexander et al., 1986). Both SMA and PMd are critical components of motor learning; implicit category learning can be thought of as a specific case of motor skill learning (Ashby et al., 1998). In recent work, the motor component of implicit learning has received more attention than originally considered (Ashby et al., 1998; Willingham, 1998). Activity of SMA and PMd are well documented to be active during motor and procedural learning paradigms (Seitz et al., 1994; Hikosaka, Rand, Miyachi, & Miyashita, 1995; Grafton et al., 1992). Indeed, patients with focal lesions in PMd or SMA perform significantly worse on sensorimotor learning tasks than even patients with primary motor cortex lesions (Halsband & Freund, 1990; Halsband & Lange, 2006). These data taken together suggest that the motor aspect of implicit category learning is critical, and

that both the striatum and premotor regions must necessarily be intact for proper function of the implicit system (Ashby et al., 1998; Alexander et al., 1986; Willingham, 1998).

Past work in neuropsychology and comparative cognition has established a compelling view of neuroanatomical dissociability between explicit and implicit systems (Ashby et al., 1998; Maddox & Filoteo, 2001). However, the present state of the imaging literature does not unanimously corroborate the previous view, but rather introduces some new ambiguities (Nomura et al., 2007; Carpenter et al., 2016; Morgan et al., 2020). Much of the neuropsychological work on implicit categorization specifically interrogate the basal ganglia as it is often the site of early deterioration such as in Parkinson's or Huntington's disease (Brown & Marsden, 1988; Knowlton et al., 1996). In the present study, we must constrain target regions of interest to locations accessible via fNIRS neuroimaging, and therefore cannot consider the basal ganglia as a region of interest. However, both the PMd and SMA are downstream of basal ganglia projections and have been shown to be densely connected, and critical for implicit learning processes (Hikosaka et al., 1995; Halsband & Freund, 1990; Matelli & Luppino, 1996).

Thus, there are four lines of evidence suggesting that a potential double-dissociation can be found by the application of neuroimaging, (1) neuropsychology, (2) comparative cognition, (3) neuroimaging studies specifying the exact statistical contrast addressing the present research question, and (4) the other relevant neuroimaging work, though perhaps not specifying the exact statistical contrast, show qualitative patterns of activity in support of a double-dissociation (Ashby & Ell, 2001; Minda & Miles, 2010; Nomura et al., 2007; Gureckis et al., 2011). Synthesizing across these existing lines of evidence, we expect the following cortical regions of interest: Explicit categorization would be associated with DLPFC activity, BA 9, 46 (Aizenstein et al., 2000; Reber et al., 2003; Gureckis et al., 2011; Helie et al., 2010; Wu et al., 2020; Nomura et al., 2007). Implicit categorization would be associated with SMA/PMd activity, BA 6 (Aizenstein et al., 2000; Reber et al., 2003; Gureckis et al., 2011; Soto et al., 2013; Helie et al., 2010). Indeed, these are both critical regions specified as belonging to distinct circuitry by COVIS (Ashby et al., 1998).

### 1.3 Functional Near-Infrared Spectroscopy

Continuous wave functional near-infrared spectroscopy (fNIRS) is an optical based neuroimaging technique to quantify changes in concentration of haemoglobin (Hb) species as a function of time (Hoshi & Tamura, 1993). Continuous wave fNIRS is incapable of measuring absolute concentrations, thus inferences of neural activity must be made on the basis of concentration changes in oxyhaemoglobin (HbO<sub>2</sub>) or deoxyhaemoglobin (HbR) between two timepoints (conditions or trials) (Quaresima & Ferrari, 2019). To perform measurements, source and detector optodes are arranged on the scalp in a grid-like fashion forming measurement channels—parabolic volumes, the troughs of which penetrate 5-8 mm into superficial cortex (Pinti et al., 2020).

Two or more wavelengths of near-infrared light flanking the isosbestic point of HbO<sub>2</sub> and HbR are selected for measurement; these wavelengths are confined to the near-infrared window of biological tissue (Smith, Mancini, & Nie, 2009). Wavelengths ranging from 650 to 1350 nm can maximally penetrate human tissue including skin, bone, and blood (Smith et al., 2009). To obtain measurements, a predetermined intensity of light is emitted from the source. Adjacent detectors spaced at approximately 3 cm record the optical density of light post-transmission through the cortex and superficial tissue (Pinti et al., 2020). Optical densities are converted to Hb concentrations via path-length

adjusted Beer-Lambert law. Finally, relative changes in Hb concentrations between conditions can be used to causally infer neuronal activity by reference to the neurovascular coupling phenomenon (Ogawa, Lee, Kay, & Tank, 1990; Huppert, Hoge, Diamond, Franceschini, & Boas, 2006).

As synaptic activity increases in response to some stimuli or process, metabolic byproducts are released into the local extracellular environment which signal surrounding astrocytes to release NO<sub>2</sub>, thereby causing the smooth muscle (pericytes) to relax and increase local cerebral blood flow (Iadecola, 2017). The full mechanism of neurovascular response, and the theories behind why it exists is beyond the scope of the present work (Iadecola, 2017). However, it has been reliably documented that an increase in neuronal processing is tightly coupled to a characteristic response profile of blood flow known as the haemodynamic response function (HRF) (Ogawa et al., 1990). The application of fNIRS is targeted to detect an HRF in the region of interest as to infer differential activity between experimentally manipulated conditions or trials (Quaresima & Ferrari, 2019).

When applied to cognitive neuroscience, fNIRS makes a different system of trade-offs compared to more typical modalities (fMRI), and therefore adds a unique contribution to active discussions in the field (Ferrari & Quaresima, 2012). The three primary disadvantages of fNIRS is the decreased spatial resolution (3 cm), the constraint of only measuring superficial (8 mm) cortical activity, and the noise component of skin, skull, CSF physiology. The two disadvantages, coarse spatial resolution and shallow cortical measurement appear to be inherent physical limitations, and are unlikely to be significantly improved upon in the near future (Pinti et al., 2020). As for the noise component, recent developments in technique have dramatically reduced the impact of superficial physiology. In particular, the introduction of short-distance (SD) channels, which are no longer than 15 mm of separation intentionally direct the emitted light beam to scatter only through the superficial physiology thus measuring the noise associated with scalp, skull, dura, and CSF, without penetrating cortex (Brigadoi & Cooper, 2015). Within the general linear model framework, the timeseries of short-distance channels can be added as a regressor to factor out the variance associated physiology. The success of SD technique has resulted in its wide proliferation and significantly ameliorates the initially severe disadvantage of employing an fNIRS paradigm (Brigadoi & Cooper, 2015).

At the cost of these aforementioned limitations, fNIRS offers several technically interesting and practical advantages. Technically, fNIRS can measure both HbO<sub>2</sub> and HbR haemospecies which offers a perspective unavailable to typical fMRI paradigms. The BOLD signal primarily measures the decreased HbR concentration locally flushing out as the neurovascular response brings in fresh blood (Ogawa et al., 1990; Huppert et al., 2006). In contrast, fNIRS can directly measure the inbound HbO<sub>2</sub> component as well as the outbound HbR component simultaneously. It has also been demonstrated that fNIRS corroborates fMRI measurements with high fidelity in both motor and prefrontal cortical regions (Sato et al., 2013; Huppert et al., 2006). Further, in the normal physiology of neurovascular response, the concentration changes of the HbO<sub>2</sub> component is of greater magnitude and potentially lends itself to more powerful detection (Pinti et al., 2020). Temporal resolution of fNIRS varies from system and often depend on the specific probe design, but in general fNIRS offers superior temporal resolution with sampling frequency at > 4 Hz compared to the standard 2 s TR time in fMRI, offering only 0.5 Hz sampling rate. This is to say, all else held equal, fNIRS has the potential to capture the HRF more accurately, such as the shape and onset of the response (Poldrack, Mumford, & Nichols, 2011). Practically, fNIRS is more portable, less sensitive to motion artifact, and more accessible in terms of funding and personnel. In general, research paradigms employing fNIRS are more naturalistic (participants are not constrained to supine position), and have the potential to be more ecologically valid. Therefore, fNIRS as an alternative modality may con-



tribute a unique perspective to the debate on the dissociation of category learning systems.

## 1.4 Objective

The primary aims of the present study are twofold. (1) To perform a direct, critical test of COVIS, specifically about the assumptions underlying the implementational level of the theory—that is at the level of functional neuroanatomy (Ashby et al., 1998; Marr, 1982). (2) To assess the feasibility of fNIRS as a modality for investigating categorization and category learning in general.

As direct tests of COVIS theory, I postulate one general hypothesis and two precise hypotheses, all of which are pre-registered on Open Science Framework OSF ([osf.io/8tbua](https://osf.io/8tbua)). Materials, scripts, raw data, figures of the present study are publicly available at OSF (DOI [10.17605, osf.io/c3kav](https://doi.org/10.17605/osf.io/c3kav)).

**H1:** explicit and implicit category learning will exhibit anatomically distinct patterns of functional cortical activity.

**H2:** explicit category learning will exhibit higher activation of dorsolateral prefrontal cortices (DLPFC; BA9, BA46), when compared to implicit category learning.

**H3:** implicit category learning will exhibit higher activation of premotor and supplementary motor areas (PMd/SMA; BA6), when compared to explicit category learning.

# Chapter 2

## Experiment I

The focus of Experiment I is to design, and validate category structures and learning task, fit for eliciting dissociation of explicit and implicit systems as measured by fNIRS neuroimaging. We employed a within-subject design consisting of two experimental conditions where participants learn a Rule-Based (RB) category structure, and subsequently an Information-Integration (II) category structure or vice versa; experimental condition order was counterbalanced across subjects. The RB and II category sets are designed to bias neural activation of either explicit or implicit category learning systems respectively, as specified by COVIS. All study procedures for Experiments 1, 2, 3 were approved by the Non-Medical Research Ethics Board at Western University (REB code: 120100).

### 2.1 Methods

The experimental stimuli used in the present study are classic sine-wave gratings that only vary meaningfully in two dimensions. Both dimensions are verbalizable: the orientation and frequency of the gratings. A total of 150 stimuli per category were generated by sampling from bivariate normal distributions. The RB condition consists of a conjunctive-rule category set in which participants must combine two single-dimensional rules regarding frequency and orientation dimensions of the stimuli (Figure 2.1a). Optimal RB classification in this condition results in a L-shaped boundary separating the two categories. If participants cannot find the optimal rule, their performance will reach a ceiling of 75%, even then, they are more likely to find a 1D rule rather than the 2D implicit boundary. The II condition consists of a 2D-linear boundary with a positive slope separating categories A and B. This optimal II boundary is extremely difficult to explicitly verbalize and favours implicit learning (Figure 2.1b).

To make robust inferences from the neuroimaging data, the behavioural task must only differ in the optimal categorization strategy while controlling for all other possible confounding variables of the category sets. The primary aim is to dissociate neural activity of explicit and implicit systems by use of RB and II experimental conditions, respectively.

Three possible confounding variables require stringent controls (Carpenter et al., 2016). First, The mathematical separation between categories; that is the Euclidean distance between category centroids divided by the mean Euclidean distance of individual exemplars from its respective category centroid along the direction across the comparison boundary. Separation is an important control variable to prevent brain activity from being ambiguously attributed to differences in perceptual

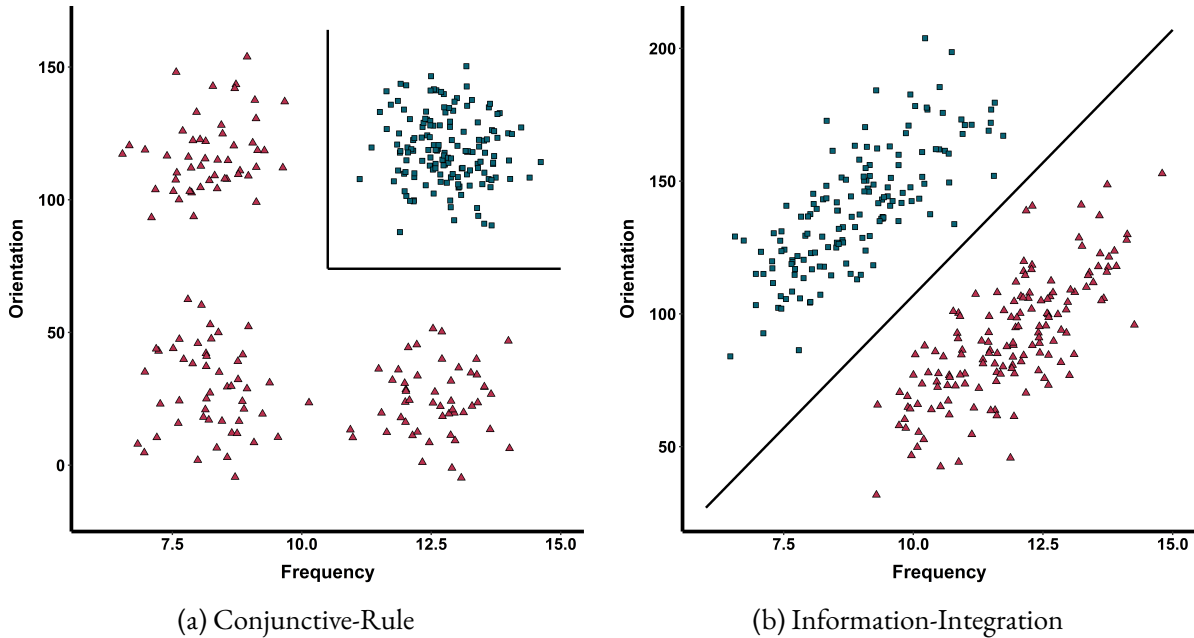


Figure 2.1: Category structures for the two experimental conditions. Frequency is given in cycles per image, orientation is given in degrees. Each point represents a single sine-wave grating stimuli. Squares denote category A, triangles denote category B. The optimal boundaries are shown as lines separating the two categories.

representation of the two category sets. In the present study, the categories are designed with the standardized separation of 8.0 arbitrary units. Second, The number of relevant dimensions. Previous designs used single-dimensional rules which are easier and faster to learn than II category sets. Participants also only require selective attention towards one relevant dimension. Thus, neural activity could be wrongly attributed to differential allocation of attention. The present task replaces the 1D rule with a 2D conjunctive-rule to match the 2D-II boundary. Third, the task difficulty, if different between conditions may result in differential recruitment of cognitive resources which confounds the primary aim of dissociating neural activity of explicit and implicit systems. While mathematical separation and number of relevant dimensions are controlled in the design of category sets, the task difficulty must be evaluated by empirical performance. The present behavioural pilot employs category sets which controls for (1) the mathematical separation, (2) number of relevant dimensions and empirically validates whether, (3) participant performance is matched in both RB and II conditions.

