2019

Subjective Measures of Implicit Categorization Learning

Audrey Zlatkin
University of Central Florida

SUBJECTIVE MEASURES OF IMPLICIT CATEGORIZATION LEARNING

by

AUDREY HILL ZLATKIN
B.S. University of Central Florida, 2012
M.S University of Central Florida, 2015

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2018

Major Professor: Corey Bohil

ABSTRACT

The neuropsychological theory known as COVIS (COmpetition between Verbal and Implicit Systems) postulates that distinct brain systems compete during category learning. The explicit system involves conscious hypothesis testing about verbalizable rules, while the implicit system relies on procedural learning of rules that are difficult to verbalize. Specifically from a behavioral approach, COVIS has been supported through demonstrating empirical dissociations between explicit and implicit learning tasks. The current studies were designed to gain deeper understanding of implicit category learning through the implementation of a subjective measure of awareness, Meta d', which until now has not been validated within a COVIS framework. Meta d' is a measure of metacognitive accuracy. This is the ability to assess the accuracy of one's own performance. These three experiments evaluated the use of Meta d' as a valid predictor of task performance within a two structure perceptual categorization task. Experiment 1 focuses on using Meta d' to parse out dissociations between awareness and performance through the phenomenon of Blind Sight and Blind Insight. Experiment 2 and 3 utilize a motor response mapping disruption to observe predicted decrements to the implicit learning system. Experiment 3 utilizes functional Near Infrared Spectroscopy (fNIRS) to measure hemodynamic changes in the Prefrontal Cortex as a function of category structure. Across the 3 experiments, Meta d' in conjunction with decision bound model fits were used to make accurate predictions about the differences in performance throughout implicit and explicit categorization tasks. These collective results indicate that metacognitive accuracy, an implicit structure, was highly sensitive to a whether a person is using the correct rule strategies through the task.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION TO CATEGORIZATION

Categorization is a pervasive aspect of human perception and cognition and, as such, it has long been a topic of debate among philosophers and scientists. Across the generations, several theories of categorization have developed to parse out the cognitive, perceptual, and neurobiological processes involved. To understand the most prominent theories of today, it is necessary to understand the historical evolution of the field.

The oldest theory of categorization, Classical Theory, was founded on the idea that every category is made up of common properties. These properties are both necessary to the category and sufficient for its representation (Smith & Medin, 1981). In time, Classical Theory began receiving criticism for oversimplification. Not all exemplars in a category contain the same properties. Additionally, the theory did not account for typicality or goodness. Typicality and goodness relate to the idea that certain category members are more typical, and therefore better examples, despite possessing the same ‘necessary and sufficient features’ as other examples (Smith & Medin, 1981). For instance, a golden retriever is a more typical example of a dog than a chihuahua. Research found that in category learning tasks, the most typical exemplars result in more accurate and faster responses than less typical examples (Rips, Shoben, & Smith, 1973). Based on this research, a new theory on categorization, Prototype Theory, began taking shape.

Prototype Theory gained traction with its ability to account for typicality effects (Rosch, 1973). Prototype Theory proposes that a category’s mental representation is distributed around the most typical category member. This theory accounted for much of past criticism of classical theory in that it did not rely on a set of specific properties. Instead, the theory states that category membership is determined based on similarity to the prototype. The more similar to the prototype, the more typical the category example. However, as with any scientific theory, there are criticisms of prototype theory. The primary concern with prototype theory is the reliance on a single prototype representation (Smith & Medin, 1981). Single prototype representation became increasingly problematic as research surfaced showing it was possible for multiple category exemplars to produce similar task performance (Medin & Schaffer, 1978). For example, in addition to the golden retriever, a labrador is also a typical exemplar for dogs. It is likely that participants would classify these two prototypical breeds of dog more quickly than a less typical breed. Additionally, prototype theory relies on 'similarity' without providing an operational definition or set of criteria for measuring it. Allowing for these developments, Exemplar theory evolved out of Prototype Theory.

Exemplar theory is an instance-based theory proposing that each decision engages a large-scale comparison of the stimulus to every previously encountered category exemplar stored in memory (Medin & Schaffer, 1978). Like prototype theory, this relies on similarity. However, exemplar theory has also been subject to criticism. It is unlikely that each classification judgment involves a global comparison to every single experienced exemplar stored in memory.

General Recognition Theory (GRT) is - like prototype theory - an abstraction-based category learning theory. However, it does not possess the same reliance on similarity-based reasoning. It is based on a perceptual space mapping instead of relying on memory representation of exemplars. This theory, also known as decision bound theory, postulates that participants learn the decision bound dividing the categories without activating exemplar or prototype memories (Ashby & Gott, 1988). The assumption is that participants divide a perceptual stimulus space into response regions; the division is represented by the decision bound. On each trial, the participant discerns the appropriate response region based on the stimulus' relationship to the decision bound and makes the corresponding response.

COVIS

Ashby and colleagues (1998) used the foundations of General Recognition theory to create a categorization theory with a neuropsychological basis. The theory, called COVIS (Competition between Verbal and Implicit Systems), suggests that there are separate learning systems mediating explicit and implicit learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). COVIS predicts that the explicit system is often given preference, despite not always being optimal to the task, causing competition between the systems.

The explicit learning system relies on a neural network comprised of the prefrontal cortex, anterior cingulate, head of the caudate nucleus, hippocampus, as well as other medial temporal lobe structures (Ashby & Maddox, 2011). Explicit-rule learning is primarily mediated by the prefrontal cortex. COVIS posits that potential rules are held in working memory to be evaluated until there is sufficient support for a rule switch or the current rule is verified through corrective feedback. Over the course of the task, many different rules could be created and tested before participants settle on the best accuracy maximizing rule (Ashby & Maddox, 2005). The explicit learning system is often tested using rule-based categorization tasks. Rule-based tasks are explicit in nature, considered easy to verbalize, and learned through hypothesis testing. For example, as seen in Figure 1 (left), it is possible for participants to use a criterion only based on length, while completely ignoring angle, to perform accurately on the task.

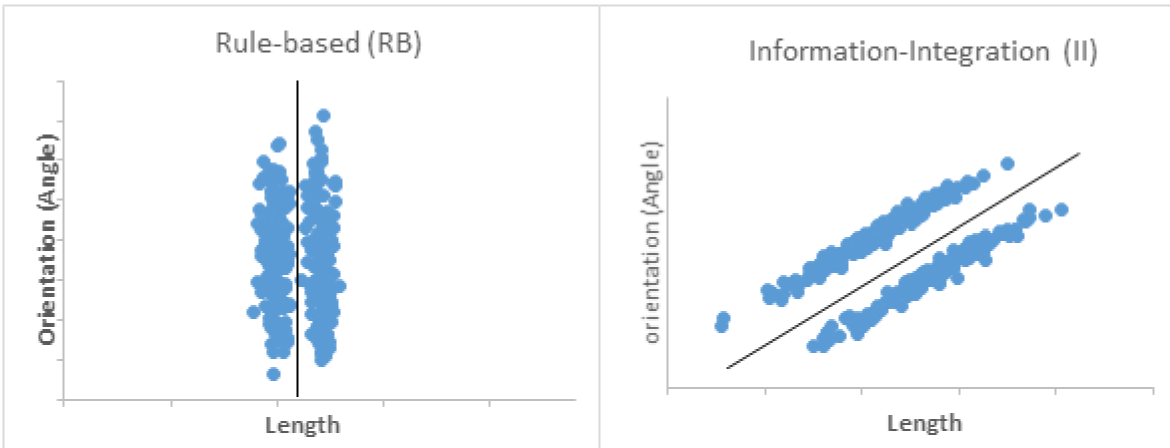


Figure 1. Stimulus values for both RB and II category structures

Scatterplots show stimulus values for both RB (left) and II category structures (right). The solid lines represent the accuracy maximizing decision bound for each structure

In contrast, the implicit learning system relies on basal ganglia structures for rule learning. The switch between potential rules is said to be specifically mediated by the head of the caudate nucleus (Ashby & Maddox, 2011). The implicit system relies on slow stimulus-response learning that is not consciously accessible to the explicit reasoning system. Implicit category-rule learning is an incremental and procedurally based process. Participants in implicit-rule learning tasks can achieve a high level of performance despite an inability to verbally state the rule. Implicit learning is often assessed through information-integration tasks. These tasks are unique in that accuracy is maximized only when the participant pulls information from two or more stimulus components, or dimensions (Figure 1, right). Traditionally, these types of rules are much more difficult, if not impossible, to verbalize. Numerous studies suggest that despite being unable to describe the rule verbally, participants can still achieve the same levels of accuracy as with the rule-based categories provided enough experience (Ashby & Maddox, 2005; Ashby, Maddox, & Bohil, 2002).

The two-system debate

The theory of COVIS, like any scientific theory, is controversial and frequently debated among the different camps within categorization research. Despite significant evidence suggesting separate system contributions to learning (Ashby et al., 1998), popular theories of categorization, namely prototype and exemplar, are based on the single system theory, and retain support among many cognitive psychologists. Within single system theories, implicit and explicit tasks are treated the same, despite the significant behavioral (Ashby, Ell, &Waldron, 2003; Maddox, Bohil, & Ing, 2004) and neurophysiological (Ashby et al., 1998) data suggesting otherwise (Ashby & Maddox, 1998).

Specifically from a behavioral approach, COVIS has been supported through demonstrating empirical dissociations between explicit and implicit learning tasks. Traditionally, this is done through various manipulations including feedback structure, response format, working memory load changes, secondary stressors, and many more. For example, Markman, Maddox, and Worthy (2009) demonstrated that through the introduction of a social pressure to an explicit categorization task, they were able to see significant performance deficits (i.e., choking under pressure). However, these deficits were not seen within the implicit task condition; surprisingly, the pressure actually improved performance. This is just one example of how these manipulations were used to consistently and reliably demonstrate behavioral differences between explicit and implicit learning tasks (Ashby & Maddox, 2011).

To date, COVIS is one of the only categorization theories that can make predictions based on the structures of the brain, allowing it to account for a much broader range of data (Ashy & Maddox, 2011). COVIS research is supported by numerous studies using populations with documented brain region injuries and various physical and mental health disorders such as Parkinson's disease, PTSD, ADD, and Schizophrenia. These neuroimaging studies offered converging evidence of separate systems mediating implicit and explicit category learning (Gluck & Poldrack, 2008). This allows for predictions to be made about which regions of the brain could be active during different types of learning tasks.

Despite the growing evidence in support of a separate systems theory of learning, there is a notable gap in the literature when it comes to the subjective experience. To my knowledge, there are very few studies that integrate subjective measures of awareness into a COVIS-based categorization learning task. In an effort to bridge this gap, pilot data was collected using an adaptation of a subjective measure of rule awareness, Scott and Dienes (2005) implicit attribution scale, in combination with our traditional two structure categorization framework (Wismer, Zlatkin, & Bohil, 2016). Participants completed an explicit (rule-based) or implicit (information integration) task. They were asked to group perceptual stimuli (lines of varying length and orientation) into a category A or category B, based on an unknown predetermined set of rules. Corrective feedback was used to help participants learn the rule as they completed the task. After each trial, participants were asked to attribute their decision response to one of 3 designations: Guessing, Intuition, or Rule. Participants in the rule-based, explicit condition overwhelmingly attributed their decisions to Rule. In the information-integration, implicit, task participants attributed their decisions to intuition significantly more than those in the Rule-based condition (Figure 2).

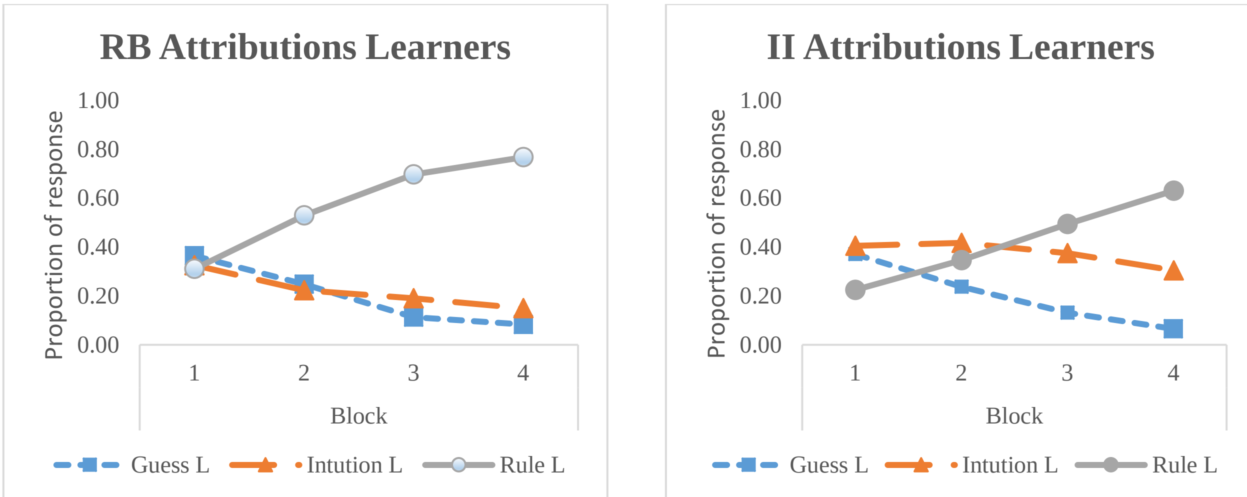


Figure 2 Wismer et al., 2016 pilot data

This Figure shows the pilot data from Wismer et al., 2016. Figure 2 (left*)* shows RB participants overwhelming attributing rule (grey circles). Figure 2 (right) shows II participants selecting intuition (orange triangles) significantly more than those in the RB condition.

These results show that not only do participants possess some level of knowledge awareness, but they believe they are employing unique strategies for each learning task. This strongly suggests that the differences between the Implicit and Explicit learning tasks are evident in the subjective data, as well as the well-established behavioral and neurophysiological results. These pilot results demonstrate the potential of subjective measures to support the separate systems theory of category learning. For these reasons, it is a notable gap in the literature, that once rectified could potentially bolster and expand on the COVIS paradigm. For these reasons, the potential usefulness of subjective measures of awareness within a categorization framework must be further explored.

CHAPTER 2: MEASURING IMPLICIT LEARNING

COVIS relies on the concept of competition between the systems. Inherent to this theory is the idea that while the explicit and implicit learning systems are separate, they interact and trade off on decisional influence on a trial by trial basis. Implicit learning theories attribute a great deal to processes that are difficult to directly measure, such as awareness. Traditionally, implicit learning has been measured through objective means such as accuracy and reaction time. However, pilot research suggests that these traditional methods leave out vital information on introspective changes that occur throughout a learning task (Wismer et al., 2016*;* Zlatkin et al., *2016).* Integrating subjective measures into the paradigm can shed valuable insight into the underlying strategies employed throughout this competition process. Despite their prevalence in the overarching field of implicit learning, subjective measures have yet to be used to assess categorization learning within the COVIS framework. I believe that integrating these proven subjective measures to traditional categorization paradigms is the next step to further understanding separate system contributions.

Subjective Measures of Implicit learning

Cognitive learning paradigms such as artificial grammar learning (AGL), serial response time (SRT), and perceptual categorization are commonly used within implicit learning research. Within the AGL and SRT frameworks, subjective measures of awareness have long been used to assess implicit learning, despite their absence in other areas of learning research. Subjective measures of awareness serve to address the underlying knowledge and strategies used throughout an implicit learning task, where it is often assumed the knowledge is unconscious and inaccessible. Essentially, they are designed to parse out the difference between ‘knowing *that* we know' and 'knowing *what* we know' (Reber, 1967). These measures are often paired with objective measures of performance such as accuracy or reaction time.

Criteria

Of course, not all measures are equal, and it is necessary to assess the quality of a subjective measure. Dienes and Colleagues (1995) created two criteria, the 'zero-correlation criterion', and the 'guessing criterion' to assess the quality of a subjective measure. Each of these criteria test for the levels of conscious and unconscious processing (Dienes, Altmann, Kwan, & Goode, 1995). The Zero-Correlation criterion is met when there is little to no correlation between the objective measure of performance and the subjective measure (self-reported awareness or task confidence), across varying degrees of difficulty. A positive correlation between these variables would indicate that there is a measurable amount of conscious awareness contributing to performance. Similarly, for the 'guessing criterion,' performance is assessed in cases where participants claim to be guessing (i.e., they believe they are answering at random). It can be assumed that if performance is at chance, there is no knowledge (conscious or otherwise) contributing to the responses. In instances where participants claim to be guessing but perform above chance, it is possible to conclude that they are relying on some form of unconscious knowledge. The most promising scale is the one that (a) shows better correlation than others between performance and awareness at varying levels of difficulty (the zero-correlation criterion), and (b) shows the least above-chance performance for trials on which participants believe they are guessing (the guessing criterion).

Measures

Some widely used subjective measures include the perceptual awareness scale (PAS), post- decision wagering (PDW), and confidence ratings (CRs). Each of these scales can be tested using the zero correlation and guessing criterion and have been validated to assess the difficult to verbalize underlying levels of task awareness.

Perceptual Awareness Scale

Ramsay and Overgaard (2004) developed the Perceptual Awareness Scale (PAS) to describe the quality of a visually presented stimulus set. The PAS originated out of a participant generated scale. Participants were shown a stimulus set made up of various shapes and colors, asked to report all features, and to report degree of clearness of the experience. They were only instructed to use a scale going from ‘no experience at all’ to ‘clear image’ and were free to structure their own scales otherwise. Nearly all participants utilized a 4-point scale with levels approximating ‘no experience', ‘brief glimpse’, ‘almost clear image’, and ‘absolutely clear image’. This formed the basis for the Perceptual Awareness Scale, which provides a deeper look at the conscious experience of the participant across the task.

This style of awareness scale has been adapted across several different paradigms of research. One adaptation, the Implicit Attribution Scale (Dienes & Scott, 2005), takes a similar approach, but with a less fluid, predetermined grading. This scale requires the participant to make trial-by-trial assessments on the bases of their judgements, using a fixed five-point scale; 1- *Guess*, 2- *intuition*, 3- *pre-existing knowledge*, 4- *Rules*, and 5- *Memory*. Attributing trial decisions to *guess* indicates that the participant had seemingly no basis or reasoning for that judgement, as if they had just flipped a coin. Attributing *intuition* indicates that the participant was somewhat confident in the decision, but they could not specifically explain why it was correct. Attributing *preexisting knowledge* indicates that the judgment was based on knowledge they possessed prior to the task, and not gained from the training phase. This attribution is not relevant within paradigms of category learning in which abstract category structures are created to purposely avoid pre-existing knowledge, and thus has been removed when adapted to these types of paradigms (Wismer et al., 2016). The *rules* attribution indicates a participant felt they based their answer on a distinct rule or set of rules acquired throughout the duration of the task, and they could verbalize these rules if asked. The *memory* attribution indicates the participant felt they based their response on specific stimulus information stored in memory. *Intuition* judgment responses were indicative of structural knowledge being more unconscious, or implicit. The *rules* and *memory* responses were indicative of explicit, and likely conscious, structural knowledge being used.

Dienes and Scott (2005) validated the Implicit Attribution Scale as a means of parsing out a participant's unconscious and conscious structural knowledge throughout an artificial grammar learning task. They took several steps to assess the validity of this scale. First, they integrated the well-used and validated measure of confidence ratings to compare. Secondly, they introduced manipulation checks, though the use of secondary tasks that interrupted conscious knowledge. This allowed them to make predictions on what types of knowledge the IAT should pick up on. These efforts provided converging evidence that the IAS was measuring what they intended.

Post-decision wagering

Post-decision wagering (PDW) is a relatively new measure of task awareness (Persaud, Mcleod, & Cowey, 2007). A task using PDW requires that after task completion, either at a trial level, block, or full duration, participants place 'wagers' on whether they performed the task correctly. This wagering relies on the confidence of the participants without explicitly asking them to report their awareness through introspection. According to Persaud and colleagues (2007), they simply perform a task that necessitates awareness of their own capabilities to be completed successfully. The researchers used a dichotomous form of wagering, with participants being able to place either a low or high wager on one of two possibilities. The degree to which a participant uses advantageous wagering to maximize gains (betting high after a correct decision, or low after an incorrect decision) is assumed to be indicative of conscious experience. For example, in a poker game, if someone knows that they have a good hand, they are more likely to place a high bet. Based on the PDW theory, the possibility to win (real or imaginary) currency provides participants with an incentive to reveal any conscious knowledge they may possess. In the same vein, failure to wager advantageously should reflect an absence of awareness. The additional motivation of reward could potentially make this a more exhaustive measure (Sandberg, Timmermans, Overgaard, & Cleeremans, 2010).

PDW scales have been consistently shown to be a valid subjective measures of knowledge awareness. Dienes and Seth (2009) put the PDW method to the test by comparing results from both PDW and verbal confidence rating on the same Artificial Grammar learning task. They found that PDW results correlate with those of the verbal confidence ratings, with each scale possessing its own strengths. Additionally, PDWs have become a popular subjective measure of implicit learning, partly due to the entertainment factor of the 'gambling' component. Several studies have utilized this method to assess underlying knowledge awareness (Haider, Eichler, and Lang, 2011; Pasquali, Timmermans, and Cleeremans, 2010).

Confidence Ratings

Confidence ratings (CRs) can be used to assess perception itself, where participants directly rank confidence in perceiving something (Cheesman & Merikle, 1986), or to assess knowledge of task performance, where they rank their confidence in whether they made the correct decision. In the former style, CRs work similarly to a PAS. In the second case, CRs used as a metacognitive judgment in which participants are requires to evaluate their own performance in a discrimination task (Dienes et al., 1995). CRs can be assessed on a variety of different scales, however, they commonly begin with ‘guessing’ or ‘no confidence’ in the as the lowest level. These scales can be dichotomous in nature, such as ‘guess/know’, as well as gradual scales. CRs are advantageous in that participants are directly asked to reflect on their conscious experience.

Ziori and Dienes (2006) sought to validate confidence ratings as a reliable subjective measure of learning. Through pairing CRs with free verbal reports and object performance ratings they were able to ascertain that CRs followed the same qualitative pattern indicating construct validity. Additionally, they found that confidence ratings were more sensitive to awareness shifts than verbal reports.

Meta d’

Meta d’ is a measure derived from confidence ratings and used to assess a participant’s metacognitive accuracy. Metacognitive accuracy is the ability to assess the accuracy of one’s own performance (Scott, Dienes, Barret, Bor, & Seth, 2014). Quantitatively this is assessed through the relationship between judgement accuracy and confidence accuracy; Meta d’ is a measure of this relationship. Meta d’ is increasingly seen within Artificial Grammar Learning (AGL) research. AGL research focuses on implicit learning of grammar strings. Participants must categorize grammar strings as either grammatical or ungrammatical based on a set of predetermined rules. Meta d’ is a measure of how accurately a participant’s confidence ratings predict task accuracy *(Maniscalco and Lau, 2012)*. Quantitatively this is assessed through the relationship between judgement accuracy and confidence accuracy. Judgement accuracy is the type 1 task while confidence accuracy, or metacognitive accuracy, represents the type 2 task. Meta d’ measures type 2 sensitivity, or metacognitive sensitivity, calculated through a ratio

(*Meta d'/d'*). If *Meta d' = d'* the participants would be considered at ideal performance (Maniscalco & Lau, 2012). Meta d' is reminiscent of signal detection theory in that it uses confidence 'hits' or 'false alarms' to compute the correlation between a participant's accuracy and their self-reported rankings. If a participant is correctly expressing high confidence in accurate decisions, or low confidence in incorrect decisions, they possess metacognitive accuracy, and will display a high Meta d' score. Those participants who express confidence in incorrect trials, or low confidence in correct trials would have a low Meta d'. Meta d' is gaining popularity as measure of metacognitive sensitivity within implicit learning tasks (Maniscalco & Lau, 2012; Scott et al., 2014; Rausch & Zehetleitner, 2016).