Participants viewed images of sine-wave grating on a computer screen and responded by pressing 'r' or 'o' on the keyboard to classify categories A and B respectively. Over the course of 15 minutes per condition, 300 trials with correct/incorrect feedback were administered for a fixed interval of 3 seconds. In each trial, the stimuli was presented for a maximum of 2 seconds. If participants respond within the 2 seconds, feedback is immediately presented for the remaining time summing to 3 seconds per trial. If participants do not respond within the 2 second stimuli presentation, the feedback "too slow!" is shown for exactly 1 second. Including both RB and II conditions, the total runtime of the behavioural task is 30 minutes exactly. These design choices were such that the paradigm could be easily converted for a later neuroimaging experiment (Experiment 3). We collected data from ( $n = 119$ ) participants, of which 11 were removed for chance performance, leaving ( $n = 108$ ) participants

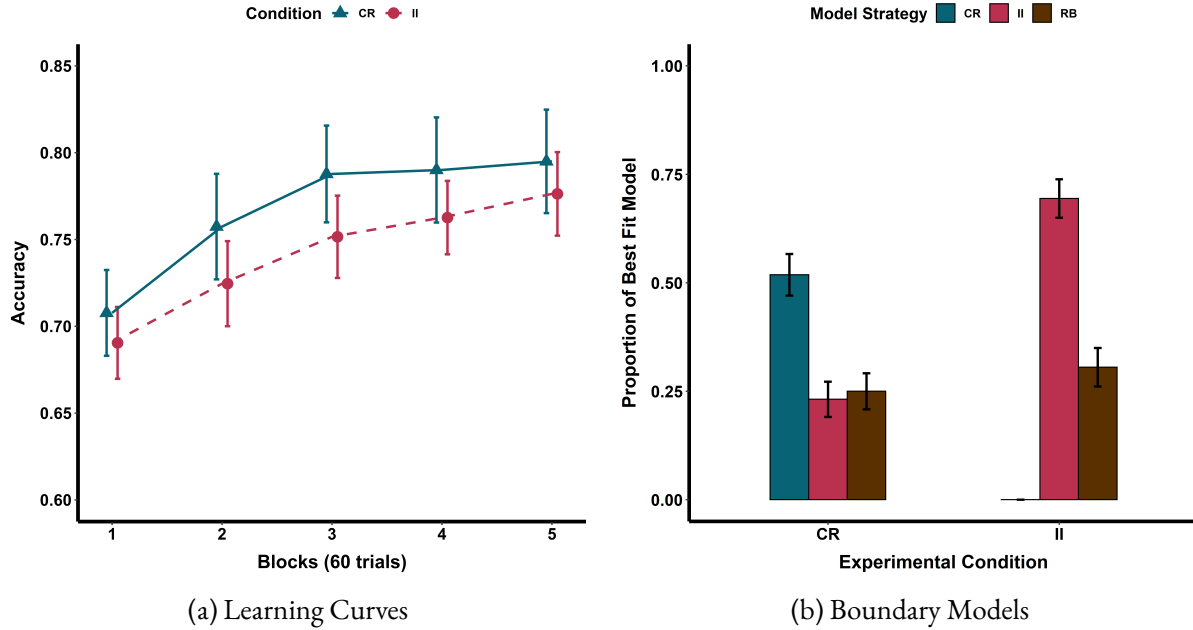


Figure 2.2: Experiment 1 Data. (a) Rule-based (RB) and Information-Integration (II) learning curves. Blue triangle solid line indicates RB category set, red circle dashed line indicates II category set. Error bars denote 95% confidence intervals of the mean. (b) Proportion of participants best fit by RB and II boundary models for respective experimental condition. CR indicates the proportion best fit by 2D general conjunctive classifier, II proportion best fit by 2D general linear classifier, RB proportion best fit by 1D linear classifier. Error bars denote standard error of proportion estimates.

in the analyses. Participants were recruited from Western University’s internal research recruitment pool. The behavioural task was designed in PsychoPy and administered online through pavlovia.org.

## 2.2 Results

As mentioned in the methods, the mathematical separation of categories and number of relevant dimensions were controlled in task design. However, task difficulty is evaluated empirically. A  $5 \times 2$  (block  $\times$  condition) mixed-design analysis of variance (ANOVA) revealed main-effect of block,  $F(3, 353) = 49.78, p < .001, \eta^2 = .051$ , and main-effect of experimental condition,  $F(1, 107) = 6.67, p < .001, \eta^2 = .009$ . Pairwise comparisons showed higher accuracy in final block compared to initial block for both Information-Integration,  $t(107) = 7.25, p < .001$ , and Rule-Based,  $t(107) = 7.67, p < .001$ , conditions indicating that participants successfully learned both category structures. Direct comparisons of RB and II conditions for each block initially revealed higher RB performance in block 2 and block 3, but these effects did not survive Holm-Bonferroni correction for multiple comparisons (Figure. 2.2a). These behavioural data empirically validate our effort to control for the confound of task difficulty. Participants were capable of learning both category structures and the overall performance did not differ between RB and II conditions.

Participants responses were also fit to General Recognition Theory (GRT) classifiers. The II condition was designed for optimal fit by 2D general linear classifier. The RB condition consists of a con-

conjunctive rule and is thus best fit by the general conjunctive classifier. However, in both conditions, participants may achieve above chance categorization by using single-dimensional rules and thus 1D linear classifiers were fit in both orientation and frequency dimensions. Optimal model fits were evaluated by Akaike's Information Criterion (AIC). The proportion of participants using the optimal strategy in congruence with the experimental condition was  $69.4 \pm 4.4\%$  for the II set and  $51.9 \pm 4.8\%$  for Conjunctive RB set (Figure 2.2b). Though not necessarily finding the optimal conjunctive-rule,  $76.8 \pm 4.5\%$  of participants in the CR condition applied some form of RB strategy.

# Chapter 3

## Experiment II

The purpose of Experiment 2 is to adapt the previously validated category structures from Experiment 1 into a suitable fNIRS protocol. Specifically, blocks of rest, and control tasks were introduced in the latter trials of category learning. The aim is to validate if the tightly controlled category structure from Experiment 1 are robust to the addition of necessary conditions for neuroimaging.

### 3.1 Methods

The category learning condition from Experiment 1 (CATG) was adapted to a blocked design (Figure 3.1). Two additional conditions were added, a control (CTRL) condition and a rest (REST) condition. The REST condition presents a fixation cross where participants are instructed to do nothing, and the duration is jittered to be one of (12, 14, 16, 18, 20) seconds as to eliminate synchronicity of noise and experimental condition frequencies. The CTRL condition has been designed to mimic the category learning condition as closely as possible with only two differences. One, the stimulus set and appropriate responses are already known and not meaningfully learned; two, the response decision is a known to participants a priori, and simple enough as to not meaningfully engage the RB category system (Carpenter et al., 2016; Stark & Squire, 2001). During CTRL, participants view uniform noise figures of 'r' or 'o' contrasted against background of different noise density and must respond by pressing the respective key upon which they will receive correct/incorrect feedback. In this way, participants are not learning new information unlike the CATG condition. Also the task is simple enough as not to elicit RB activation, it is closer to a detection problem rather than a categorization problem.

The first 9 minutes of the experiment is identical to Experiment 1, and consists of 180 uninterrupted 3 s trials of category learning to allow participants sufficient time to develop familiarity with the learning task. When transitioning into the imaging phase, a 20s block of rest is introduced. In the remaining 11 minutes, CTRL and REST conditions are interleaved with the original CATG task. REST blocks are jittered to have variable duration's and are presented exactly in the following order: (20, 20, 12, 18, 14, 16, 16, 14, 18, 20, 12) s. The conditions are presented in the following order: 24 trials x 2 s = 48 s of CATG, jittered REST, 24 trials x 2 = 48 s of CTRL, jittered REST. This block order constitutes one cycle of 128 s on average. The cycle is repeated 5 times for a total of 11 minutes.

Experiment 2 tests whether the results of Experiment 1 are robust to the addition of interleaved CTRL and REST blocks. Online participants ( $n = 72$ ) were recruited from the university's research

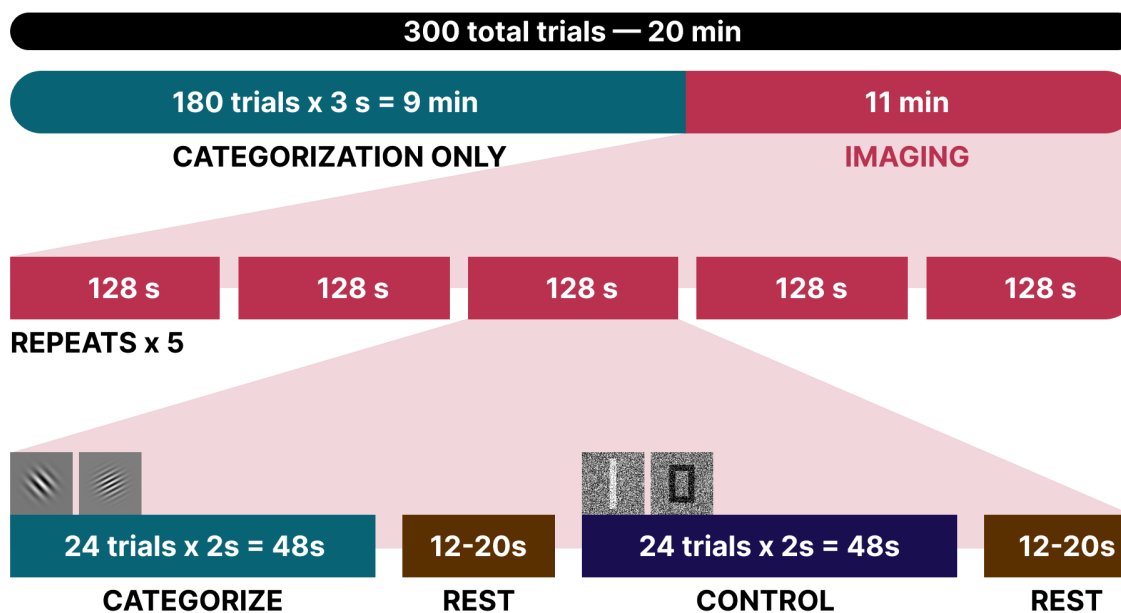


Figure 3.1: Learning and imaging protocol for experiments 2 and 3. In Experiment 2 the design was between-subjects so each participant completed 20 minutes of either RB or II conditions. In experiment 3 the design was within-subjects so the same participant completed 20 minutes of both RB and II counterbalanced for a total scanning time of 40 minutes per participant.

pool to perform either one of RB or II versions of the task, of which 14 were removed for chance responding, leaving ( $n = 58$ ) for the present analyses.

## 3.2 Results

Repeated-measures  $5 \times 2$  (block  $\times$  condition) mixed-design ANOVA revealed main-effect of block  $F(3, 143) = 25.1, p < .001, \eta^2 = .17$ , and no effect of condition. Pairwise comparisons showed higher accuracy in final block compared to initial for both RB  $t(31) = 6.04, p < .001$ , and II  $t(25) = 4.11, p = .004$  conditions. These results indicate that learning performance on the category sets designed in Experiment 1 are robust to the addition of interleaved CTRL and REST conditions.

GRT models showed that within the II condition,  $70 \pm 4.4\%$  were best fit by II strategy, and  $30 \pm 4.4\%$  used RB strategy. Within the RB condition,  $36 \pm 4.6\%$  of participants were best fit by the optimal 2D CR strategy,  $14 \pm 3.4\%$  were best fit by the suboptimal 1D RB strategy, and  $50 \pm 4.8\%$  were best fit by II strategy. This significant incongruence between categorization strategy and experimental condition in Experiment 2 indicates that even tightly controlled designs of category set may not robustly elicit the appropriate optimal strategy in participants and therefore must be accounted for in the analyses of neuroimaging data.