Graded versus dichotomous self-report

Significant research within the field of Artificial Grammar Learning (AGL) supports the idea that awareness during implicit learning tasks is graded, not dichotomous (Sandberg et al., 2010). This is evident through the validation of graded subjective awareness scales such as the Perceptual Awareness Scale (Ramsay & Overgaard, 2004) and the Implicit Attribution Scale (Dienes & Scott, 2005).

CHAPTER 3: EXPERIMENTS

Research suggests both conscious and unconscious contributions to performance on implicit learning tasks. Despite their lack of use thus far within COVIS research, initial pilot data suggests that subjective measures of awareness could be a promising addition to the paradigm. Wismer and colleagues (2016) demonstrated clear differences in how participants verbalized or acknowledged their strategies between explicit and implicit tasks when using a subjective Implicit Attribution Scale. These results, in addition to the significant behavioral and neurophysiological data, serve as converging evidence supporting the separate system theory of category learning. Up until this point, there have been few attempts to integrate this type of measure within a categorization framework. This is clearly a significant gap in the COVIS literature. I believe that this offers the opportunity to provide significant converging evidence for the COVIS paradigm. I seek to demonstrate the usefulness of subjective measures, in tandem with behavioral and neurophysiological measures, in parsing out the underlying strategies and processes of implicit and explicit category learning.

The research will be carried out within the GRT framework, utilizing both an explicit rule-based and implicit information-integration decision bound approach. The studies will utilize simple perceptual stimuli made up of lines that vary in length and angle of orientation from trial to trial. The categories are based on those created by Ashby and Waldron (1999) and are composed of numerous category examples preventing memorization as a response strategy. On each trial of the category learning experiment, participants will be presented with a stimulus, sampled from one of the two categories, and must determine the category to which the stimulus belongs. Corrective feedback will be presented following each response, which provides the basis for learning the accuracy-maximizing decision rule.

Experiment 1: Blind Sight and Insight in Categorization

Rationale

According to COVIS, the implicit and explicit learning systems interact and compete for decisional control on each trial. Despite both systems always being active, this competition causes strategy switches over trials as participants test potential rules. Significant research shows that participants utilize different strategies and possess varying levels of strategy awareness when decisions are mediated by each of these learning systems (Wismer et al., 2016*;* Zlatkin, Patel, Wismer, Bohil, 2016). This decoupling of knowledge awareness and task performance has been termed *blind sight* and *blind insight* within the Artificial Grammar Learning literature. However, these phenomena have yet to be specifically examined within a categorization paradigm. *Blind Sight* occurs when participants achieve above chance accuracy on a task but are not able to verbalize or accurately assess their performance (Dienes, et al., 1995). In these instances, both the Guessing Criterion and the Zero Correlation Criterion would be satisfied. In a sense, they are 'blind' to their own knowledge or strategies used to accomplish the task.

Blind insight is the complementary phenomenon that has been shown to occur when participants who perform at chance possess high metacognitive accuracy (Scott et al., 2014). This is viewed as the reverse of Blind sight. Metacognitive Accuracy occurs when the Zero Correlation Criterion is violated, meaning that participants achieve a high correlation between their confidence and their decision accuracy, despite doing poorly in the task. For example, Scott and colleagues (2014) asked participants to group grammar strings as belonging to grammar A or grammar B in an Artificial Grammar Learning task. They measured both judgement accuracy and metacognitive accuracy (Meta d'). The results indicated that participants expressed high confidence in their correct trials and low confidence in their incorrect trials. In instances of *Blind Insight*, participants seem to possess a degree of awareness of their own poor performance, indicating some level of implicit knowledge of the category structures. I am interested in whether these phenomena can be seen within a decision bound categorization paradigm, and therefore offering further support to the separate systems explanation of COVIS.

Method

Participants

For this experiment, I collected a total of 65 participants. This number is based on pilot data (Wismer et al., 2016) which showed clear differences between the RB and II conditions. All participants were recruited through the University of Central Florida's online research hub (from Sona Systems) and received course credit for participation.

Variables and Measures

The independent variable is category structure (RB or II). The dependent variables are decision accuracy and metacognitive accuracy. Decision accuracy was recorded for each trial and averaged within each block. Metacognitive accuracy was computed using trial accuracy and confidence ratings to create a Meta d' score for each participant for each block.

Procedure

For this experiment, a two-category structure, between-subject categorization task was used. Participants were placed within either an explicit rule-based or implicit information-integration task utilizing two dimensional stimuli made up of lines of varying length and orientation. The experiment was presented on a desktop computer and programmed using MATLAB. The study is made up of six blocks of 80 trials each. There were four blocks of training, followed by two test blocks (Figure 3). In the training blocks, participants were presented with a stimulus sampled from one of the two category structures and required to classify it as either category A or category B. This method required participants to press a key labeled ‘A’ to classify a stimulus as category A and a key labeled ‘B’ to classify a stimulus as category B. Responses needed to be made within five seconds, during which the stimulus remained on the screen. Immediately following the response, corrective feedback was presented (‘Correct-this is A/ Incorrect-this is B’). The two test blocks followed, utilizing the same stimuli-response method, however, there was no feedback provided in these blocks. Instead, participants were prompted to rate their decision confidence on a 1-5 scale (1 being very unsure/ 5 very confident) immediately following decision response on number keys indicated on the keyboard. Two test blocks were used to ensure that there were enough trials to measure learning patterns.

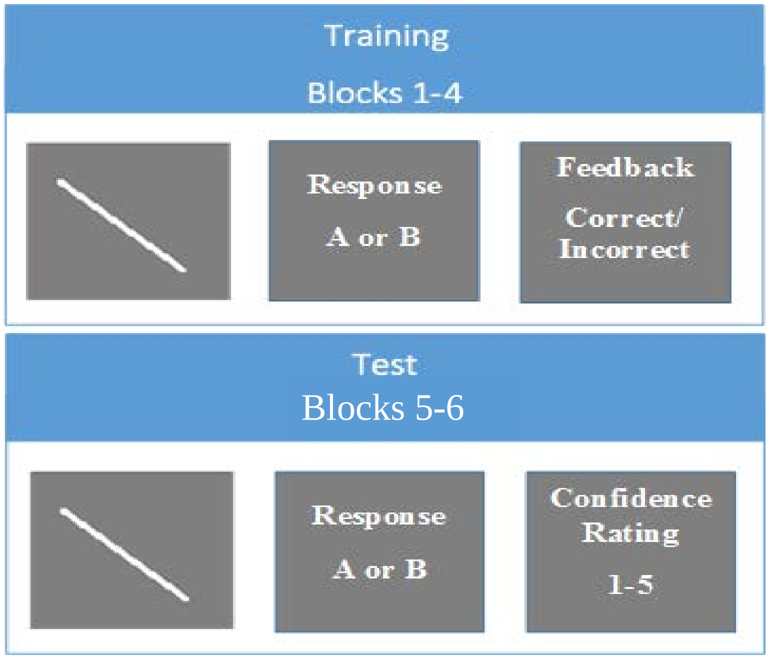


Figure 3 Experiment 1 Design

This Figure indicates the task structure for each trial in training and test blocks respectively

Results

I expected that learners (>55% accuracy) in both the RB and II category structures would possess high Meta d' scores. However, despite this overall prediction, I also made predictions based on individual differences in awareness. I hypothesized that more participants would display *blind sight* in the II structure, compared to the RB structure. Participants qualify as having blind sight if they achieve high performance despite low Meta d'. I based these predictions on previous research in this area showing participants in an implicit task can perform well despite not being able to verbalize or otherwise express their strategies.

Additionally, I predicted that some participants, despite being labeled 'non-learners', will possess higher than random Meta d' in the II structure, indicating the occurrence of *blind insight*. I do not expect this effect in the RB structure. I expected that RB non-learners would have low accuracy paired with low Meta d'. However, if incidents of blind insight do occur within RB non-learners, it could potentially be an indicator of participants using suboptimal implicit strategies for an explicit task.

Traditionally, RB tasks are considered easy to verbalize and are learned using explicit (consciously aware) category structure knowledge (Ashby & Maddox, 2005), for these reasons, I believe it is unlikely to see the same dissociations between awareness and performance that I predict in the II category structure.

Accuracy

Classification accuracy across blocks can be seen in Figure 4. Accuracy was condensed to provide more stable estimates of performance, from six 80 trial blocks, to three 160 trial blocks. The first two blocks are training blocks and block 3 is a test block.

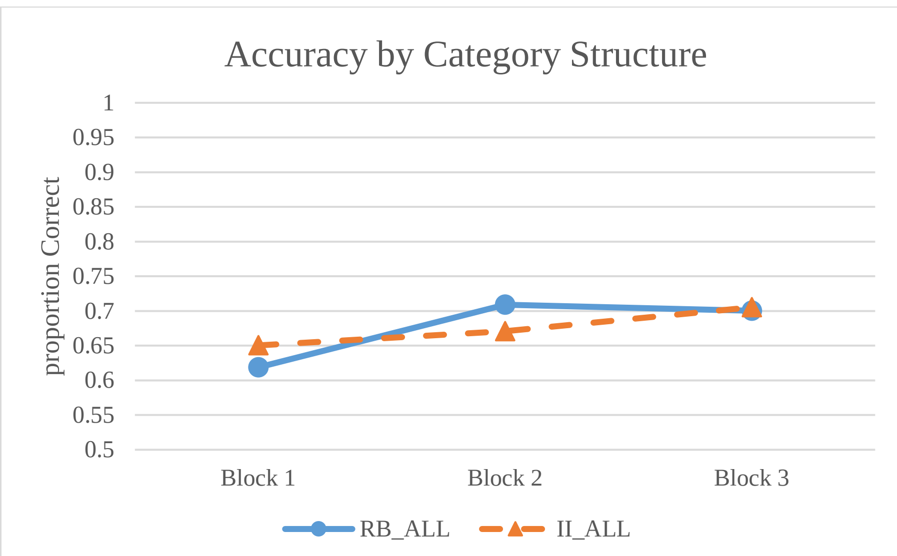


Figure 4 Experiment 1: Accuracy across blocks

Accuracy across blocks for both RB (blue) and II (orange) category structures shows the effect of asymptotic equivalence between structures.

There was a significant main effect of block on accuracy, $F(2,126) = 20.02$, $p < .001$, *partial* $\eta^2 = .241$, with accuracy increasing across blocks for both RB and II structures. This categorization task was designed under the assumption that both task conditions are roughly the same level of difficulty, as indicated by equivalent performance after learning reaches asymptote. The lack of a significant main effect of condition on accuracy, $F(1, 63) = 0$, $p = .989$, *partial* $\eta^2 = .000$, verifies that the category structures achieved asymptotic equivalence. As shown in Figure 4, both RB and II conditions reach around 70% accuracy by the final block. There was a significant interaction between block and condition, $F(2, 126) = 4.75$, $p < .05$, *partial* $\eta^2 = .070$. This is evident from the non-overlapping learning curves presented for RB and II accuracy. Pairwise comparisons show significant changes in accuracy from block 1 ($M = .62$) to 2 ($M = .71$) for RB ($p < .001$), however, a significant change does not occur for the II condition until block 2 ($M = .65$) to block 3 ($M = .71$, $p < .01$). Traditionally, as replicated here, explicit learning occurs earlier in the task and presents as a spike, then plateaus. In contrast, implicit learning has been shown to occur much more gradually across a task duration.