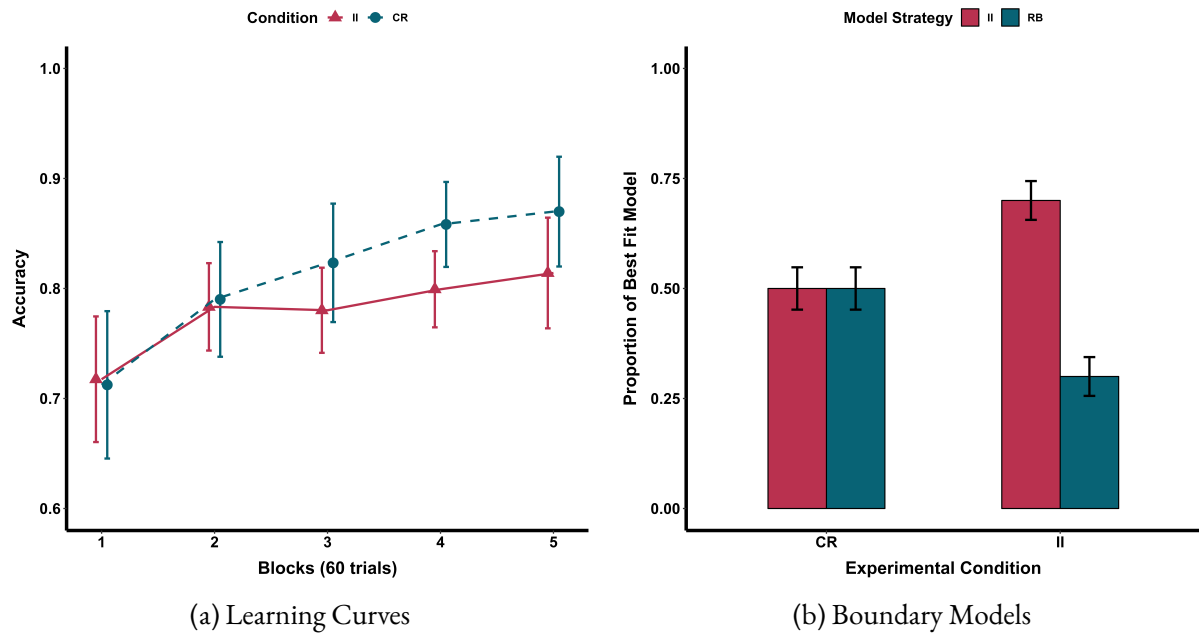


Figure 3.2: Experiment 2 Data. (a) Rule-based (RB) and Information-Integration (II) learning curves. Blue triangle solid line indicates RB category set, red circle dashed line indicates II category set. Error bars denote 95% confidence intervals of the mean. (b) Proportion of participants best fit by RB and II boundary models for respective experimental condition. II proportion best fit by 2D general linear classifier, RB proportion best fit by 1D linear classifier. Error bars denote standard error of proportion estimates.



# Chapter 4

## Experiment III

The focus of Experiment 3 is to acquire functional neuroimaging data associated with the process of learning the empirically controlled category structures reported in Experiment 1 & 2. To this end, we employed optical based fNIRS imaging to measure regional changes in haemoglobin concentrations while participants learned the aforementioned RB and II category structures.

### 4.1 Methods

#### 4.1.1 Design

The RB and II categorization tasks that were finalized in Experiment 2 contained three conditions (CATG, CTRL, REST). The exact tasks were used in Experiment 3, with only a minor modification to send triggers from the experimental computer to the fNIRS hardware (Figure 3.1). We collected data from ( $n = 43$ ) participants which were recruited from an undergraduate internal research pool at Western University, and the local community. The design was within-subjects so the same participant completed both the RB and II task, in contrast to Experiment 1 & 2 where each participant completed exclusively the RB or II task. The duration of each learning task was 20 minutes exactly, the total scanning time per participant was 40 minutes for both tasks. The order of the tasks were counterbalanced such that every subsequent participant completed the tasks in reversed order with respect to the immediately previous participant. However, order effects of a within-subjects design still may remain, and will therefore be addressed in the analyses.

#### 4.1.2 fNIRS Hardware

Source and detector positions were selected to maximize exposure of measurement channels to two cortical regions-of-interest (ROI). These are the DLPFC and PMd/SMA. However, considering the spatial proximity of DLPFC and PMd/SMA and that the stated aim is to dissociate the two, the number of optodes and channels were selected as to maintain a chasm of non-measurement between DLPFC and PMd/SMA ROI. The probe is designed with emphasis on the minimization of cross-talk between channels of different ROI, but at the cost of forgoing maximum coverage. Optimal optode location were determined with the aid of fOLD toolbox, a photon transport simulation-based software tool which generates a specificity-optimized probe layout given user-defined ROI (Morais, Balardin, & Sato, 2018). The DLPFC was specified by Brodmann's areas BA 9, 46; PMd/SMA was

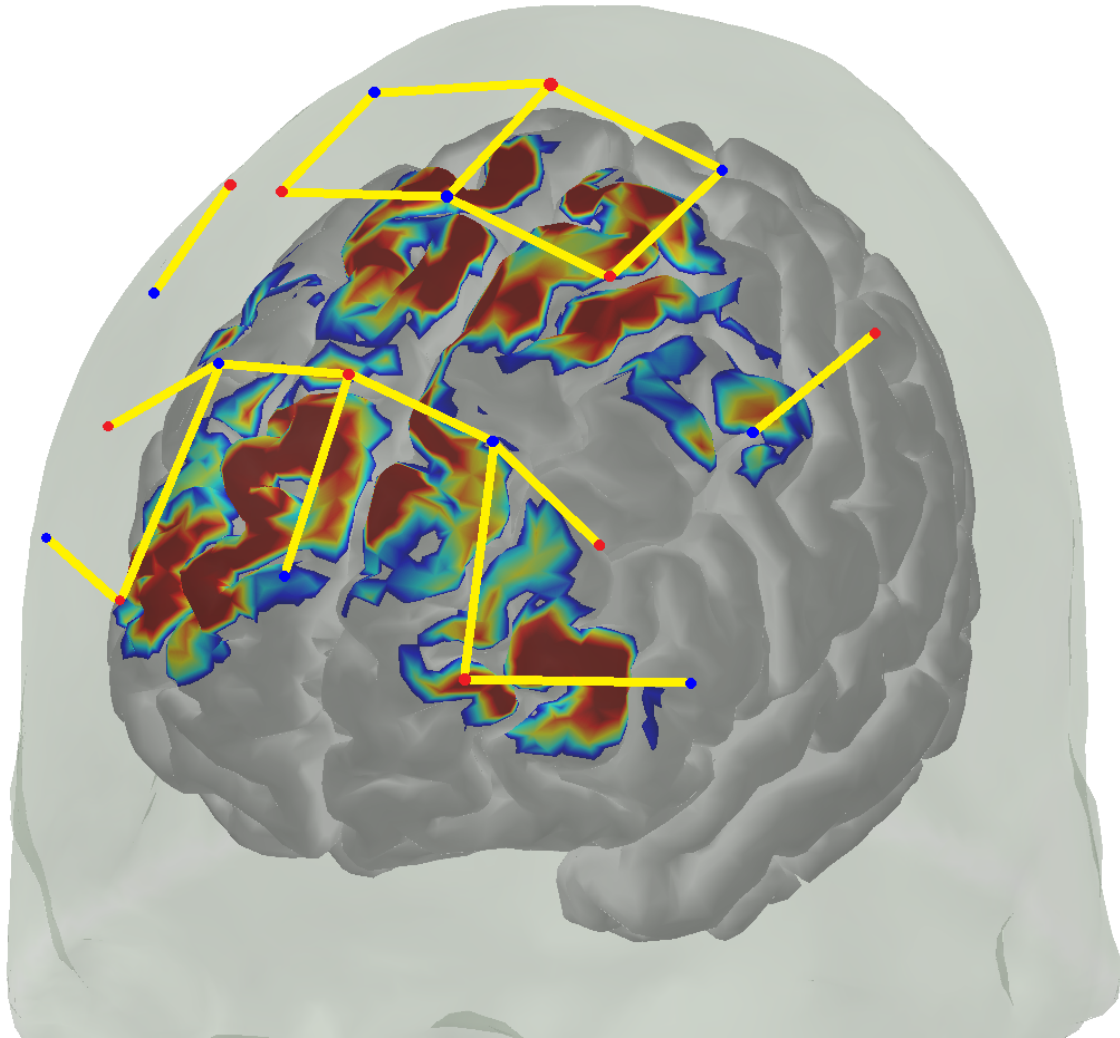


Figure 4.1: Cortical sensitivity of probe layout. Red points are sources, blue points are detectors, yellow lines joining the points represent the measurement channels between optodes. Note that the true measurement channels take the form of a concave volume passing through the cortex. Sensitivity of the probe layout is displayed as heatmap on the surface of cortex ranging from  $0.25 \text{ mm}^{-1}$  (blue) to  $1.0 \text{ mm}^{-1}$  (red)

Table 4.1: Optode Location and ROI Specificity

Channel	Source	Detector	ROI	Side	Specificity (%)
1	Cz	FCz	6	C	83.8
2	FC2	C2	6	R	82.5
3	FC1	C1	6	L	81.8
4	FC1	FCz	6	L	73.2
5	FC2	FCz	6	R	63.0
6	C3	FC3	6	L	61.7
7	C4	FC4	6	R	56.9
8	Cz	C1	6	L	56.5
9	Cz	C2	6	R	55.4
10	Fz	F2	9	R	68.9
11	F4	F2	9	R	68.4
12	F3	F1	9	L	66.6
13	Fz	F1	9	L	63.2
14	Fz	AFz	9	C	61.8
15	AF4	F2	9	R	51.5
16	AF3	F1	9	L	48.4
17	AF3	F5	46	L	49.3
18	AF4	F6	46	R	47.4

specified by BA 6. Retaining channels with  $specificity_{ROI} > 45\%$  resulted in a 10x10 source-detector layout with 18 recording channels (Table 4.1). Locations of sources and detectors are placed in accordance to the 10-20 EEG system. In the AtlasViewer environment, the specified probe layout was inputted into a Monte Carlo photon migration simulation to generate a estimate of measurement sensitivity on the cortical surface (Figure 4.1) (Aasted et al., 2015). The simulation confirms that both of our specified ROI: the DLPFC and PMd/SMA receive adequate coverage by our probe design. To measure physiological noise correlates originating from non-cortical tissue, four short channels were placed at 8mm separation on sources at AF3, AF4, C3, C4. These were selected to be in the corners of the probe as to allow sampling of physiology across the surface of the head. Data were recorded on the NIRScout system from NIRx Medical Technologies. Two-wavelength (760nm, 850nm) LED sources measured HbO<sub>2</sub> concentrations at a sampling frequency of 6.51Hz. Experimental triggers for start of task conditions were sent via Cedrus c-pod device to the fNIRS hardware.

### 4.1.3 fNIRS Analysis

Raw data NIRx datafiles were imported and converted to .nirs and subsequently .snirf files for pre-processing within the Homer3 of the OpenfNIRS environment (Tucker et al., 2023; Huppert, Dia-

mond, Franceschini, & Boas, 2009). For each subject, raw voltages were converted to optical density values. Channels with  $\text{SNR} < 5$  were pruned. Motion artifacts were identified by thresholding  $\text{SD} > 5$ , and masked over a timerange of 0.5 by recursive PCA correction. A low pass filter at 0.5 Hz was applied before converting optical density to concentration via Modified Beer-Lambert's Law without partial path length adjustment (1.0) for both wavelengths. Concentration units are reported in  $\mu\text{M} \cdot \text{mm}$ . Haemodynamic response function (HRF) was extracted by GLM which included the following regressors: highest correlated short-separation channel noise, and consecutive gaussian basis set to model HRF. GLM was solved by iterated-least squares (Barker, Aarabi, & Huppert, 2013). Subject-level statistics for experimental conditions were exported for group-level analysis in R.

Subject-level t-contrasts were performed on extracted betas, collapsed across all subjects for both CR and II conditions. Two contrasts served to discern neural effects of category learning (CATG > REST), (CATG > CTRL), and one contrast served as a positive control, (CTRL > REST). The betas representing task-related activation is tested against 0 to determine general relative activation of the categorization task. In the case of (CATG > REST) the difference in betas is tested against 0. Multiple comparisons were adjusted for false-discovery rate (FDR).

To analyze categorization conditions of CR and II, full mixed-effects models were computed channel-wise at the group level including the following fixed effects: condition, order, z-scored performance, best fit GRT boundary, and participant as a random effect. Model fits were backwards tested by sequentially removing fixed effects, and computing AIC to optimize the parsimonious model explaining variance in haemodynamic response. Subject level contrasts (CATG > REST) represented the dependent haemodynamic response. Since the only difference of CATG between CR and II conditions is the stimuli set learned, and all other parameters were equivalent, there is no need to account for the (CATG > CTRL) contrast at the group level. Multiple comparisons were adjusted for FDR.

For GRT boundaries, participants were partitioned as RB if their responses were best fit by either a single-dimensional general linear classifier (1D-RB), or a conjunctive-rule (2D-CR), otherwise if they were partitioned as II if they were best fit by a two-dimensional general linear classifier (2D-II).

Model statistics were exported and visualized using the visbrain environment in Python 3 (Combrisson et al., 2019). Channels were treated as visbrain source objects localized to estimated MNI coordinates and projected to the surface of the cortical mesh. Projections represent the statistical properties obtained from channels mapped to standard 10-20 EEG locations, not to be confused with interpolation of activation clusters on the cortex.

## 4.2 Results

### 4.2.1 Behavioural

Repeated-measures  $5 \times 2$  (block  $\times$  condition) mixed-design ANOVA on performance revealed main-effect of block,  $F(3, 129) = 30.1, p < .001, \eta^2 = .14$ , main effect of condition,  $F(1, 41) = 33.4, p < .001, \eta^2 = .07$  (Figure 4.2a). Pairwise comparisons showed higher accuracy in final block compared to initial for both RB,  $t(43) = 8.6, p < .001$ , and II,  $t(43) = 4.8, p < .001$ , conditions indicating learning in both categories. Pairwise comparisons of condition show higher performance in RB conditions for blocks 2,  $F(1, 43) = 9.4, p < .001$ , through block 5,  $F(1, 43) = 31.6, p < .001$ . There was a significant interaction effect of block and condition,  $F(4, 164) = 4.6, p < .001, \eta^2 = .02$ , indicating participants learned RB category at a faster rate.

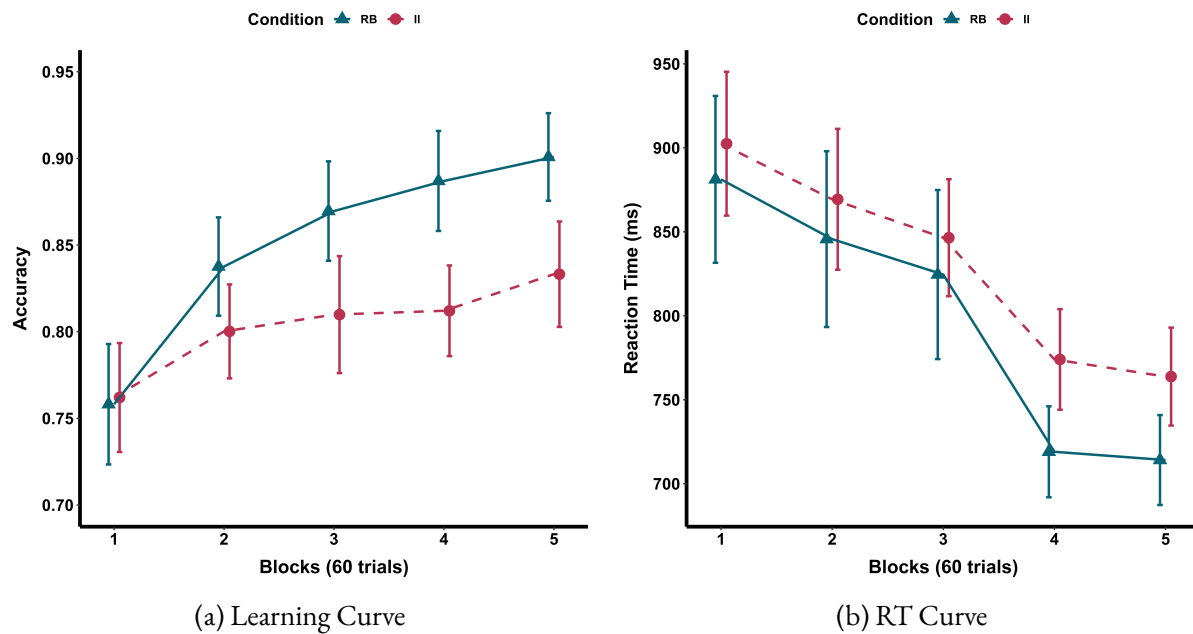


Figure 4.2: Experiment 3 Behavioural Data. (a) RB and II. Blue triangle solid line indicates RB category set, red circle dashed line indicates II category set. Errorbars denote 95% confidence intervals of the mean. (b) Reaction times denoted in milliseconds. Errorbars denote 95% confidence intervals of the mean.

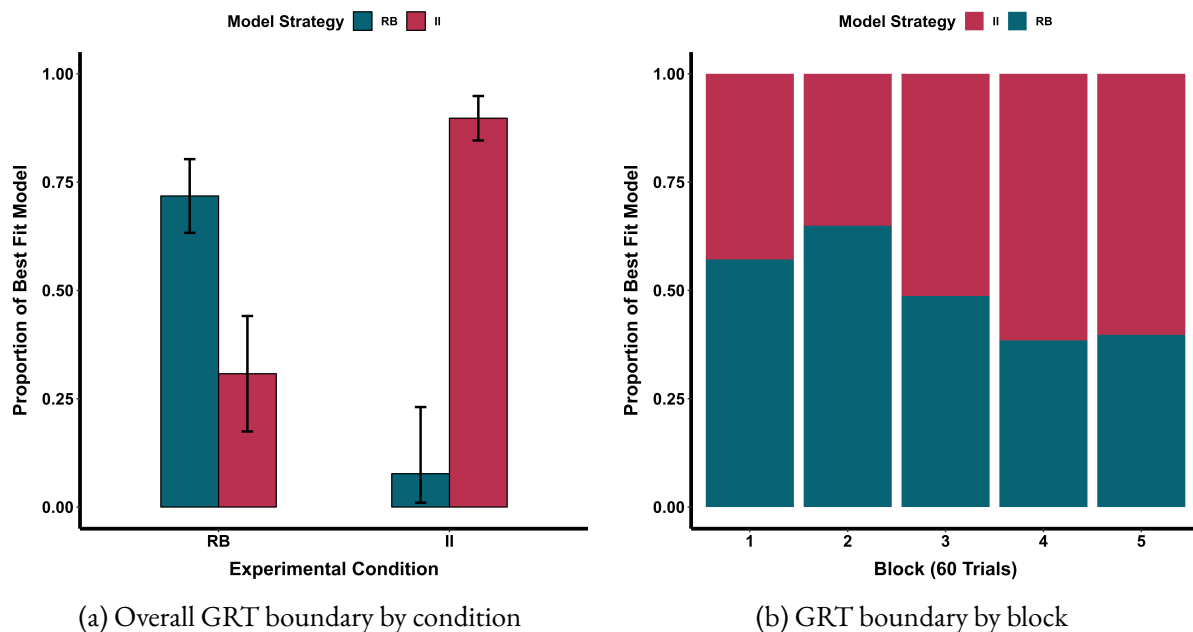


Figure 4.3: GRT modeling results (a) Depicts the best fit GRT boundary by experimental condition. Errorbars denote standard error of the proportion. (b) Depicts the relative proportion of all participants' best fit respective boundary by cumulative block, all proportions sum to one.

Repeated-measures  $5 \times 2$  (block  $\times$  condition) mixed-design ANOVA on reaction time data revealed main-effect of block,  $F(2, 84) = 52.1, p < .001, \eta^2 = .18$ , main effect of condition,  $F(1, 41) = 6.26, p < .001, \eta^2 = .01$  (Figure 4.2b). Pairwise comparisons showed lower reaction time in final block compared to initial for both RB,  $t(46) = 7.2, p < .001$ , and II,  $t(43) = 7.2, p < .001$ , conditions indicating decrease of reaction time in both conditions. Pairwise comparisons of condition showed lower reaction times in RB conditions for blocks 4,  $F(1, 44) = 35.4, p < .001$ , and block 5,  $F(1, 44) = 25.3, p < .001$ . There was no significant interaction effect, indicating the rate of reaction time decrease did not differ by condition.

Overall, performance showed moderate negative correlation with reaction time,  $r(456) = -.34, p < .001$ , indicating the effect of learning—as performance increases, the response time to decision decreases.

GRT models showed that within the II condition,  $90 \pm 5.1\%$  were best fit by II strategy, and  $7.7 \pm 15.4\%$  used RB strategy (Figure 4.3a). Within the RB condition,  $72 \pm 8.5\%$  of participants were best fit by a RB strategy,  $31 \pm 13.3\%$  were best fit by II strategy. Of the participants within the RB condition, implementing a RB strategy,  $84 \pm 2.0\%$  learned the optimal CR boundary. These empirical incongruities between experimental condition and inferred strategy use is nontrivial and must therefore be addressed. By the final block of the experiment and across both conditions, participants exhibited a bias towards using an II bound to perform categorization (Figure 4.3b), the relative proportions of participant's best fit GRT boundary shifts throughout the experiment. Block 2 showed the highest proportion ( $65 \pm 4.2\%$ ) of RB categorization and decreased significantly,  $\chi^2(1, N = 78) = 9.9, p = .002$ , by block 5 ( $40 \pm 4.3\%$ ) suggesting that participants attempted RB strategies by default, but biased towards II strategies ( $60 \pm 4.3\%$ ) by the end of the experiment.

#### 4.2.2 Subject-level fNIRS

Collapsed across all sessions for all subjects. A t-contrast of (CATG > REST) revealed broad activation of the frontal lobe, with 9 of 18 measurement channels passing significance threshold (Table 4.2). Adjustments by FDR were applied channel-wise to contrast statistics using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). These results were corroborated by the positive control t-contrast (CTRL > REST), which revealed 6 of 18 channels as significant, all of which were also active in the (CATG > REST) contrast. These contrasts indicate that both the CATG and CTRL components of the behavioural task are capable of eliciting a broad and similar haemodynamic response relative to rest in PMd/SMA and DLPFC regions bilaterally. These are expected effects for any given cognitively demanding task that also includes a motor response component such as pressing a button. To discern possible effects of categorization in and of itself, and unrelated to the motor and perceptual aspects of responding to the behavioural task, we performed a direct t-contrast of (CATG > CTRL), and found two significant channels. Both channels were localized to R DLPFC and showed significantly decreased haemodynamic response in CATG relative to CTRL. The direction of the result is unexpected but nonetheless suggests differential activity due to cognition of categorization, as compared to the control task which endeavors to hold all other extraneous variables constant.

#### 4.2.3 Group-level fNIRS

We performed mixed-effects modeling on subject level betas from (CATG > REST) contrasts. Categorization condition, participant performance, and best fit GRT bound, were modeled as fixed-effects

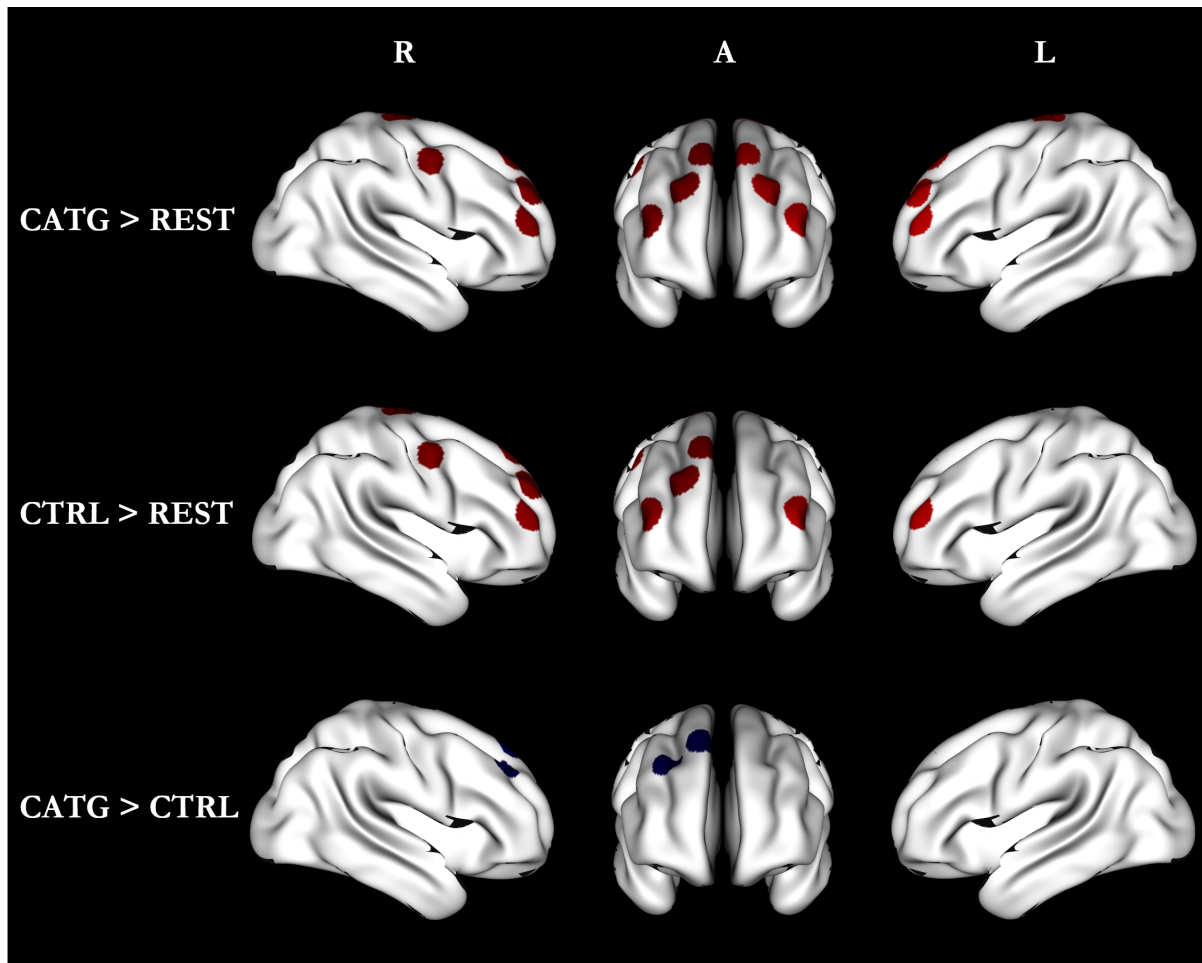


Figure 4.4: Projection of fNIRS channel activation for task condition contrasts. Red represents relative activation, blue represents relative deactivation. Images are thresholded displaying only channels with FDR-adjusted p-values  $< .05$  (For full statistics, see Table 4.2) Each disc projection represents the estimated trough of a measurement channel and spanning 20 mm in diameter. 'A' denotes anterior view.

Table 4.2: Subject-Level Contrasts. Coordinates denote estimated MNI coordinates based on optode placements according to standard 10-20 EEG positions. BA denote Brodmann’s areas.

Channel	x	y	z	ROI	BA	Side	t	p	$p_{FDR}$
CATG > REST									
10	10	41	50	DLPFC	9	R	3.81	<.001	.004
15	22	52	33	DLPFC	9	R	3.22	.002	.010
18	40	50	16	DLPFC	46	R	3.20	.002	.010
7	52	-4	48	PMd/SMA	6	R	3.00	.004	.015
9	17	-21	75	PMd/SMA	6	R	2.89	.005	.016
16	-23	52	32	DLPFC	9	L	2.46	.016	.036
8	-17	-20	74	PMd/SMA	6	L	2.42	.018	.036
13	-9	41	50	DLPFC	9	L	2.43	.018	.036
17	-39	50	17	DLPFC	46	L	2.32	.022	.040
CTRL > REST									
10	10	41	50	DLPFC	9	R	4.09	<.001	.002
7	52	-4	48	PMd/SMA	6	R	3.51	<.001	.005
15	22	52	33	DLPFC	9	R	3.43	<.001	.005
9	17	-21	75	PMd/SMA	6	R	3.32	.001	.05
18	40	16	17	DLPFC	46	R	2.85	.005	.018
17	-39	17	17	DLPFC	46	L	2.43	.017	.046
CATG > CTRL									
10	10	41	50	DLPFC	9	R	-3.66	<.001	.007
11	30	40	41	DLPFC	9	R	-3.29	.002	.012

of interest. Order effects were modeled as a nuisance fixed-effect. With respect to GRT modeling, participants that were best fit by any 1D-rule or the optimal CR rule were classified as RB learners, participants that were best fit by GLC were classified as II learners. Participants were modeled as random effects. Two effects survived corrections for multiple comparisons. Channel 11, which corresponds to R DLPFC, BA 9 showed a significant effect of best fitting GRT bounds (Figure 4.5a). Participants showed significantly higher haemodynamic response when best fit by RB bound,  $\beta = 12.7$ ,  $95\%CI[5.4, 20.0]$ ,  $t = 3.74$ ,  $p = .001$ ,  $p_{FDR} = .020$ , relative to participants that were best fit by II bound (Figure 4.5b). Channel 18, corresponding to R DLPFC, BA 46 revealed an effect of task performance (Figure 4.6a). Z-scored performance exhibits a significantly negative relationship to haemodynamic response,  $\beta = -10.1$ ,  $95\%CI[-4.5, -15.7]$ ,  $t = -3.54$ ,  $p < .001$ ,  $p_{FDR} = .011$ , during categorization relative to rest (Figure 4.6b). We did not find any significant effects due to the RB/II experimental conditions.