Meta d'

Confidence ratings from each trial in the test block (N=160) were transformed into a single Meta d' score using hit rate (HR) and false alarm rates (FAR). To preserve as much information as possible from the 1-5 confidence rating scale, a median-split frequency analysis was applied to each participant's responses to determine whether rankings were coded as a high confidence or low confidence. A hit occurred when a participant made a correct categorization response and rated their confidence as high. A false alarm occurred when a participant rated their confidence as high despite giving the incorrect categorization response for that corresponding trial. A participant's hit rate is the proportion of hit responses in a block ($N^{hits}/160$). The proportion of false alarms in a block is the False alarm rate ($N^{FA}/160$). These proportions were used to find Meta d' for each participant during the test block 3.

The formula for Meta d' is a modification of the standard d' formula used within signal detection theory.

$$\textit{Meta } d' = Z_{Hit} - Z_{FA}$$

When compared across all participants, there was no significant difference in the Meta d' scores of RB and II, $t(63) = .181$, $p = .857$. Participants were grouped as either learners or non-learners based in accuracy criterion of 55% (learners > 55% > non-learners). Comparisons between the learner and non-learner groups revealed a significant difference between groups on Meta d'. Non-learners ($M = 1.22$) had significantly lower Meta d' scores than those of learners ($M = 1.89$) $t(63) = 8.798$, $p < .001$. This is in line with my prediction that learners in both conditions would have high Meta d' scores.

Blind Sight

To examine variations in awareness at the individual level, I searched for incidents of blind sight within the data. A participant qualified as having blind sight by being a learner (>55% test accuracy) in conjunction with a low Meta d' score (Meta d' < 1). While there were cases of blind sight in both conditions, there were significantly greater proportion of blind sight in the II (.19, N= 7) condition compared to the RB (.05, N= 2) condition ($z=1.65$, $p<.05$). This supports my hypothesis that there would be a higher likelihood of dissociations between performance and awareness in the II condition.

Blind insight

To look for incidents of blind insight, a series of comparisons were done within only the non-learner group. Despite there being no significant difference in accuracy between RB and II non-learners, there was a significant difference in Meta d', $t(23)=3.64$, $p < .01$. As evident in Figure 5, II non-learners show greater accuracy at predicting their own performance than non-learners in the RB condition, indicated by the significantly higher Meta d' despite equivalent accuracy. This is evidence of implicit knowledge influencing behavior despite it not yet being discernable in task accuracy.

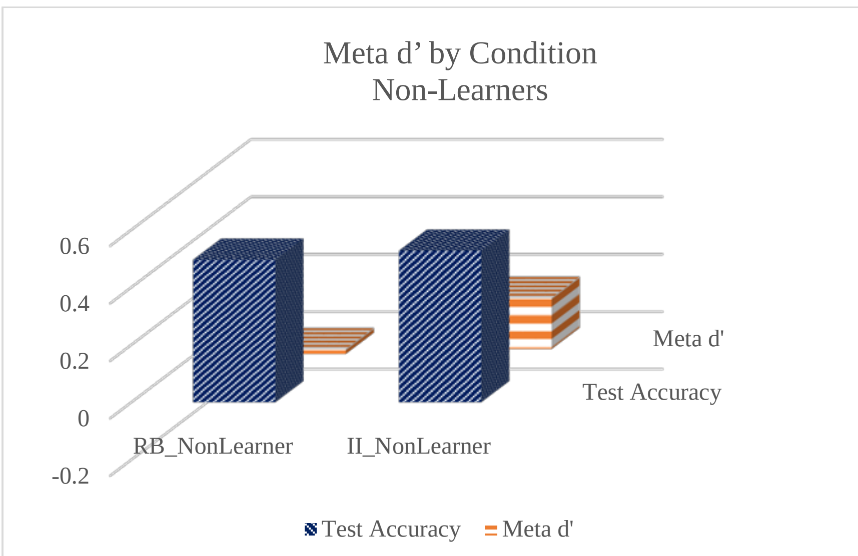


Figure 5 Experiment 1: Meta d' Non-learners

This Figure displays the block 3 accuracy and Meta d' for non-learners in each category structure. There were no differences between accuracy in the RB and II structure (blue), however, II Meta d' was significantly higher (orange).

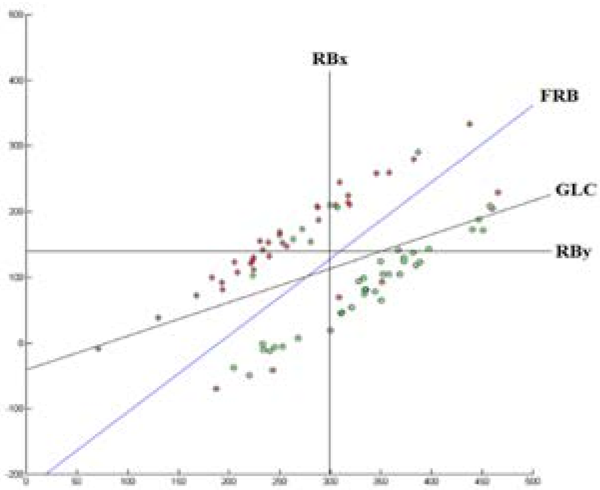
Modeling

I fit decision bound models to the data to instantiate and test hypotheses on decision strategies used throughout the task (Table 1). These models use maximum-likelihood estimation to estimate parameters of the user's decision rule along one or more dimensions to describe their performance. For example, these models indicate whether a participant is using the correct or incorrect 1-dimensional bound, or if they are integrating information from both dimensions. RBx and RBy are selective attention models in that accurate performance only requires the participant to attend to one dimension. Use of selective attention rules implies the participant is using conscious knowledge of the rule. FRB and GLC are information integration models (Figure 6).

Table 1 Summary of decision bound models

Model	**Slope**	**Intercept**	**Decision Bound**
RBx	infinite (vertical)	free	1D, vertical
RBy	0	free	1D, horizontal
FRB	Optimal	Free	1D for RB, slope = 1 for II
General Linear Classifier	free	free	1D for RB, slope = 1 for II
Optimal	Optimal	Optimal	1D for RB, slope = 1 for II
Random Responder	n/a	n/a	n/a

To perform well on an information-integration task, participants must pull information from multiple dimensions. Therefore, using one of these integration rules suggests a participant is relying on the implicit learning system. The random-responder model was fit to the data with the intention of serving as a criterion for grouping non-learners and learners. Pairing these modeling results with the performance data and Meta d’ scores allows assessment of the accuracy of each individual’s learning strategies, and further understanding of the relationships between objective performance and subjective experience throughout the learning task. For example, if a participant used a correct 2D integration bound, but possessed a low Meta d’ it could be indicative of blind sight.



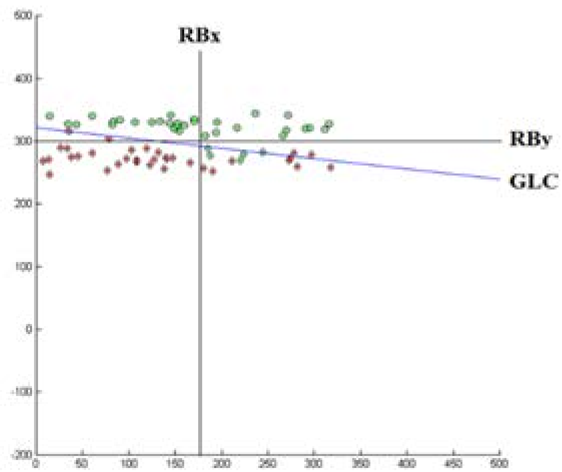


Figure 6 Decision Bounds by Model type

This figure shows examples of decision bound models for the information-integration category structures (left) and selective attention rule-based category structures (right).

The random-responder model corresponded highly with the accuracy criterion of 55% for learners/non-learners. However, the random responder model was not consistently effective at detecting non-learners who perseverated on one response throughout the task (i.e. pressing A repeatedly for every trial regardless of stimulus presentation). For this reason, as noted above, the accuracy criterion was a more useful grouping criterion for learners/non-learners than the random responder model.

These decision bound models were used to group participants based on whether they were optimal-learners, sub-optimal learners, or non-learners. Optimal-learners in the RB condition were those who achieved greater than 55% accuracy and were best fit by the OPT, or RBy RBy model. This indicated that they were utilizing a one dimensional bound (selectively attending to one dimension) required to do well on the task. All non-learners were classified best by the random responder model for RB. Those participants who achieved above chance performance (>55% accuracy) but were still best fit by the random responder model were considered sub-optimal learners.

Optimal-learners in the II condition were those best fit by the OPT, FRB or GLC models. Both the FRB and GLC are integration models, meaning that the decision bound utilizes information from both the length and orientation dimensions. II participants who were best fit by RBy, RBx were classified as sub-optimal. RBy and RBx are selective attention models and indicate that a participant was only utilizing one dimension in order to perform the task, which is a sub-optimal strategy within the II task. An example would be paying attention to whether the line was short or long and disregarding orientation. It is still possible to achieve above chance accuracy using an incorrect 1D rule in the II task. Participants in this case (best fit by RBx or RBy) were considered sub-optimal learners, as they achieved above chance accuracy while using a less than optimal strategy.

Accuracy by modeling comparisons

As described above, results from fitting the computational models to individual-participant response data were used to separate participants into three groups based on the type of model (decision bound) they were best fit by throughout the task: optimal-learners, sub-optimal learners, and non-learners. Accuracy comparisons were made using these new distinctions. A repeated-measures 2 (category structure) x 4(block) x 3(model grouping) ANOVA was run to examine the effects of model group on accuracy. This relationship between accuracy and model grouping between the two category structures is displayed in Figure 7.

A significant effect for model grouping on accuracy was shown $F(2, 60) = 77.73$, $p<.001$, *partial* $\eta^2 = .722$. Pairwise comparisons indicate that collapsed across structures, optimal-learners (M= .78) had significantly higher accuracy than both the sub-optimal learners (M= .62) and the non-learners (M= .52, $p< .001$). Pairwise comparisons showed the difference between sub-optimal learners and non-learners was less pronounced, but still approached significance (p=.055).

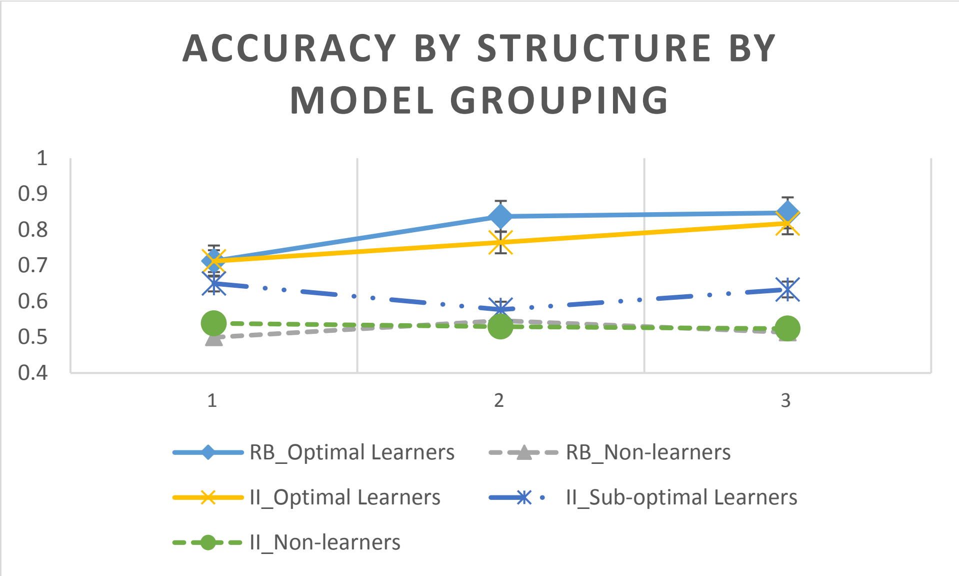


Figure 7 Accuracy by model group across structures

This figure shows accuracy broken down by model group across category structure. There were no RB Sub-optimal learners and therefore that group is not represented on the figure.

There was a significant interaction between the effect of block and model group on accuracy $F(2,60)=9.00$, $p<.001$ *partial* $\eta^2 = .231$. There was not a significant interaction between structure, model group and block, however, pairwise comparisons showed distinctions in the timeline of learning across structures, even when in the same model group. Within both the RB and II structures there were significant differences for optimal-learners across block 1 ($M_{RB}= .71$, $M_{II}=.71$) and 2 ($M_{RB}= .83$, $M_{II}= .75$), $p<0.001$), indicating that they were already learning the rule early on. Within the RB optimal-learners, there was no significant difference across blocks 2 ($M=.84$) and 3 ($M=.84$), indicating a plateau in accuracy. In contrast, within II optimal-learners there is still a significant change between blocks 2 ($M= .77$) and 3 ($M= .81$), indicating the slower learning curve. Interestingly, there were no significant changes across blocks 1 (M= .65), 2 (M= .57), or 3 (M= .63) for II sub-optimal learners. This suggests that despite achieving above chance performance, that learning for those using sub-optimal rules was slower and more gradual than that of the II optimal-learners.

Meta d’ by modeling comparisons

A 2 (structure) x 2(model group) ANOVA was run to examine the effects of model group on Meta d’. There was a significant main effect of model group on Meta d’ $F(2, 60) = 44.78$, $p<.001$, *partial* $\eta^2 = .559$. Optimal-learners (M= 1.94) had significantly higher Meta d’ scores than both the suboptimal-learners (M=.68) and non-learners (M=.131). There was not a significant interaction between model group and structure on Meta d’, $F(1, 60) = .200$, $p = .657$, *partial* $\eta^2 = .003$.

Discussion

The primary goal of experiment 1 was to validate the use of a novel measure of subjective awareness, Meta d’, within a categorization framework to examine the relationship between awareness and task performance. I predicted that Meta d’ would vary as a function of task performance, with lower Meta d’ being seen for non-learners and high Meta d’ for learners across structures. The data supported this, indicating that overall, Meta d’ is a good predictor of accuracy within a categorization framework.

Previous use of subjective awareness scales in implicit learning research have revealed the phenomenon of blind insight and blind sight. The secondary goal of experiment 1 was to use Meta d’ as a means to determine if these phenomena could be replicated within a perceptual categorization task. Participants qualified as having blind sight if they achieved high performance, despite low Meta d’. Results supported my hypothesis that more participants would display blind sight in the II structures, compared to the RB structures. These results are in line with implicit learning theories that describe information-integration tasks as unconscious and non-verbalizable. This indicated that Meta d' provided valuable information about differences in implicit and explicit learning previously unexplored within the COVIS categorization framework.

Meta d' also proved useful in determining cases of blind insight among the non-learners. I predicted that some participants, despite being labeled 'non-learners', would possess higher Meta d' in the II structure, indicating the occurrence of *blind insight*. Results supported this hypothesis, with II non-learners having significantly higher Meta d' scores compared to RB.

Further parsing out the data using decision bound models allowed me to make predictions on Meta d' based on whether participants were in optimal or suboptimal model groups. There was evidence to suggest that Meta d' is sensitive enough to show variations across modeling groups, however, these interpretations are limited in that there were no sub-optimal learners in the RB structure. Experiment 2 expands on the usefulness of Meta d' and provides a wider range of manipulations.

Experiment 2: Motor Response Mapping Disruption

Rationale

Many COVIS studies equate implicit learning with procedurally learned response encoding (e.g., consistent response button location learning). This effect has been shown through the manipulation of response location mapping in various ways. Research suggests that changes or interference with the motor response patterns of an implicit task can negatively influence implicit learning. Ashby, Ell, and Waldron (2003) found that by switching response key locations they could disrupt learning in an information-integration task, but not in a rule-based task. COVIS attributes implicit learning to the perceptual mapping of stimuli to response locations. Therefore, by removing the associated response positions, it is possible to significantly interfere with implicit learning strategies. Maddox, Bohil, and Ing (2004) expanded upon this by going one step further. Instead of just switching the response mapping, they designed an experiment that removed it entirely. Traditionally, research in this area is conducted using a forced-choice method, where participants are required to press a key designated for Category A or a key designated for Category B on a keyboard. Research suggests that this style of response allows participants to encode responses to a consistent location (e.g., A is right key, B is left key), enabling a stronger implicit connection to the categories (Ashby, Ell, &Waldron, 2003). Maddox and colleagues (2004) utilized a variable response mapping condition that had participants answer, 'yes or no' on designated keys on a keyboard (e.g., 'is this an A?"), instead of traditional A or B, while categorizing perceptual stimuli in RB or II tasks. This prevented a consistent stimulus mapping and greatly reduced implicit learning in II conditions. This change had no effect on the rule-based conditions.

Experiment 2 was designed to parse out the contributions of response location encoding on the implicit learning system, specifically when this procedural encoding system is disrupted. I was interested in further exploring where and how these changes interfere with the implicit learning process.

Method

Participants

For this experiment, I collected a total of 159 participants (N_{RB_AB}= 41; N_{RB_YN}=41; N_{II_AB}= 40; N_{II_YN}= 37). I conducted a power analysis using the effect size from Maddox, Bohil, and Ing’s 2004 study on variable response mapping due its similar design. The recommended sample size was 30 per cell based on statistical test ANOVA: Fixed effects, special, main effects and interactions; alpha = .05, power = .8, effect size = .535, # groups = 2). All participants were recruited through the University of Central Florida’s online research hub (from Sona Systems), and received course credit for participation

Variables and Measures

The independent variables are category structure (RB or II) and motor response mapping condition (A/B or yes/no). The dependent variables are decision accuracy and Meta d’. Decision accuracy was recorded for each trial and averaged across block. Meta d’ was computed using the same method described in experiment 1.

Procedure

For the purposes of this experiment, a between subjects two-structure category task, like that of experiment 1, was used. The experiment was presented on a desktop computer and programmed using MATLAB. The experiment was made up of six blocks of 80 trials each. The experiment alternated between training and test blocks (three of each). Participants were placed within either an explicit rule-based, or implicit information-integration task. However, I also added a motor response disruption condition, which relied on a variable response mapping method, in addition to the typical A or B response control condition. Those participants placed in the variable response mapping condition were asked to categorize simple perceptual stimuli as either category A or B, however, this condition used a ‘yes/no' response method. Participants were presented with a stimulus from one of the two category structures, followed by a prompting question of “Is this an A?” or “Is this a B?” They were required to press a key denoting ‘Yes’ or a key denoting ‘No’. The question asked on each trial was randomized, preventing participants from mapping the response locations consistently. Training block trials presented a stimulus, required a classification response, and then provided immediate corrective feedback on each trial. Test block trials presented the stimulus, required a classification response, and then immediately asked participants to rate their decision confidence on a 5-point scale (1 being very unsure/ 5 being very confident). The control condition utilized the same ‘A or B’ response method used in experiment 1.

I predicted that those in the disrupted Yes/No condition for II category structure would sufferer a major impairment in learning, resulting in lower accuracy, lower Meta d’ and less optimal learning strategies. This is based on the procedurally learned motor response encoding associated with implicit learning (Maddox et al., 2004). By removing the potential for consistent response mapping, I expected that the implicit learning system would be limited. I expected that this difference would be present in the II structures, but not within RB structures, due to previous research indicating that explicit tasks do not have the same reliance on procedural motor response location encoding. Therefore, it was expected that participants in the RB category structure would have high performance and high Meta d’ across both the AB and disrupted YN conditions.

I expected that participants in the II structure, disrupted motor response mapping condition (II_YN) would exhibit lower Meta d’ (e.g., less accurate at assessing their own performance), than those in the control condition (II_AB). I expected this subjective measure of implicit learning to mirror other validated behavioral measures across similar task conditions.

Additionally, I expected to see more incidents of sub-optimal learning strategies in the II structures. As in experiment 1, I expected to see a dissociation in performance and awareness in the II structure, evident through blind insight and blind sight.

Results

Accuracy

Accuracy across the three test blocks for both RB and II can be seen in Figure 8. There was a significant main effect of block on accuracy F (2, 310) = 35.101, p <.001, *partial* η^2 = .185. Accuracy significantly increased from block 1 (M= .612), block 2 (M= .679), and block 3 (M=.695). There was not a significant interaction between block and structure, F(2,310)= .116, p= .890, *partial* η^2 = .001. There was not a significant interaction between block and response method, F(2,310)= .057, p = .945 *partial* η^2 = .001. There was a significant main effect of structure F(1, 155) = 6.178, p <*.05*, *partial* η^2 = .038. Accuracy was higher in the RB conditions (M=.69) than in the II conditions (M=.63). There was not a significant main effect of response method, F(1, 155)= 1.987, p = .161, *partial* η^2 = .017. There were no significant differences between the RB_YN (M= .66), II_AB (M= .63) and II_YN (M= .67) conditions. However, the RB_AB (M= .73) condition showed significantly higher accuracy than the RB_YN (p <.05) and the two II structure conditions (p<.01). This is counter to my expectation that there would be no difference between the two response methods in RB. There was no interaction between structure and response method on accuracy F(1,155) = 2.632, p = .107, *partial* η^2 = .017.

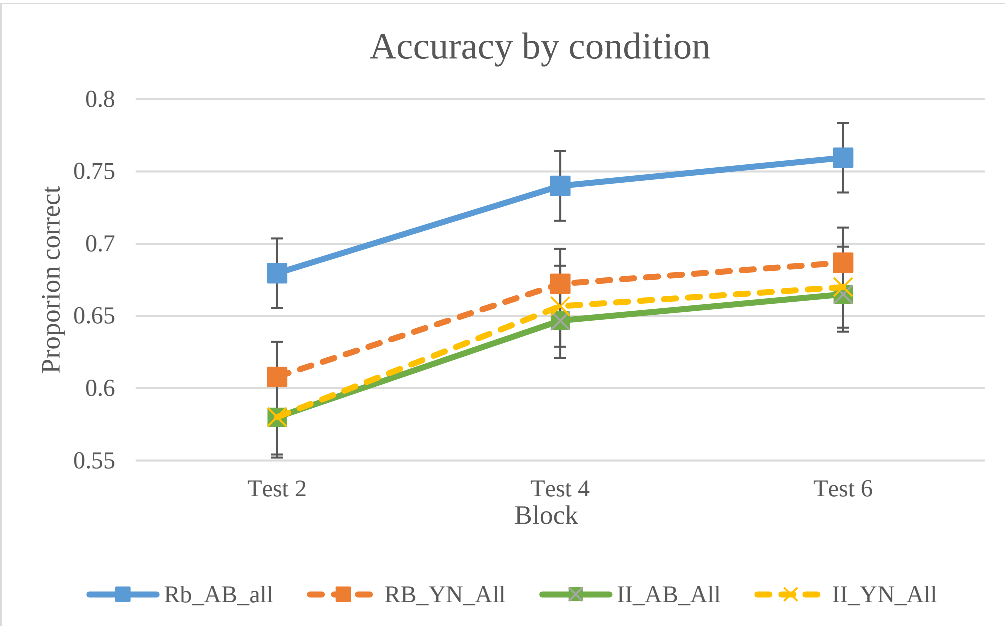


Figure 8 Accuracy by condition

This figure shows the accuracy across the three test blocks for each of the four category structures. RB_YN was significantly higher than the other three.