These data support our broad hypothesis **H<sub>1</sub>**: explicit and implicit category learning will exhibit anatomically distinct patterns of functional cortical activity; and our narrow hypothesis **H<sub>2</sub>**: Explicit



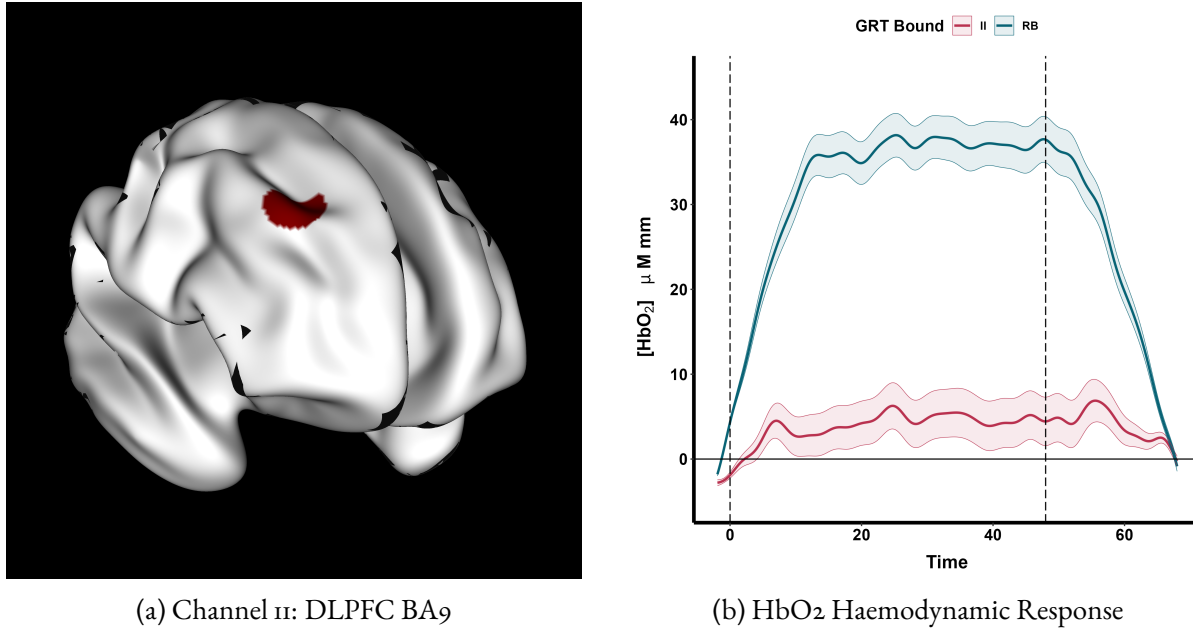
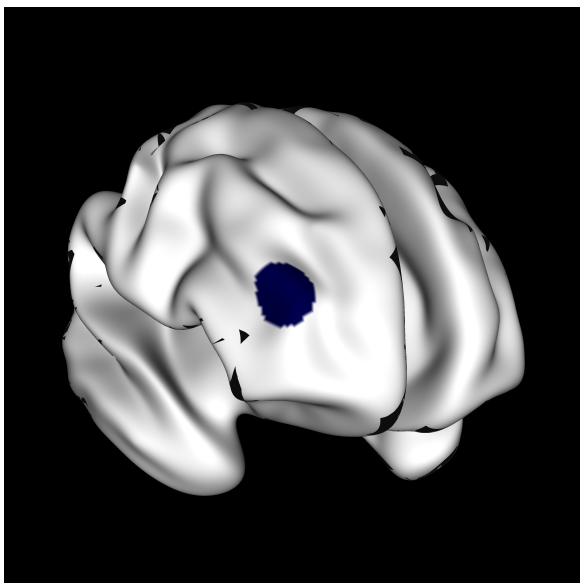


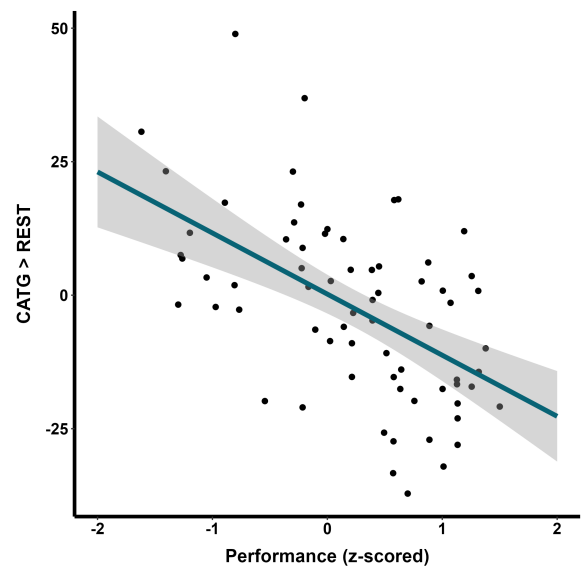
Figure 4.5: (a) Channel II of R DLPFC exhibited a significant effect of categorization strategy as best fit by GRT modeling. (b) Average estimated HbO<sub>2</sub> haemodynamic response during CATG blocks for participants partitioned as RB or II learners by best fitting GRT models. Dotted lines indicate onset and end of CATG trials. Band indicates standard error of mean haemoglobin concentration.

category learning will exhibit higher activation of DLPFC; BA9, BA46, when compared to implicit category learning. The strategy effect in channel II of DLPFC; BA9 clearly demonstrate higher response to explicit category learning of a RB boundary relative to implicit category learning of an II boundary. However, the present data fails to support our narrow hypothesis **H3**: Implicit category learning will exhibit higher activation of PMd/SMA; BA6, when compared to explicit category learning. No significant differences in neural response were localized to either hemisphere of the PMd/SMA ROI.

We also attempted to fit mixed effects models by partitioning participants into three levels; 1D-RB, 2D-CR, 2D-II, but these models did not explain a significantly increased proportion of HRF variance compared to the two-partition (RB, II) model, and were therefore abandoned.



(a) Channel 18: DLPFC BA46



(b) HbO<sub>2</sub> Haemodynamic Response

Figure 4.6: (a) Channel 18 of R DLPFC exhibits significant negative relationship between performance and haemodynamic response. (b) Z-scored performance plotted against haemodynamic response. Band denotes standard error of the estimate.

# Chapter 5

## Discussion

The two primary aims of the study were: (1) To dissociate brain activity of explicit and implicit category learning as specified by COVIS theory. (2) To ascertain the feasibility of fNIRS as a neuroimaging modality for investigating category learning generally.

The present study attempted to dissociate neural activity of COVIS specified Explicit and Implicit category learning systems via neuroimaging. I designed and validated a tightly controlled behavioural task in Experiment 1, then modified it to be compatible with an imaging protocol in Experiment 2. Finally, in Experiment 3 I employed fNIRS to evaluate the activity of DLPFC and PMd/SMA. I did not observe a double dissociation but found a single dissociation in the R DLPFC, showing higher activation in participants best fit by GRT RB > II boundaries. Further, I localized an negative relationship between performance and haemodynamic response in the R DLPFC.

### 5.1 Neural Dissociation

The fixed effects of best fit GRT decision boundary supports one of the narrow hypotheses: in the DLPFC, RB category learning elicits higher neural activation relative to II category learning. Such an result would be expected in the context of COVIS theory, which specifies only the involvement of DLPFC in explicit category learning and not implicit category learning. My data of RB > II contrast corroborates previous work on the topic (Helie et al., 2010; Nomura et al., 2007; Soto et al., 2013; Aizenstein et al., 2000; Gureckis et al., 2011; Morgan et al., 2020; Wu et al., 2020).

However, these results fail to replicate some findings by Carpenter and colleagues, which share the most similar experimental design as the present work and reports no significant dissociation between RB and II category learning conditions (Carpenter et al., 2016). In all other aspects of the experiment, our designs are very similar, as I have modeled many of the controls after those specified in the Carpenter's publication. Both of our studies failed to find an effect of a experimental condition alone and I share their sentiment that the lack of effect is due to the discrepancy between the experimental condition and the actual categorization boundary learned by the participant (Carpenter et al., 2016). Experiments 1, 2, and 3 consistently present a nontrivial incongruency between the learned strategy and the experimental condition, a frequent result in studies that model the decision boundary (Ashby, 2014; Hélie et al., 2017; Donkin, Newell, Kalish, Dunn, & Nosofsky, 2015). Carpenter's study, like ours, also modeled haemodynamic response by best fitting GRT models but failed to find a dissociation between participants utilising RB and II boundaries. We offer the following two ex-

planations for this discrepancy. (1) Carpenter's analysis only compared participants who were best fit by CR against participants that were best fit by II and therefore excluded some of the individual variation in neural activity from participants who learned a suboptimal 1D-rule. In the context of COVIS, explicit and implicit learning is mediated by discrete anatomical circuitry and would predict that even 1D-rule use should elicit dissociable neural activity from implicit categorization (Ashby et al., 1998; Minda & Miles, 2010). Indeed a 1D-rule was previously reported to have elicited significantly dissociated activity, which initiated the discussion on the present issue (Nomura et al., 2007). By discarding participants who were best fit by 1D-rule in the analysis, Carpenter decreased the degrees of freedom and therefore the statistical power in the comparison (Carpenter et al., 2016). (2) Carpenter's study used fMRI as the imaging modality which is based on the decreased concentration of HbR in response to task onset. In comparing modalities, it has been well documented that BOLD signal in fMRI is most greatly correlated with decreased relative HbR concentration as measured by fNIRS (Huppert et al., 2006; Sato et al., 2013). Further, relative changes in  $HbO_2$  concentrations tend to have a larger magnitude than corresponding relative HbR concentration changes (Pinti et al., 2020; Huppert et al., 2006; Sato et al., 2013). Therefore, Carpenter's report of a null-effect of categorization strategy could reflect the intrinsic property of the BOLD signal only indirectly measuring  $HbO_2$  through the quantification of HbR, whereas fNIRS measures  $HbO_2$  chromophore concentrations directly (Pinti et al., 2020; Huppert et al., 2006; Carpenter et al., 2016).

I reported extensive activation of bilateral DLPFC and PMd/SMA at subject-level contrasts of CATG > REST across all categorization conditions. These results are in broad agreement with almost all neuroimaging studies on category learning or any tasks involving sustained attention and recruitment of working memory (Owen, McMillan, Laird, & Bullmore, 2005; Helie et al., 2010; Aizenstein et al., 2000; Nomura et al., 2007; Carpenter et al., 2016; Sato et al., 2013; Gureckis et al., 2011; Soto et al., 2013; Morgan et al., 2020; Wu et al., 2020). My positive control contrast of CTRL > REST revealed a similar though less extensive activation across the same regions. Unexpectedly, the CATG > CTRL revealed a decreased relative activation of CATG in two channels localized to the DLPFC. I suspect the inclusion of the positive control condition in conjunction with the experimental design introduced an unexpected task-switching and response inhibition effect in the CTRL condition thereby explaining the relative lesser activation of the main CATG task (Kane & Engle, 2002; Hyafil, Summerfield, & Koehlin, 2009). The imaging protocol specifies that participants were given 120 trials of uninterrupted categorization to familiarize themselves with the categorization task, though they were provided with instructions prior to the experiment and were aware of the alternate control condition, the CTRL trials are likely to be perceived as novel upon first introduction after trial 168 of CATG. Moreover, participants may come to view CATG as the main task—rightfully so, and consequently view CTRL as a distractor thereby compelling participants to maintain as much of the previously learned boundary information whilst concurrently attending to the CTRL task in a perfunctory manner. This mode of task switching has been documented to tax working memory and therefore activate DLPFC relative to maintaining a constant performance across continuous uniform trials (Kane & Engle, 2002; Hyafil et al., 2009; Wang, Zhu, Rehman, & You, 2020). In any case, the addition of a positive control was important for the present study since the stated aim is to apply a different modality (fNIRS) than the majority of past work on the subject matter. However, I recommend that a positive control in the form of a control task may be unnecessary for blocked designs in future work, since the comparison of RB and II conditions will have already accounted for all of the same perceptual and motor confounds on account of the fact that it would be impossible to directly interleave RB and II category learning within the same run.