Meta d'

Confidence ratings were transformed into Meta d' scores for each of the three test blocks. This was done using the same method described in experiment 1. Results indicated a significant main effect of block on Meta d' F (2, 310) = 36.342, p <.001, *partial* η^2 = .190. Consistent with the pattern shown in experiment 1 data, Meta d' increased across blocks for all conditions. Pairwise comparisons revealed significant changes from blocks 1 (M= .60) to 3 (M= 1.17) for all conditions, however, the II_YN condition did display the smallest increase. This is potential evidence of a disruption in metacognitive accuracy and is explored further in the breakdown by learners and non-learners below. There was no significant interaction between block and structure on Meta d', F(2, 310)= .325, p =.723, *partial* η^2 =.002. There was also no significant interaction between block and response method, $F(2, 310) = .552$, $p = .576$, *partial* $\eta^2 = .004$. There was a significant main effect of structure, $F(1, 155) = 6.529$, $p < .05$, *partial* $\eta^2 = .512$. Meta d' was higher in the RB structures (M=1.13) than in the II structures (M=.75) There was no significant main effect of response method on Meta d', $F(1, 155) = 1.184$, $p = .278$, *partial* $\eta^2 = .008$. As predicted, there was no significant difference between the Meta d' of RB_AB (M= 1.30) and RB_YN (M=.96). Contrary to predictions, there were also no significant differences between RB_YN and II_YN (M= .77) or II_AB (M=.72) and II_YN. There was no significant interaction between structure and response method on Meta d', $F(1,155) = 2.011$, $p = .158$, *partial* $\eta^2 = .013$.

To better explore the interaction of Meta d' on structure and response method participants were grouped as either learners or non-learners using the same criteria as experiment 1 (<.55% accuracy). A 2(structure) x 2 (response method) x 3(block) ANOVA was conducted using learners only. There was a significant effect of block on Meta d', $F(2, 212) = 45.396$, $p < .001$, *partial* $\eta^2 = .300$. There was not a significant interaction between block and structure, $F(2,212) = .865$, $p = .571$, *partial* $\eta^2 = .008$. There was no significant interaction between block and response method, $F(2,212) = .562$, $p = .571$, *partial* $\eta^2 = .005$. There was a significant effect of structure on Meta d' in learners, $F(1,106) = 12.987$, $p < .001$, *partial* $\eta^2 = .109$. Meta d' was higher in the RB conditions (M=.1.63) than in the II conditions (M=.1.09). There was not a significant effect of response method, $F(1,106) = 3.076$, $p = .082$, *partial* $\eta^2 = .109$, or a significant interaction between structure and response method, $F(1,106) = 2.988$, $p = .087$, *partial* $\eta^2 = .027$. RB_AB learners (M= 1.91) had significantly greater Meta d' than RB_YN (M=1.36), II_AB (M= 1.16), and II_YN (M= 1.13). However, planned comparisons revealed that II_YN is the only condition not to have a significant increase of Meta d' between blocks 2 and 3 (Figure 9). This does provide some support to the hypothesis of disrupted implicit knowledge. The II_YN participants seem to be showing less improvement in predicting their own performance compared to the non-disrupted control (II_AB) and RB conditions.

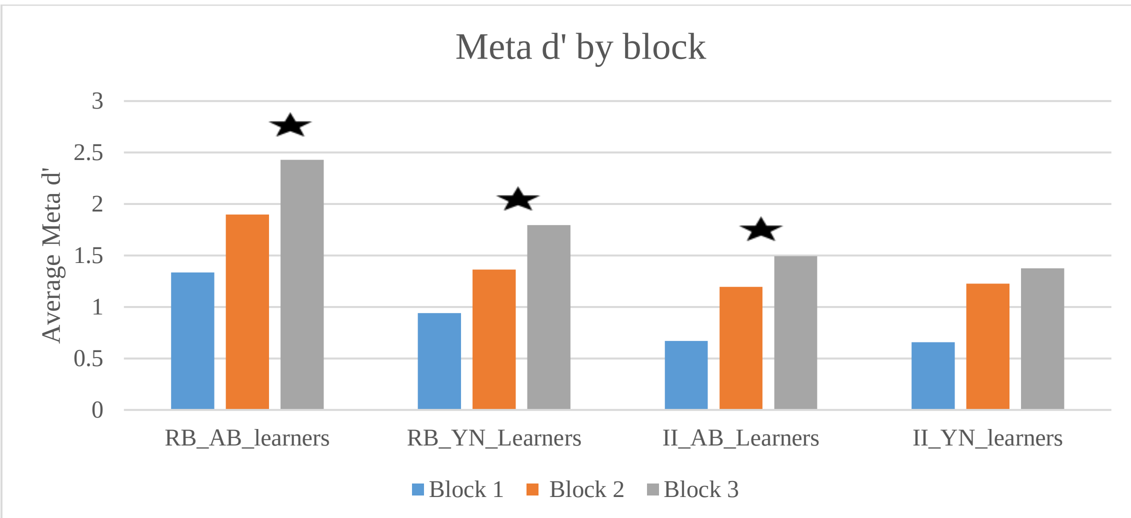


Figure 9 Meta d' by block across conditions

This figure shows the increase of Meta d' across blocks for all 4 conditions. Significant changes from block 3 to 4 are denoted with a star.

Blind Sight

I examined variations in awareness at the individual level by searching for incidents of blind sight within the data. As in experiment 1, a participant qualified as having blind sight by being a learner (>55% test accuracy) in conjunction with a low Meta d' score (Meta d' < 1). Results indicated that there were significantly greater proportion of participants displaying blindsight across the two RB conditions (.05) compared to both II (.29), $z= 4.83$ $p<.01$. Additionally, there were significantly greater proportion of blind sight cases in II_YN (.30) compared to RB_YN (.09) ($z= 2.35$, $p<.05$). This also held true for the AB condition with II_AB

(.175) having a greater proportion of blind sight than RB_AB (0) conditions z= 3.79, *p*<.01 (Figure 10). This data supports my hypothesis that there would be a greater dissociation between awareness and performance in II conditions, resulting in a higher proportion of blind sight.

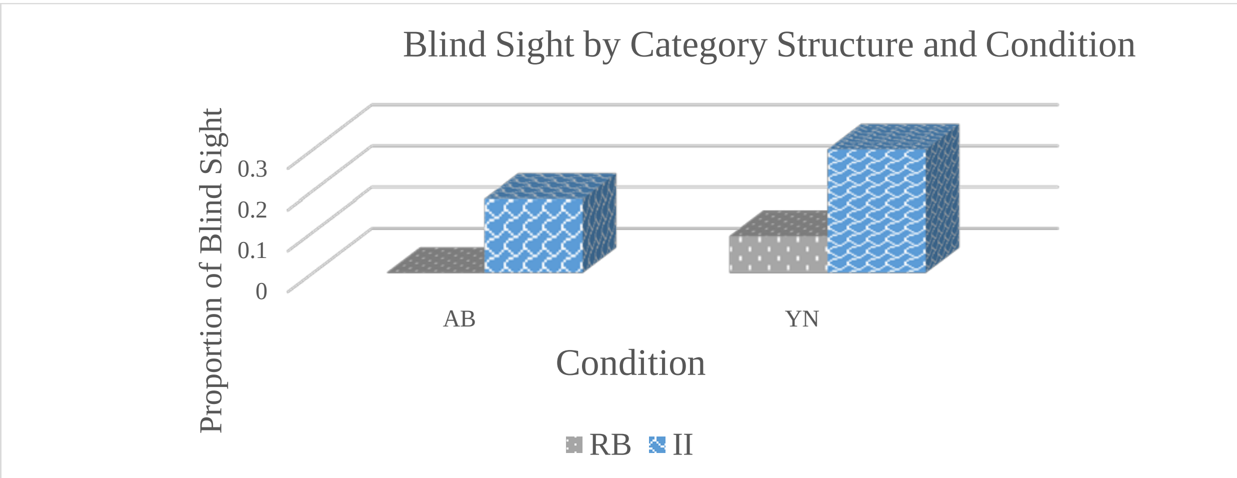


Figure 10 Blind sight by structure and condition

Blind Insight

There were no significant differences between the Meta d' values of RB_AB (*M*= -.13) and II_AB (*M*= -.20), *t*(22)= .615, *p*=.55.The difference between the Meta d' of non-learners in the YN condition were also nonsignificant, *t* (21) = 1.95, *p*= .06, with the RB_YN (*M*= -.10) Meta d' being lower than that of the II_YN (*M*=.07) condition.

Modeling

I fit all individual response data using the same computational models utilized for experiment 1. Participants were grouped as optimal learners, sub-optimal learners, and non-learners and a 2 (structure) x 2(response method) x 3 (block) x 3 (model group) ANOVA was conducted on the data. There was a significant effect on Meta d' for model group F (1, 148) = 57.04, p<.001, *partial* η^2 = .435. Collapsed across conditions, optimal learners (M= 1.37) had significantly greater Meta d' than both sub-optimal learners (M= .89) and non-learners (M=.04). There was a significant interaction between model group and block F (4, 296) = 15.197, p<.001, *partial* η^2 = .161. Planned comparisons revealed that there was a significant difference in the Meta d' of optimal and sub-optimal rule users in block 3 (with higher Meta d' for optimal learners) for all conditions, with the exception of II_YN (p=.09), as evident in Figure 11. This provides further evidence of an implicit disruption of metacognitive accuracy, supporting my predictions.

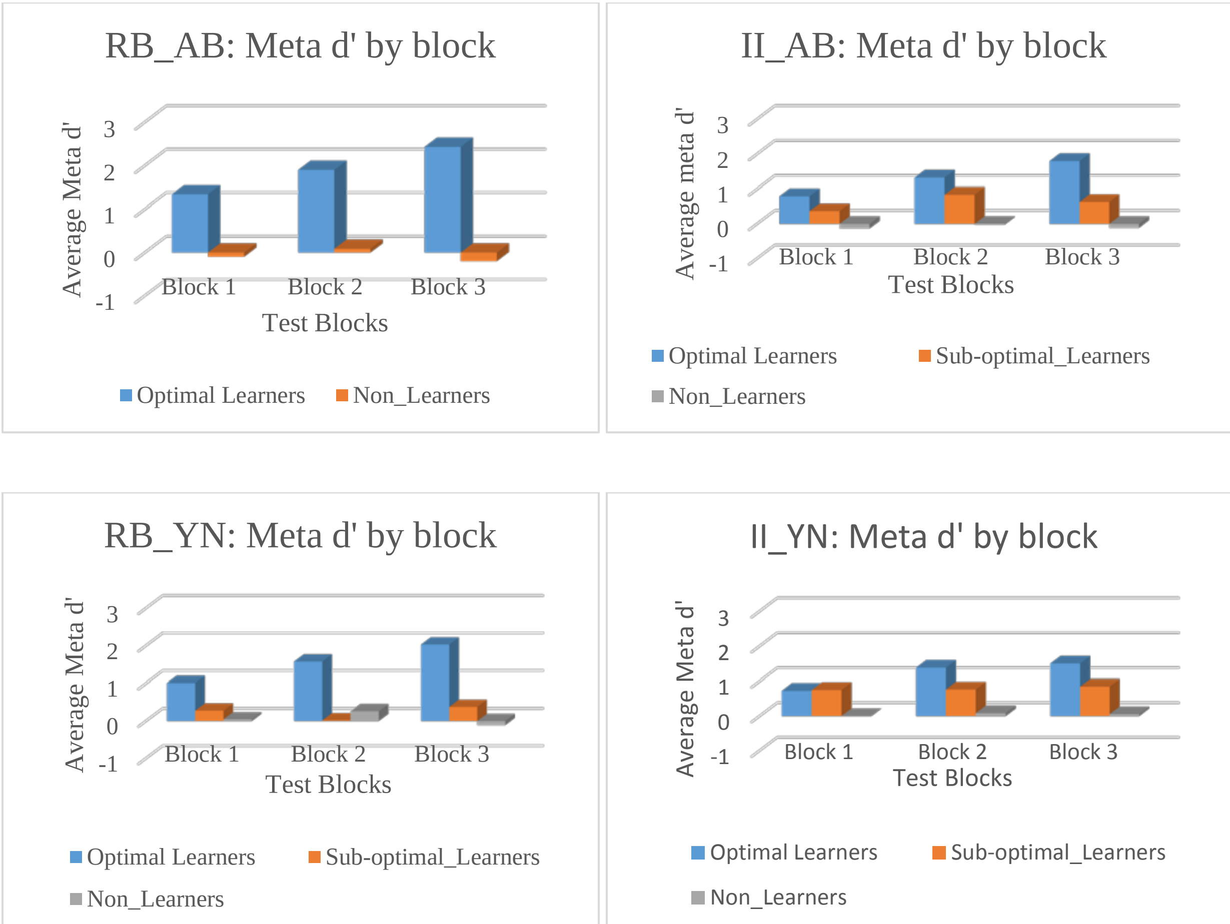


Figure 11 Meta d' by block across conditions

This figure shows the increase of Meta d' across blocks for all 4 conditions.

Discussion

The primary objective of experiment 2 was to examine the relationship between awareness and performance in an implicit task under the manipulation of a disruption to the procedurally based motor response mapping. Previous research (Maddox et al., 2004) demonstrated a severe impairment to implicit learning when motor response mapping was removed and replaced with a Yes/No variable response method. My attempts to replicate these findings with the addition of a subjective measure, Meta d', were not successful. There was no significant difference in accuracy between the II_AB condition, using a consistent motor response mapping method, and the II_YN condition, using the disrupted variable response mapping method. This was contrary to my predictions on performance based on condition. One potential explanation for this unexpected result is that the introduction of a confidence rating scale after each trial allowed participants more time to think about each trial, by requiring the stimuli to be held in working memory throughout the rating phase. This would introduce some level of explicit processing, therefore negating any potential implicit decrement if participants utilized selective attention rules. As it is possible to achieve above chance performance on an II task using selective-attention rules, this could explain the lack of performance decrement within the II_YN condition. Further investigation into the effects of a confidence scale on categorization learning is needed.

Despite the failure to replicate the implicit learning disruption, interesting effects emerged when examining the difference in Meta d’ across these manipulations. There is reason to suggest that while there was no visible decrement to type 1 task performance (accuracy) in II_YN due to the response method manipulation, there seems to be a metacognitive accuracy impairment as seen in the Meta d’ scores between conditions. Meta d’ scores for those in the II_YN condition with disrupted variable response mapping had significantly lower Meta d’ than the RB counterparts. This suggests there is a disruption in the implicit awareness of category structure. RB_YN participants were significantly better at predicting their own performance. This effect was further supported when I looked for cases of blind sight among the groups. There were significantly more participants experiencing blind sight in the II_YN condition compared to the RB_YN. This supports my hypothesis that there would be a greater dissociation in awareness and performance in the II conditions, particularly II_YN.

Meta d' proved even more informative when looked at across groups designated by computation modeling fits. Decision bound models allowed me to group participants as either optimal-learners, suboptimal-learners, or non-learners. Meta d' proved a sensitive measure of metacognitive accuracy, with differences between the three model groupings being significant across both structures. This showed that there is a great deal of information being missed out on when performance is just assessed at the accuracy level. This provides further evidence supporting the use of Meta d' as a valid means of assessing metacognitive accuracy through a perceptual categorization task. Additionally, Meta d' varied uniquely across category structures, providing valuable information in support of the COVIS model of implicit and explicit learning.

Experiment 3: Motor Response Mapping Disruption with fNIRS

Rationale

As previously emphasized, COVIS is the only categorization learning theory that integrates a neurophysiological component allowing for predications based on brain structure. The lack of subjective measures being used within the COVIS literature means that no one has looked at the neurophysiological changes associated with implicit and explicit learning in combination with confidence ratings as a measure of task awareness. I seek to rectify this gap.

Method

Experiment 3 was a replication on the design of experiment 2, with a neurophysiological expansion. Due to the greater trial count necessary for a neuroimaging experimental design, I utilized a within subjects two-structure categorization task with participants completing both the

RB and II structure tasks. Experiment 3 utilized only the disrupted variable response mapping method (Yes/No responding, no A/B responding conditions). The experiment was presented on a desktop computer and programmed using MATLAB.

Participants

I conducted a power analysis using the effect size from Maddox, Bohil, and Ing's 2004 study on variable response mapping due its similar design. The recommended sample size was 30 based on statistical test ANOVA: Fixed effects, special, main effects and interactions; alpha = .05, power = .8, effect size = .535, # groups = 2). For this experiment, I collected a total of 42 participants. Of these participants, 33 yielded useable fNIRS data. Nine participants were removed due to poor or incomplete data. All participants were recruited through the University of Central Florida's online research hub (from Sona Systems) and received course credit for participation. Participants were screened for handedness, only right-handed participants were used.

Variables and Measures

The independent variable was category structure (RB/ II). The dependent variables were decision accuracy, Meta d', and hemodynamic response in the PFC. Metacognitive accuracy (Meta d') was computed using the same method described for both experiment 1 and 2.

Procedure

This experiment was made up of four blocks of 160 trials each. Each block contained alternating sets of category learning trials and basic perceptual baseline measurement trials (classifying lines as blue or yellow). Like in experiment 2, there were alternating training and test blocks. Based on the accuracy plateauing after block 4 in experiment 2, only 4 blocks were used for this experiment. This reduction in blocks allowed me to keep each session under 3 hours, as it was within subjects. Training block trials presented a stimulus, required a response, and then provided immediate corrective feedback. This corrective feedback provided the basis for learning the accuracy maximizing rule. Test blocks presented a stimulus, required a response, and then immediately asked participants to rate their decision confidence on a 5-point scale (1 being very unsure/ 5 being very confident). There was a jittered ITI, averaging four seconds, due to the design required for neuroimaging. I collected neurophysiological data using functional Near Infrared Spectroscopy, with a focus on the Prefrontal Cortex (details provided below).

Functional Near Infrared Spectroscopy

I evaluated task-related differences in prefrontal cortex activity using functional near infrared spectroscopy (fNIRS). fNIRS neuroimaging is comparable to functional magnetic resonance imaging (fMRI); both measure hemodynamic changes (the Blood Oxygen Level Dependent, or BOLD, response; Cui, Bray, Bryant, Glover, & Reiss, 2011) in response to neural activity. However, fNIRS measures changes in light absorption while fMRI measures changes in magnetic properties of hemoglobin. Despite a weaker signal to noise ratio than fMRI, fNIRS measurements correlate strongly with fMRI measurements across a large range of cognitive tasks (Cui et al., 2011).

fNIRS sources send pulses of near-infrared light into the top layer of the cortex through the scalp. Detectors collect the amount of light reflected back to the surface. The light absorption rate differs for oxygenated and deoxygenated hemoglobin, and therefore provides an indication of neural activity changes.

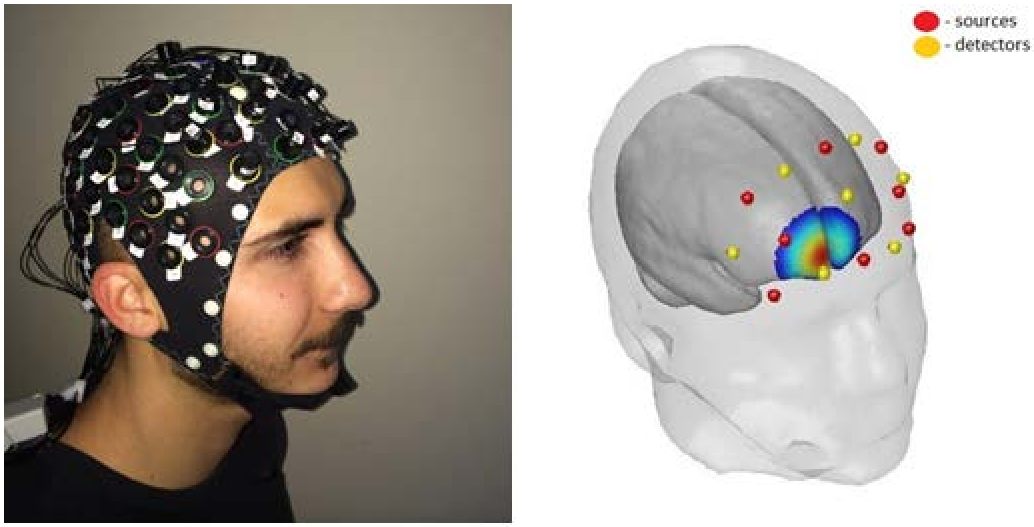


Figure 12 fNIRS cap and probe layout

Figure 12 (left) depicts a participant wearing an fNIRS cap (from NIRx Medical Technologies, LLC). The topographic image (right) shows a probe mapping used to measure activity in the prefrontal cortex.

fNIRS is portable and relatively non-invasive, allowing for research to take place in a greater variety of operational environments. Generally, the lightweight headgear consists of flexible caps or bands with re-configurable light sources and detectors (Figure 12). Participants report that the caps are snug fitting but with little discomfort. Trained researchers can complete the fNIRS set-up process in approximately 20 minutes.

fNIRS is relatively robust to movement artifacts, in contrast to fMRI which requires complete immobility of study participants. fNIRS is also relatively unobtrusive. It is silent and is highly tolerable to participants. fNIRS has been validated as an effective tool for studying both normal brain function and a variety of pathologies within a diverse range of populations, including sensitive populations such as older adults and young children (Ferrari & Quaresima, 2012).

fNIRS data acquisition.

Specific to my study, neuroimaging data was acquired using a 20-channel functional near-infrared spectroscopy system from NIRx Medical Technologies, LLC (NIRSport 88 system). Eight light sources and eight detectors combine to provide 20 measurement channels located over the prefrontal cortex. Source LEDs emit near-infrared light at wavelengths of 760nm and 850nm. The sampling rate for fNIRS data acquisition is 7.81Hz. Data was collected using NIRStar acquisition software (http://nirx.net/nirstar-1).fNIRS channels measure task-related concentration changes in oxy Hb, Hb, and total hemoglobin. For the proposed study, measurement channels recorded bilaterally over superior frontal gyrus (Brodmann area 8), dorsolateral prefrontal cortex (DLPFC; BA 9), anterior PFC (APFC, BA10), ventrolateral PFC (VLPFC; inferior frontal gyrus, BA 45), and middle frontal gyrus (BA 46).

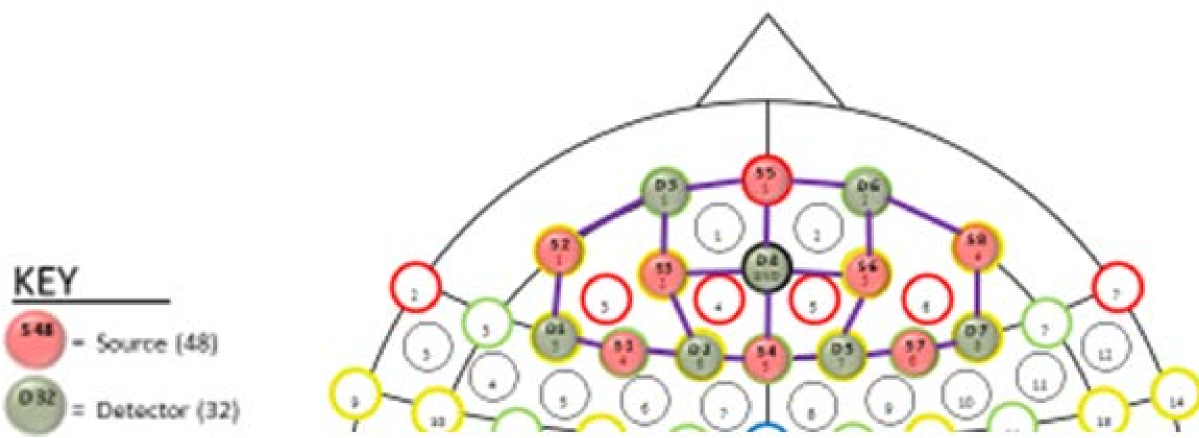


Figure 13 fNIRS data collection channel layout.