## 5.2 Performance Haemodynamics

We reported a negative relationship between task performance and neural response in R DLPFC 4.6b. This effect was not predicted a priori and was therefore not a pre-registered hypotheses. Experiments 1 and 2 made efforts to control for task performance between conditions, but failed to preserve congruent performance in Experiment 3. Therefore, performance was included as a factor in the mixed-effects model. The simplest explanation of the negative relationship between performance and neural response is due to decreased processing time (Poldrack, 2000; Kelly & Garavan, 2005). This view is supported by the reaction time data, which shares a moderate negative correlation with the overall performance. This particular confluence of improved performance over time, decreased neural response, and decreased reaction time has been previously reported by an investigation of visuospatial WM which localized a negative relationship between task performance and neural response to DLPFC, BA46 (Garavan, Kelley, Rosen, Rao, & Stein, 2000). An fNIRS based, load-dependence working memory study found that relationship between task performance and neural response varied as a function of working memory load (Meidenbauer, Choe, Cardenas-Iniguez, Huppert, & Berman, 2021). They reported that HbR response magnitude held a negative linear relationship to task accuracy in a high working memory load condition, which corresponds to a positive relationship between neural response and task accuracy (Meidenbauer et al., 2021; Sato et al., 2013; Huppert et al., 2006). However, the relationship was reversed in the low working memory load condition, HbR response concentration held a positive relationship with task performance, and therefore indicates a negative relationship between task performance and neural response, congruent with the observation of the present report (Meidenbauer et al., 2021; Sato et al., 2013; Huppert et al., 2006). A particular feature of the presently observed negative relationship between performance and neural response is that it is agnostic to categorization condition or strategy since there were no significant interaction effects with condition. Therefore, implicit learners who are not recruiting significant Working memory resources are also contributing to this effect. Indeed, work in visuomotor associative learning of arbitrary stimuli, which is highly analogous to visual implicit category learning, has demonstrated this negative performance relationship localized to the DLPFC (Toni, Ramnani, Josephs, Ashburner, & Passingham, 2001; Ashby et al., 1998).

While there is precedence in the aforementioned literature describing the negative relationship between performance on working memory type tasks and neural response in the DLPFC, there is also a substantial literature suggesting a generic effect in the reversed direction—that task performance should positively correlate neural response in DLPFC (Owen et al., 2005; Di Rosa et al., 2019; Mencarelli et al., 2019; Holmes et al., 2019; Webler et al., 2022; Ogawa, Kotani, & Jimbo, 2014). Generally, working memory load is associated with increased performance, it is assumed that higher performing individuals on working memory tasks are capable engaging a higher working memory load (Owen et al., 2005; Mencarelli et al., 2019; Ogawa et al., 2014). Further, interventions aimed at increasing activity of DLPFC: transcranial direct current stimulation, transcranial magnetic stimulation, and transcranial infrared laser stimulation have been reported to improve task performance on working memory type tasks (Webler et al., 2022; Di Rosa et al., 2019; Holmes et al., 2019; Blanco, Saucedo, & Gonzalez-Lima, 2017). Two of these studies validated stimulation driven activity by concurrent imaging (Webler et al., 2022; Di Rosa et al., 2019). However the majority of these studies are specifically focused on working memory load, and do not necessarily account for influence of learning effects over the course of an experiment (Kelly & Garavan, 2005; Poldrack, 2000).

The discrepancy between the presently observed negative relationship and some of the working

memory literature suggests that there may be another explanation unique to category learning which accounts for this effect. Participants are defaulting to a RB strategy early on, thus engaging working memory resources, then gradually abandoning the RB strategy for the alternative II strategy which concurrently reduces working memory load and increases overall performance (Ashby et al., 1998). It is well documented that there are individual differences in the default categorization strategy as well as the duration in which participants attempt their default strategy before switching to an alternative (Ashby et al., 1998; Smith & Minda, 2002; Le Pelley, Newell, & Nosofsky, 2019; Nosofsky, 1986; Hélié et al., 2017). Regardless of stimuli type, the majority of participants default to an explicit rule-based strategy, and incrementally adjust towards the optimal boundary as they are presented with more trials (Ashby et al., 1998; Hélié et al., 2017). Indeed, GRT modeling by block in the present study show an early preference for RB categorization which peaks by block 2, and decays towards a bias for implicit categorization by block 5 (Figure 4.3b). Notably, default categorization strategy may be a function of culture, though the influence is unreliable and sparsely documented, and doesn't confound the present data since it was broadly sampled entirely within a Western context (Norenzayan, Smith, Kim, & Nisbett, 2002; Klein, 2005; Murphy, Bosch, & Kim, 2017). Taken together, the negative relationship between task performance and DLPFC activity may be capturing the default tendency to initiate categorization with a suboptimal RB strategy, thereby explaining low performance and high DLPFC response, and the gradual adjustment to an optimal strategy, explaining the higher performance and lower DLPFC activity. Astute readers may notice that most learners within the RB condition successfully learned the CR boundary, and would therefore still be using a RB strategy by block 5 and thereby eliciting DLPFC activity. However, the learning curve (Figure 4.2a) shows that the mean accuracy has exceeded the performance ceiling (.75) of suboptimal 1D rule by block 2. Therefore, successful participants in the RB condition have already discovered the underlying optimal CR rule by block 2, and subsequent blocks reflect the refinement in the precise position of the decision bound and not the initial resource-intensive RB learning process. It has been documented that explicit learning tends to exhibit a step-wise discrete increase in classification accuracy once the underlying rule has been discovered (Smith et al., 2004; Minda & Miles, 2010). In sum, the cognitive process incurring the highest recruitment of working memory as specified a priori by COVIS—the formulation and hypothesis testing of rules—has been abandoned by participants who successfully discover the CR bound, but penalizes participants' performance who continue applying the suboptimal 1D bound, thereby explaining the negative relationship between task performance and DLPFC activity (Ashby et al., 1998; Minda & Miles, 2010; Meidenbauer et al., 2021).

### 5.3 Multiple Comparisons

One major limitation of the present study is that there are several other interesting effects that failed to survive corrections for multiple comparisons (See Appendix Af2a, Af2b), and were therefore not reported in the main text. This situation can be attributed to one of four possibilities: (1) The design of the experiment is under-powered, (2) the effects are spurious, (3) corrections for multiple comparisons are unnecessarily conservative, or (4) non-independence of fNIRS channels. I acknowledge that an under-powered design may be a contributing factor, and suggest solutions to improve upon the present design. I also argue that corrections for multiple comparisons may be unnecessarily conservative for the interpretation of channel-wise fNIRS data. The justification I provide is due to the non-independence problem of measurement that is uniquely associated with the modality of fNIRS.

First, it is certainly possible that the present experiment is under-powered. This effect is attributable to the sample size, inter-subject variability and behavioural task design. Examining individual participant's raw time-series revealed large variability of haemodynamic response as measured by fNIRS across subjects. The most likely culprit is the large variability of hair features across populations such as color, texture, length, thickness, and density (Khan et al., 2012; Orihuela-Espina, Leff, James, Darzi, & Yang, 2010). These features directly impact the critical optode-scalp contact that is necessary for the collection of high quality signal, therefore contributing to inter-subject HRF variability (Khan et al., 2012). A large HRF variability will, in turn, demand a higher sample size to ascertain a positive result, especially if the effect size is not intrinsically large. We recruited a sizable sample ( $n = 43$ ) compared to the typical neuroimaging experiment, this allowed us to detect two critical main effects, but jettisons several initially significant effects that do not survive FDR correction (Szucs & Ioannidis, 2020). In regards to the behavioural task, the choice of block design was justified as to provide a consistent format between Experiments 1, 2, and 3 but isn't optimized for statistical power with respect to GLM models. Further, the inclusion of a control (CTRL) condition was to test our secondary aim, the feasibility of fNIRS for studying categorization. If one were to optimize statistical power to dissociate Explicit and Implicit category learning systems, it would be reasonable to drop the CTRL condition completely and double the CATG blocks, since the task is identical across RB and II conditions which only differ at the pixel-scale by category structure.

Second, it is always a possibility that results do not survive multiple comparisons simply because they are Type I errors. If the underlying philosophy is to minimize Type I errors at all costs, it is indeed prudent to only report results that have survived corrections for multiple comparisons, even if Type II errors are inflated as a consequence.

Third, there is no canonical interpretation of multiple comparisons corrections that is considered universally acceptable and agnostic to discipline (Saville, 1990; Rothman, 1990). In the case of fNIRS, the problem of multiple comparisons remains very much an unsolved problem and solutions are often idiosyncratic depending on study design (Yücel et al., 2021; Singh & Dan, 2006; Uga et al., 2015). The only generally recommended action is to transparently report the approach, which I presently endeavor to do, presenting both corrected and uncorrected p-values (Yücel et al., 2021). However, the present study specified two ROI bilaterally and targeted all channels to be optimized towards those ROI. That is to say, several channels are measuring the same functional region, and the sensitivity profiles are overlapping. Therefore, applying channel-wise corrections may be conservative because I are statistically accounting for each ROI more than once, whereas our interpretations are largely conducted at the resolution of a typical Brodmann's area. Further, the Benjamini-Hochberg method assumes independence, but fNIRS channels are non-independent as a consequence of spatial correlation. There exists corrections that account for arbitrary assumptions of independence (Benjamini-Yekutieli), but are even more conservative than the standard Benjamini-Hochberg FDR correction (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001).

Fourth, the problem of non-independent fNIRS channels can be further illustrated by the following sketch: A single source is flanked by 2 detectors separated at 90 degrees forming a L-shape (Appendix Af3a). Though measurement channels are often depicted as edges joining two nodes, this is an oversimplification. Physically, the measurement channels are better approximated by spherical volumes where the origin is centered at the midpoint of the measurement channel, since this is the region with the highest density of photons that are back-scattered towards the surface detectors. Neighboring channels of standard 3cm separation, as in the present study, would measure spherical volumes that have overlapping sensitivity to the same brain region thus violating the assumption

of independence. If a cluster of neural tissue, centered in the overlapping region were to activate, both channels would report a correlated effect, though weaker than a full effect because channel sensitivity decreases as an exponential function of the measurement depth (Appendix Af3b) (Boas, Dale, & Franceschini, 2004). Therefore an application of FDR or FWER which assumes independence would inflate Type II error, an activated cluster located between neighboring channels would be deemed insignificant simply because the peak of the activation was not localized at an optimal location with respect to the probe design. In fMRI literature, this exact issue is often addressed by the application of random field theory, which corrects activation statistics based on cluster thresholds rather than for each individual voxel (Friston, Holmes, Worsley, & others, 1994). Some work has been done to translate these methods into fNIRS by interpolating the data onto inhomogenous gaussian random fields and computing the excursion probability (p-values) using Sun's tube formula (Ye, Tak, Jang, Jung, & Jang, 2009). However, in the absence of a corresponding anatomical MRI or precise 3d coordinates, interpolation methods cannot improve the resolution beyond the standard approach. Using cluster-based thresholds by interpolation may resolve the ambiguity surrounding dependency of measurement channels, but effectively trades-off one ambiguity for another (Ye et al., 2009). Interpolating without precise 3d-coordinates or an anatomical scan introduces false-spatial resolution about the location and size of activation. In the present study I specify ROI at the resolution of individual Brodmann's areas, thereby preventing false confidence in the spatial resolution. Though I do provide estimated MNI coordinates of the channel, they do not necessarily represent the precise location of the activated cluster. We therefore argue that there may be additional effects (interaction effects, activations) that are informative to the question of dissociating multiple category learning systems, but are presently sequestered by a conservative approach to channel-wise corrections.

## 5.4 Feasibility of fNIRS

The majority of neuroimaging research in category learning is done by the modality of fMRI (Nomura et al., 2007; Carpenter et al., 2016; Soto et al., 2013). Undoubtedly, fMRI offers the highest spatial resolution for non-invasive study of human participants, but also contains some physical and practical constraints due to the inherent properties of the modality. In comparison, fNIRS offers many practical advantages and a handful of technical advantages, at the cost of spatial resolution (Quaresima & Ferrari, 2019; Pinti et al., 2020). Examples include, greater participant comfort, portability, ecological validity, accessibility with special populations such as patients or children, reduced expense, and simultaneous recordings from two or more participants known as hyperscanning (Czeszumski et al., 2020).

To date, only one published study uses fNIRS to study category learning (Wu et al., 2020). In the present study I have demonstrated the feasibility of using fNIRS to study category learning. I was able to ascertain neural activity of a general categorization network with simple contrasts on a block design, collapsed across all runs (CATG > REST). Further, I was capable of detecting subtler effects of categorization strategy and performance effects based on a mixed-effects modeling approach. I therefore conclude, that fNIRS is a feasible modality for studying categorization effects, but with several important caveats.

The fNIRS technology as it currently stands: (1) is incapable of studying subcortical regions (Quaresima & Ferrari, 2019). Much of the discussion around category learning happens in deep regions such as basal ganglia, and thalamus (Ashby et al., 1998). Therefore, the present study cannot



make even the weakest claims on the category learning dynamics of these deeper subcortical regions. (2) likely does not provide high enough spatial resolution to study the effects of representation. The estimated resolution (which depends on idiosyncratic design of the probe), is approximately 2 cm (Pinti et al., 2020). Therefore, fNIRS as a modality is unlikely to contribute to the discussion of category representations, notwithstanding exceedingly creative applications of the technology. (3) Individual variability of categorization is further compounded by large differences in fNIRS haemodynamic response. Recent work has demonstrated humans have certain default preferences and strategies when category learning (Shen & Palmeri, 2016). If the analytic approach treats individual variance as a nuisance factor (eg. between-subjects design), this problem may be exacerbated by the high individual variance of fNIRS signal and haemodynamic responses. Idiosyncratic factors such as hair length, color, texture, ethnicity (skull shape), in addition to unknown physiological confounds such as heart rate or blood pressure in the skin, skull, CSF affect the quality of fNIRS signal and consequently the measured haemodynamic response. Researchers will need to be careful in accounting for the high inter-subject variability especially when studying categorization (eg. within-subjects, high N, many trials etc).

## 5.5 Conclusion

I sought to test the feasibility of using fNIRS to find a double dissociation of the neural activity underlying explicit and implicit categorization as specified by COVIS. I observed broad activation across bilateral DLPFC and PMd/SMA regions in response to general categorization task. I localized a simple single-dissociation of explicit learning eliciting higher neural response in R DLPFC. I also detected a linear relationship between overall task performance and the magnitude of neural response in R DLPFC across all subjects. I therefore conclude that explicit and implicit category learning systems can be dissociated by cortical activity and that fNIRS is indeed a feasible modality for studying categorization.

# Bibliography

- Aasted, C. M., Yücel, M. A., Cooper, R. J., Dubb, J., Tsuzuki, D., Becerra, L., ... Boas, D. A. (2015, April). Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *Neurophotonics*, 2(2), 020801.
- Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., & Carter, C. S. (2000, November). Complementary category learning systems identified using event-related functional MRI. *J. Cogn. Neurosci.*, 12(6), 977–987.
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.*, 9, 357–381.
- Aristotle. (1938). *Aristotle: The categories. on interpretation. prior analytics* (H. P. Crooke, Ed.). Harvard University Press.
- Ashby, F. G. (2014). *Multidimensional models of perception and cognition*. Psychology Press.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998, July). A neuropsychological theory of multiple systems in category learning. *Psychol. Rev.*, 105(3), 442–481.
- Ashby, F. G., & Ell, S. W. (2001, May). The neurobiology of human category learning. *Trends Cogn. Sci.*, 5(5), 204–210.
- Ashby, F. G., & Maddox, T. W. (2011). *Human category learning 2.0* (Vol. 1224) (No. 1).
- Barker, J. W., Aarabi, A., & Huppert, T. J. (2013, July). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomed. Opt. Express*, 4(8), 1366–1379.
- Benjamini, Y., & Hochberg, Y. (1995, January). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, 57(1), 289–300.
- Benjamini, Y., & Yekutieli, D. (2001, August). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.*, 29(4), 1165–1188.
- Blanco, N. J., Saucedo, C. L., & Gonzalez-Lima, F. (2017, March). Transcranial infrared laser stimulation improves rule-based, but not information-integration, category learning in humans. *Neurobiol. Learn. Mem.*, 139, 69–75.
- Boas, D. A., Dale, A. M., & Franceschini, M. A. (2004). Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *Neuroimage*, 23 Suppl 1, S275–88.
- Bopp, K. L., & Verhaeghen, P. (2005, September). Aging and verbal memory span: a meta-analysis. *J. Gerontol. B Psychol. Sci. Soc. Sci.*, 60(5), P223–33.
- Bozoki, A., Grossman, M., & Smith, E. E. (2006). Can patients with alzheimer’s disease learn a category implicitly? *Neuropsychologia*, 44(5), 816–827.
- Brigadoi, S., & Cooper, R. J. (2015, May). How short is short? optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy. *NPh*, 2(2), 025005.

- Brown, R. G., & Marsden, C. D. (1988). An investigation of the phenomenon of “set” in parkinson’s disease. *Mov. Disord.*, 3(2), 152–161.
- Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. (2016, October). A comparison of the neural correlates that underlie rule-based and information-integration category learning. *Hum. Brain Mapp.*, 37(10), 3557–3574.
- Combrisson, E., Vallat, R., O’Reilly, C., Jas, M., Pascarella, A., Saive, A.-L., ... Jerbi, K. (2019, March). Visbrain: A Multi-Purpose GPU-Accelerated Open-Source suite for multimodal brain data visualization. *Front. Neuroinform.*, 13, 14.
- Cools, A. R., van den Bercken, J. H., Horstink, M. W., van Spaendonck, K. P., & Berger, H. J. (1984, May). Cognitive and motor shifting aptitude disorder in parkinson’s disease. *J. Neurol. Neurosurg. Psychiatry*, 47(5), 443–453.
- Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., ... König, P. (2020, February). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Front. Hum. Neurosci.*, 14, 39.
- Davies, M. (2010, October). The corpus of contemporary american english as the first reliable monitor corpus of english. *Lit Linguist Computing*, 25(4), 447–464.
- Di Rosa, E., Brigadoi, S., Cutini, S., Tarantino, V., Dell’Acqua, R., Mapelli, D., ... Vallesi, A. (2019, November). Reward motivation and neurostimulation interact to improve working memory performance in healthy older adults: A simultaneous tDCS-fNIRS study. *Neuroimage*, 202, 116062.
- Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015, July). Identifying strategy use in category learning tasks: a case for more diagnostic data and models. *J. Exp. Psychol. Learn. Mem. Cogn.*, 41(4), 933–948.
- Dum, R. P., & Strick, P. L. (2005, February). Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. *J. Neurosci.*, 25(6), 1375–1386.
- Ferrari, M., & Quaresima, V. (2012, November). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage*, 63(2), 921–935.
- Friston, K. J., Holmes, A. P., Worsley, K. J., & others. (1994). Statistical parametric maps in functional imaging: a general linear approach. *brain mapping*.
- Garavan, H., Kelley, D., Rosen, A., Rao, S. M., & Stein, E. A. (2000). Practice-Related functional activation changes in a working memory task. *Microsc. Res. Tech.*, 51(1), 54–63.
- Gershberg, F. B. (1997, January). Implicit and explicit conceptual memory following frontal lobe damage. *J. Cogn. Neurosci.*, 9(1), 105–116.
- Grafton, S. T., Mazziotta, J. C., Presty, S., Friston, K. J., Frackowiak, R. S., & Phelps, M. E. (1992, July). Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *J. Neurosci.*, 12(7), 2542–2548.
- Gureckis, T. M., James, T. W., & Nosofsky, R. M. (2011, July). Re-evaluating dissociations between implicit and explicit category learning: an event-related fMRI study. *J. Cogn. Neurosci.*, 23(7), 1697–1709.
- Halsband, U., & Freund, H. J. (1990, February). Premotor cortex and conditional motor learning in man. *Brain*, 113 ( Pt 1), 207–222.
- Halsband, U., & Lange, R. K. (2006, June). Motor learning in man: a review of functional and clinical studies. *J. Physiol. Paris*, 99(4-6), 414–424.

- Helie, S., Roeder, J. L., & Ashby, F. G. (2010, October). Evidence for cortical automaticity in rule-based categorization. *J. Neurosci.*, *30*(42), 14225–14234.
- Hélie, S., Turner, B. O., Crossley, M. J., Ell, S. W., & Ashby, F. G. (2017, June). Trial-by-trial identification of categorization strategy using iterative decision-bound modeling. *Behav. Res. Methods*, *49*(3), 1146–1162.
- Hikosaka, O., Rand, M. K., Miyachi, S., & Miyashita, K. (1995, October). Learning of sequential movements in the monkey: process of learning and retention of memory. *J. Neurophysiol.*, *74*(4), 1652–1661.
- Holmes, E., Barrett, D. W., Saucedo, C. L., O'Connor, P., Liu, H., & Gonzalez-Lima, F. (2019, October). Cognitive enhancement by transcranial photobiomodulation is associated with cerebrovascular oxygenation of the prefrontal cortex. *Front. Neurosci.*, *13*, 1129.
- Hoshi, Y., & Tamura, M. (1993, February). Detection of dynamic changes in cerebral oxygenation coupled to neuronal function during mental work in man. *Neurosci. Lett.*, *150*(1), 5–8.
- Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009, April). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.*, *48*(10), D280–98.
- Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006, January). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage*, *29*(2), 368–382.
- Hyafil, A., Summerfield, C., & Koehlin, E. (2009, April). Two mechanisms for task switching in the prefrontal cortex. *J. Neurosci.*, *29*(16), 5135–5142.
- Iadecola, C. (2017, September). The neurovascular unit coming of age: A journey through neurovascular coupling in health and disease. *Neuron*, *96*(1), 17–42.
- Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989, June). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behav. Neurosci.*, *103*(3), 548–560.
- Kane, M. J., & Engle, R. W. (2002, December). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.*, *9*(4), 637–671.
- Kelly, A. M. C., & Garavan, H. (2005, August). Human functional neuroimaging of brain changes associated with practice. *Cereb. Cortex*, *15*(8), 1089–1102.
- Khan, B., Wildey, C., Francis, R., Tian, F., Delgado, M. R., Liu, H., ... Alexandrakis, G. (2012, May). Improving optical contact for functional near-infrared brain spectroscopy and imaging with brush optodes. *Biomed. Opt. Express*, *3*(5), 878–898.
- Klein, H. A. (2005). Cultural differences in cognition: Barriers in multinational collaborations. *How professionals make decisions*, 243–253.
- Knowlton, B. J., Paulsen, J., Swenson, M., & Butters, N. (1996). Dissociations within nondeclarative memory in HD. *Neuropsychology*, *10*(4), 538–548.
- Laozi. (1868). *The speculations on metaphysics, polity, and morality, of the old philosopher, lau-tsze* (J. Chalmers, Ed.). Trübner.
- Le Pelley, M. E., Newell, B. R., & Nosofsky, R. M. (2019, September). Deferred feedback does not dissociate implicit and explicit Category-Learning systems: Commentary on smith et al. (2014). *Psychol. Sci.*, *30*(9), 1403–1409.
- Lewohl, J. M., Wang, L., Miles, M. F., Zhang, L., Dodd, P. R., & Harris, R. A. (2000, December). Gene expression in human alcoholism: microarray analysis of frontal cortex. *Alcohol. Clin.*

- Exp. Res.*, 24(12), 1873–1882.
- Maddox, W. T., & Filoteo, J. V. (2001, September). Striatal contributions to category learning: quantitative modeling of simple linear and complex nonlinear rule learning in patients with parkinson's disease. *J. Int. Neuropsychol. Soc.*, 7(6), 710–727.
- Maddox, W. T., Todd Maddox, W., & Gregory Ashby, F. (2004). *Dissociating explicit and procedural-learning based systems of perceptual category learning* (Vol. 66) (No. 3).
- Marr, D. (1982). The philosophy and the approach. *Visual perception: Essential readings*, 104–123.
- Matelli, M., & Luppino, G. (1996, August). Thalamic input to mesial and superior area 6 in the macaque monkey. *J. Comp. Neurol.*, 372(1), 59–87.
- Meidenbauer, K. L., Choe, K. W., Cardenas-Iniguez, C., Huppert, T. J., & Berman, M. G. (2021, April). Load-dependent relationships between frontal fNIRS activity and performance: A data-driven PLS approach. *Neuroimage*, 230, 117795.
- Mencarelli, L., Neri, F., Momi, D., Menardi, A., Rossi, S., Rossi, A., & Santarnecchi, E. (2019, September). Stimuli, presentation modality, and load-specific brain activity patterns during n-back task. *Hum. Brain Mapp.*, 40(13), 3810–3831.
- Miles, S. J., & Minda, J. P. (2009). Learning new categories: Adults tend to use rules while children sometimes rely on family resemblance. *Proceedings of the Cognitive Science Society*, 31(31).
- Milton, F., & Pothos, E. M. (2011, October). Category structure and the two learning systems of COVIS. *Eur. J. Neurosci.*, 34(8), 1326–1336.
- Minda, J. P., Desroches, A. S., & Church, B. A. (2008, November). Learning rule-described and non-rule-described categories: a comparison of children and adults. *J. Exp. Psychol. Learn. Mem. Cogn.*, 34(6), 1518–1533.
- Minda, J. P., & Miles, S. J. (2010, January). Chapter 3 - the influence of verbal and nonverbal processing on category learning. In *Psychology of learning and motivation* (Vol. 52, pp. 117–162). Academic Press.
- Morais, G. A. Z., Balardin, J. B., & Sato, J. R. (2018, February). fNIRS optodes' location decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Sci. Rep.*, 8(1), 1–11.
- Morgan, K. K., Zeithamova, D., Luu, P., & Tucker, D. (2020). *Spatiotemporal dynamics of multiple memory systems during category learning* (Vol. 10) (No. 4).
- Murphy, G. L., Bosch, D. A., & Kim, S. (2017, November). Do americans have a preference for rule-based classification? *Cogn. Sci.*, 41(8), 2026–2052.
- Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., ... Reber, P. J. (2007, January). Neural correlates of rule-based and information-integration visual category learning. *Cereb. Cortex*, 17(1), 37–43.
- Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002, September). Cultural preferences for formal versus intuitive reasoning. *Cogn. Sci.*, 26(5), 653–684.
- Nosofsky, R. M. (1986, March). Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.*, 115(1), 39–61.
- Ogawa, Kotani, K., & Jimbo, Y. (2014, July). Relationship between working memory performance and neural activation measured using near-infrared spectroscopy. *Brain Behav.*, 4(4), 544–551.
- Ogawa, Lee, T. M., Kay, A. R., & Tank, D. W. (1990, December). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U. S. A.*, 87(24), 9868–9872.

- Orihuela-Espina, F., Leff, D. R., James, D. R. C., Darzi, A. W., & Yang, G. Z. (2010, July). Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. *Phys. Med. Biol.*, *55*(13), 3701–3724.
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005, May). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.*, *25*(1), 46–59.
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020, March). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann. N. Y. Acad. Sci.*, *1464*(1), 5–29.
- Poldrack, R. A. (2000, July). Imaging brain plasticity: conceptual and methodological issues—a theoretical review. *Neuroimage*, *12*(1), 1–13.
- Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional MRI data analysis*. Cambridge University Press.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal Of Experimental Psychology*, *77*(3), 353–363.
- Quaresima, V., & Ferrari, M. (2019, January). Functional Near-Infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in Natural/Social situations: A concise review. *Organizational Research Methods*, *22*(1), 46–68.
- Rabi, R., & Minda, J. P. (2016, March). Category learning in older adulthood: A study of the shepard, hovland, and jenkins (1961) tasks. *Psychol. Aging*, *31*(2), 185–197.
- Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003, May). Dissociating explicit and implicit category knowledge with fMRI. *J. Cogn. Neurosci.*, *15*(4), 574–583.
- Robinson, A. L., Heaton, R. K., Lehman, R. A., & Stilson, D. W. (1980, October). The utility of the wisconsin card sorting test in detecting and localizing frontal lobe lesions. *J. Consult. Clin. Psychol.*, *48*(5), 605–614.
- Rothman, K. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*(1), 43–46.
- Sato, H., Yahata, N., Funane, T., Takizawa, R., Katura, T., Atsumori, H., ... Kasai, K. (2013, December). A NIRS–fMRI investigation of prefrontal cortex activity during a working memory task. *Neuroimage*, *83*, 158–173.
- Saville, D. J. (1990). Multiple comparison procedures: The practical solution. *Am. Stat.*, *44*(2), 174–180.
- Seitz, R. J., Canavan, A. G., Yágüez, L., Herzog, H., Tellmann, L., Knorr, U., ... Hönberg, V. (1994, December). Successive roles of the cerebellum and premotor cortices in trajectory learning. *Neuroreport*, *5*(18), 2541–2544.
- Shen, J., & Palmeri, T. J. (2016, November). Modelling individual difference in visual categorization. *Vis. cogn.*, *24*(3), 260–283.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.
- Singh, A. K., & Dan, I. (2006, November). Exploring the false discovery rate in multichannel NIRS. *Neuroimage*, *33*(2), 542–549.
- Smith, Mancini, M. C., & Nie, S. (2009, November). Bioimaging: second window for in vivo imaging. *Nat. Nanotechnol.*, *4*(11), 710–711.
- Smith, & Minda, J. P. (2002, July). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *J. Exp. Psychol. Learn. Mem. Cogn.*, *28*(4), 800–811.

- Smith, Minda, J. P., & Washburn, D. A. (2004, September). Category learning in rhesus monkeys: a study of the shepard, hovland, and jenkins (1961) tasks. *J. Exp. Psychol. Gen.*, *133*(3), 398–414.
- 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* (Vol. 71).
- Stark, C. E., & Squire, L. R. (2001, October). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc. Natl. Acad. Sci. U. S. A.*, *98*(22), 12760–12766.
- Szucs, D., & Ioannidis, J. P. (2020, November). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990-2012) and of latest practices (2017-2018) in high-impact journals. *Neuroimage*, *221*, 117164.
- Toni, I., Ramnani, N., Josephs, O., Ashburner, J., & Passingham, R. E. (2001, November). Learning arbitrary visuomotor associations: temporal dynamic of brain activity. *Neuroimage*, *14*(5), 1048–1057.
- Tucker, S., Dubb, J., Kura, S., von Lühmann, A., Franke, R., Horschig, J. M., ... Luke, R. (2023, January). Introduction to the shared near infrared spectroscopy format. *Neurophotonics*, *10*(1), 013507.
- Uga, M., Dan, I., Dan, H., Kyutoku, Y., Taguchi, Y.-H., & Watanabe, E. (2015, January). Exploring effective multiplicity in multichannel functional near-infrared spectroscopy using eigenvalues of correlation matrices. *Neurophotonics*, *2*(1), 015002.
- Virag, M., Janacsek, K., Horvath, A., Bujdoso, Z., Fabo, D., & Nemeth, D. (2015, July). Competition between frontal lobe functions and implicit sequence learning: evidence from the long-term effects of alcohol. *Exp. Brain Res.*, *233*(7), 2081–2089.
- Wang, Z., Zhu, R., Rehman, A. U., & You, X. (2020, October). Dorsolateral prefrontal cortex and Task-Switching performance: Effects of anodal transcranial direct current stimulation. *Neuroscience*, *446*, 94–101.
- Webler, R. D., Fox, J., McTeague, L. M., Burton, P. C., Dowdle, L., Short, E. B., ... Nahas, Z. (2022, May). DLPFC stimulation alters working memory related activations and performance: An interleaved TMS-fMRI study. *Brain Stimul.*, *15*(3), 823–832.
- Willingham, D. B. (1998, July). A neuropsychological theory of motor skill learning. *Psychol. Rev.*, *105*(3), 558–584.
- Wu, J., Fu, Q., & Rose, M. (2020, February). Stimulus modality influences the acquisition and use of the rule-based strategy and the similarity-based strategy in category learning. *Neurobiol. Learn. Mem.*, *168*, 107152.
- Ye, J., Tak, S., Jang, K., Jung, J., & Jang, J. (2009). *NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy* (Vol. 44) (No. 2).
- Yücel, M. A., Lühmann, A. V., Scholkmann, F., Gervain, J., Dan, I., Ayaz, H., ... Wolf, M. (2021, January). Best practices for fNIRS publications. *Neurophotonics*, *8*(1), 012101.

# Appendix

Table A1: Basic Participant Demographics. Age displays means, all other variables display percentages relative to sample size for each respective experiment.

Variable	Exp.I	Exp.II	Exp.III
Age	19.7	19.3	20.4
Gender			
Female	62	75	56
Male	37	24	44
Other	1	1	0
Ethnicity			
White	46	50	44
Black	5	1	2
Asian	36	43	42
Hispanic	3	1	0
Other	10	5	10
Education			
Secondary	2	1	0
Diploma	8	9	2
Bachelors	89	86	79
Masters	1	2	4
Doctorate	0	1	2



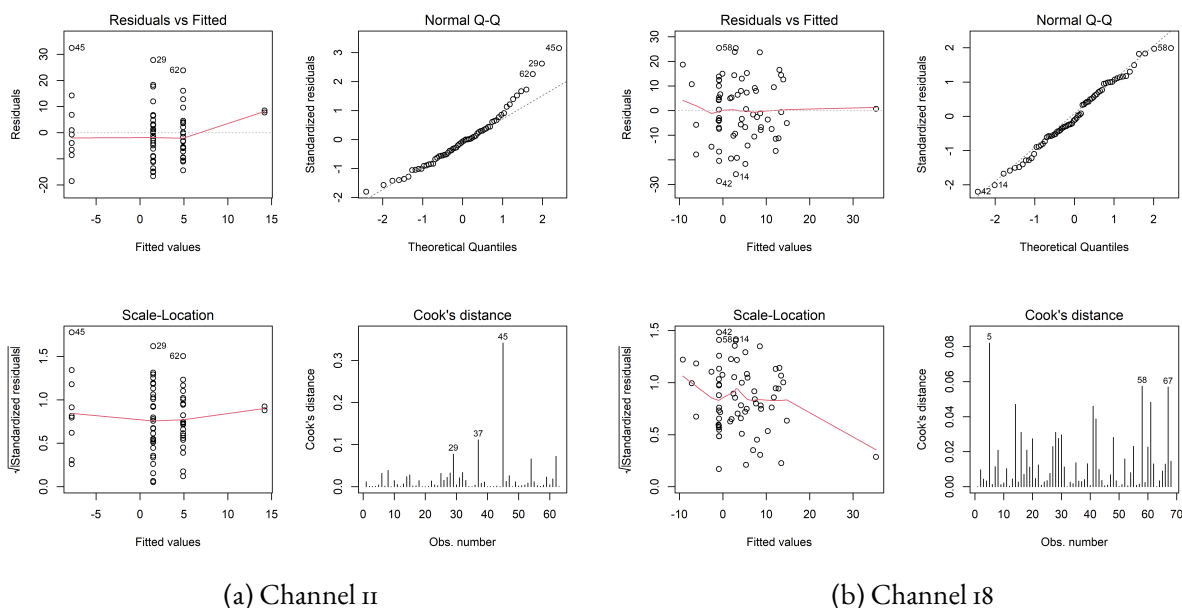


Figure Af1: Model diagnostic plots for the significant channels exhibit group effects. Note that model residuals appear to violate assumption of normality. However with the sample size is  $>30$  ( $n = 86$ ), the violation has relatively minimal influence on whether OLS is the best linear unbiased estimator.

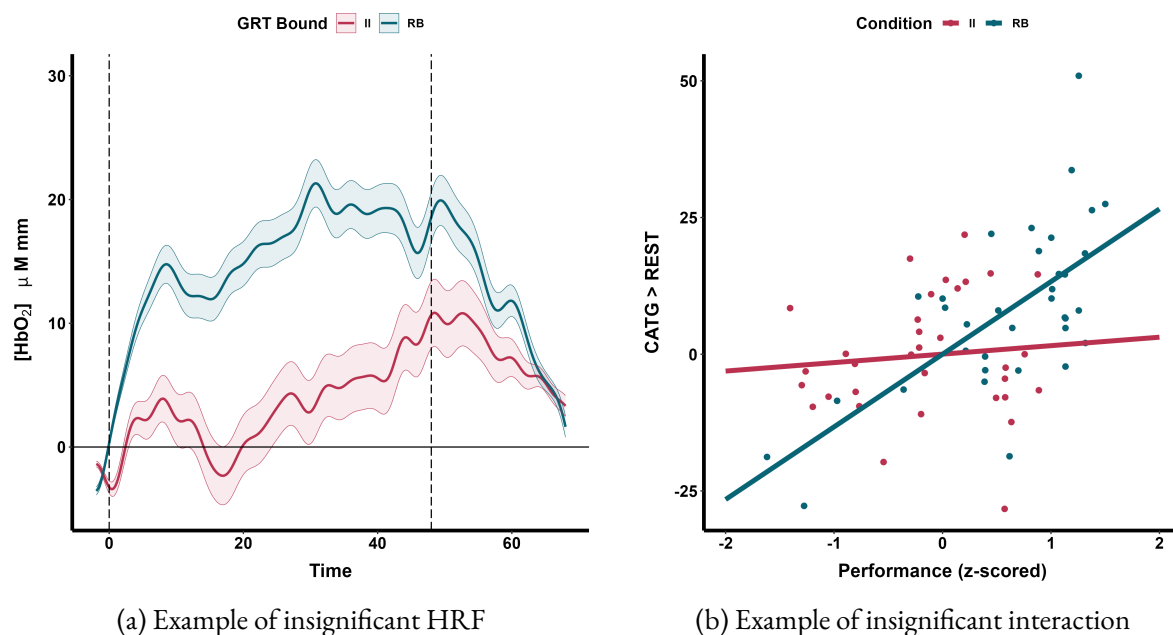
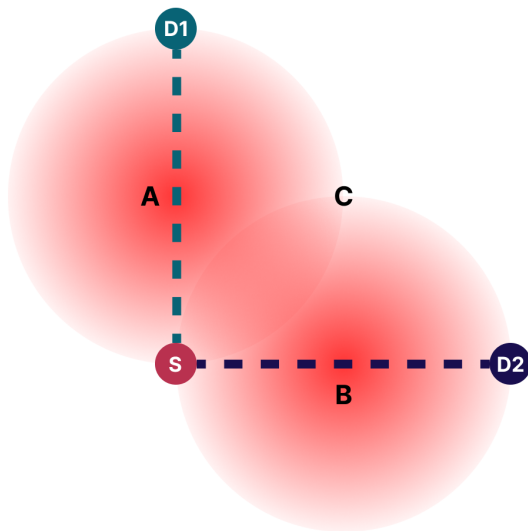
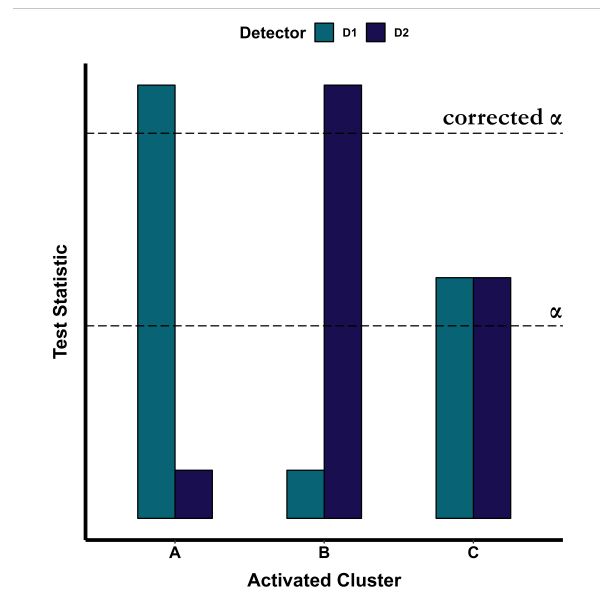


Figure Af2: Examples of interesting qualitative effects that are initially significant, but sequestered by FDR corrections for multiple comparisons. (a) Estimated HRF on channel 14, L DLPFC, BA9, shows a marked difference between RB and II condition,  $t = -2.54$ ,  $p = .04$ ,  $p_{FDR} = .18$ , but fails to survive FDR correction. (b) Performance  $\times$  condition exhibits a visible interaction effect,  $t = 2.56$ ,  $p = .013$ ,  $p_{FDR} = .17$ , but fails to survive FDR correction.



(a) Diagram of neighboring channels



(b) Relative signal for respective activation clusters

Figure Af3: Schematic of fNIRS channel non-independence. (a) One source and two detectors,  $D_1$ ,  $D_2$  creates two channels of measurements where the sensitivity profile overlaps. Instances of cluster activation may happen at a relatively independent measurement location; A, B; or at a non-independent locations; C. (b) Example of non-independent measurements C failing to pass significance thresholds post-correction for multiple comparisons. Dashed line of  $\alpha$  represents the significance threshold of the test-statistic pre and post corrections. Note that cluster C would otherwise pass an uncorrected threshold.

# Curriculum Vitae

**Name:** Tim Qiu

**Post-Secondary Education and Degrees:** University of Western Ontario  
London, ON  
2021 - 2023, MSc

University of Western Ontario  
London, ON  
2017 - 2021, BSc

**Honours and Awards:** Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII-GSST)  
2022

Global Undergraduate Awards Highly Commended in Psychology  
2021

NSERC Undergraduate Student Research Award  
2020

NSERC Undergraduate Student Research Award  
2018

## Publications:

Qiu, T. T., & Minda, J. P. (2023). Category Learning DLPFC Single-Dissociation by fNIRS. (In prep)

Qiu, T. T., & Minda, J. P. (2023, June 24). Novel Concentric-Circle Technique Interrogates Implicit Category Learning. <https://doi.org/10.31234/osf.io/ps5aq>

Qiu, T. T., & Minda, J. P. (2022). Psychedelic Experiences and Mindfulness are Associated with Improved Wellbeing. *Journal of psychoactive drugs*, 1-11. <https://doi.org/10.1080/02791072.2022.2060773>

Qiu, T. T., & Minda, J. P. (2021). Recreational Psychedelic Users Frequently Encounter Mystical Experiences, Trip Content and Implications for Wellbeing. *PsyArXiv*. <https://doi.org/10.31234/osf.io/xrbzs>

Qiu, T. T., Nielsen, E., Guimaraes, E., & Minda, J. P. (2022). Trait, but not State Mindfulness Improves Resistance to Cognitive Biases. PsyArXiv. <https://doi.org/10.31234/osf.io/n33pt>