Red circles represent optode locations. Green circles represent detector locations. Blue lines indicate the data collection channels created by each Source /Detector pair.

There are substantial differences in the anatomical brain structures between participants. Without structural MRI scans, the localization to Brodmann areas is only approximate. To minimize inaccuracies, participants were measured and fitted with caps sized to provide a snug fit. Cap placement used anatomical landmarks based on the international 10-20 system commonly used for EEG data collection.

Preprocessing

Several preprocessing filters were applied to the data before statistical analysis. Spike artifacts and discontinuities are detected and eliminated automatically using NIRSLab analysis software. A low pass filter was applied to reduce interference from heart rate and respiration. Finally, a motion artifact reduction filter was applied. After preprocessing, raw optical data values were transformed into estimates of oxygenated hemoglobin (oxyHb) concentration changes using the modified Beer-Lambert Law (MBLL) in nirsLAB v2016.01 (http://nirx.net/nirslab-1). Subjects with greater than 50% channel oversaturation were eliminated from the analysis due to poor signal quality.

General linear model analysis

I conducted a general linear model analysis, using nirsLAB's version of SPM8 (Bang et al., http://www.fil.ion.ucl.ac.uk/spm/software/spm8/2016). At the individual subject level analysis (level 1), I included regressors for training phase and testing phase for each of the 4 blocks. For each individual data set, a canonical HRF was convolved with a boxcar function to model task-related activity. I removed serial correlation by precoloring with a Gaussian kernel (FWHM = 4s). Regression ß values estimated during level 1 SPM were extracted to be used for statistical comparisons at the group level in SPSS.

In conjunction with the predictions from Experiment 2 for accuracy and confidence, I expected that participants would exhibit higher PFC activation in the II structure due to the motor disrupted response method (Yes/No), compared to the RB structure. I believe that by disrupting the motor response location encoding (an implicit structure), it would force participants to fall back on the less optimal explicit learning system to complete the task. I expected this to result in a longer sustained activation pattern throughout the test blocks of the task (mostly visible in terms of higher PFC activation in the final test block in the Yes/No condition) (see *Figure 14*). I believe this would add to the separate systems debate by showing coinciding changes in confidence and BOLD response due to implicit disruptions, findings primarily supported by behavioral accuracy data up until now. These predictions are in part based on the procedurally learned motor response encoding associated with implicit learning (Maddox et al., 2004), as well as numerous studies showing neurophysiological dissociations between implicit and explicit learning tasks (Ashby & Maddox, 2005). By removing the potential for consistent response mapping, I predicted that the participant will be forced to fall back on a less optimal explicit learning strategy, therefore resulting in a greater working memory load, indicated through sustained activation throughout the task.

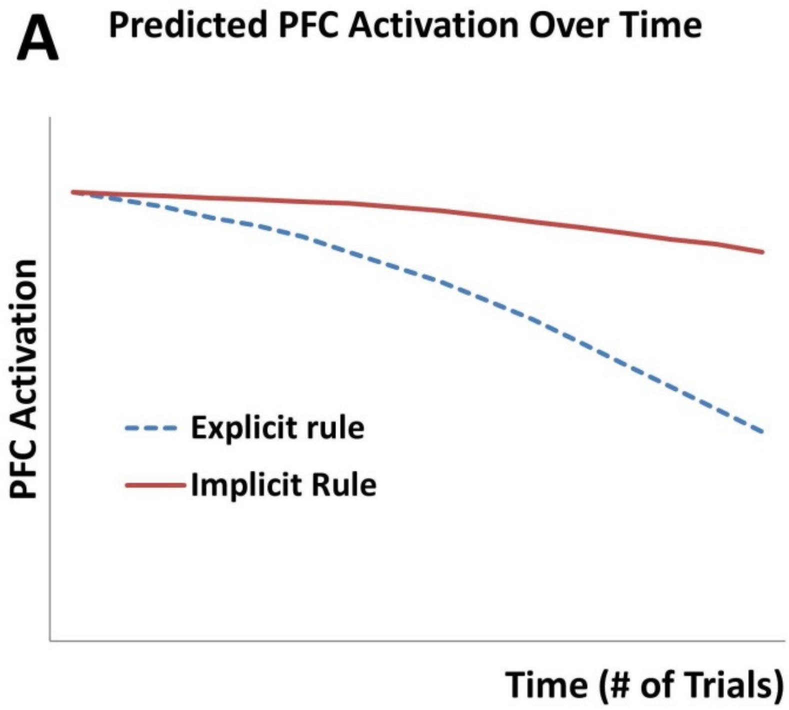


Figure 14 PFC activation predictions

The red line is representing PFC activation for the II condition. It is predicted that II activation will remain high across the task due to the interference in motor response encoding severely disrupting learning. The blue dashed line represents RB activation, which is predicted to be unaffected by the motor response disruption and drop drastically upon learning the rule.

Results

Accuracy

Accuracy across blocks for both RB and II category structures can be seen in Figure 15. This includes all participants for each structure. There was significant main effect of block on accuracy F (3,96) = 17.403, p<.05, *partial* η^2 = .352.

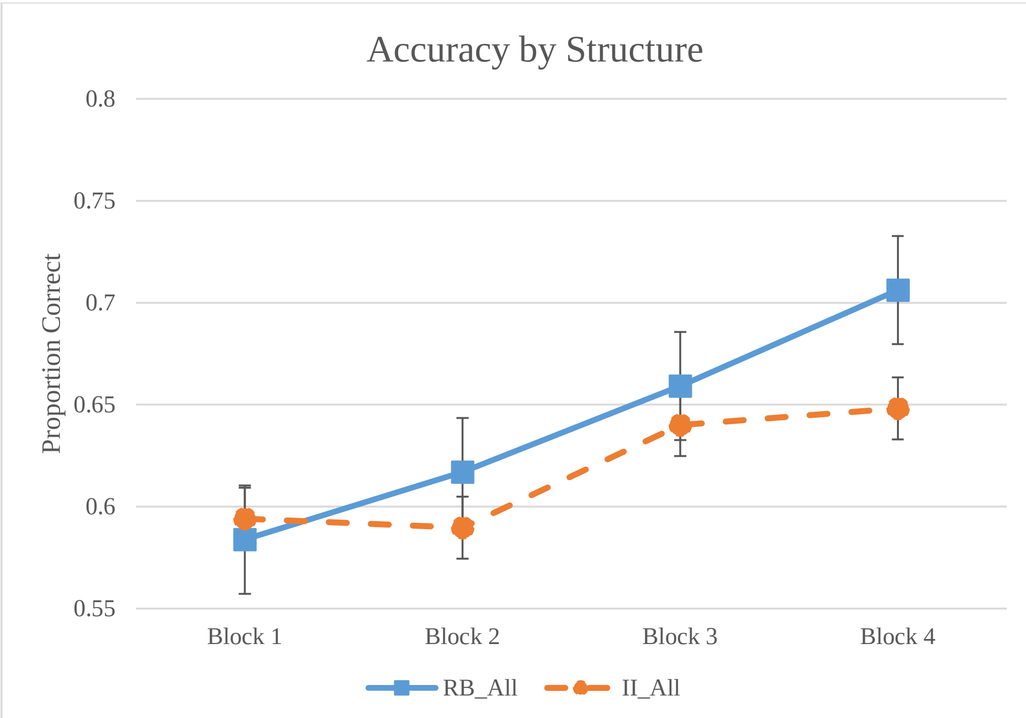


Figure 15 Accuracy by Structure across blocks

There was not a main effect of structure, $F(1, 32) = .937$, $p=.34$, *partial* $\eta^2 = .028$ The predicted interaction between structure and block approached significance, $F(3, 96)= 2.41$, $p=.084$, *partial* $\eta^2 = .180$. Pairwise comparisons revealed that there were significant changes for the RB structure between block 1($M= .58$) and block 3 ($M=.65$, $p <.05$)), and block 3 and 4 ($M= .71$, $p<.01$). There were no significant changes between blocks 3 and 4 in the II structure. This indicates that there was much less learning across II overall. This supports my hypothesis that the disrupted motor mapping would impair implicit learning in the II structure but not effect RB performance. This matches my initial hypothesis about these conditions and replicates Maddox and colleague's 2004 study.

When broken up into learners and Non-learners, based on the same 55% accuracy criterion used in previous experiments, the accuracy trends remain the same. This break down can be seen in Figure 16.

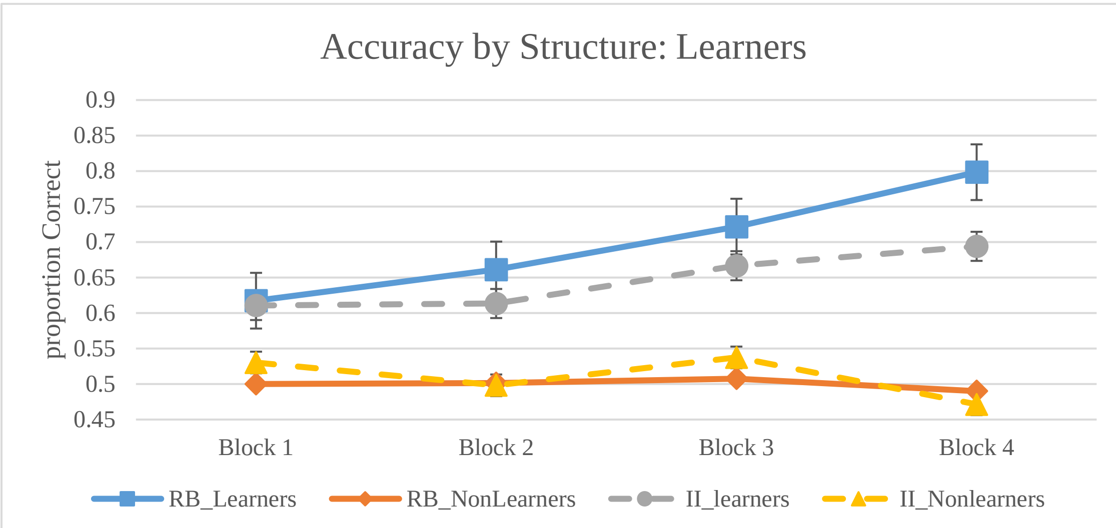


Figure 16 Accuracy by category structure and learner/non-learner designation

Meta d'

Confidence ratings were transformed into Meta d' scores for both test blocks (block 2 and 4). This was done using the same method described in experiment 1 and 2. A 2 (structure) by 2 (block) ANOVA was conducted and indicated a significant main effect of block on Meta d' $F(1, 32) = 21.120$, $p < .01$, *partial* $\eta^2 = .398$. Meta d' increased across blocks for both category structures (Figure 17). This is consistent with the results of experiment 1 and 2. There was a marginally significant main effect for structure, $F(1,32) = 4.075$, $p = .052$, *partial* $\eta^2 = .113$, with the RB structure ($M = .989$) have greater meta d' than that of the II structure ($M = .685$). This is in support of my hypothesis of the II structure performance suffering a decrement due to the use of variable response mapping. There was not a significant interaction between category structure and block, $F(1, 32) = .001$, $p= .4975$, *partial* $\eta^2 = .000$.

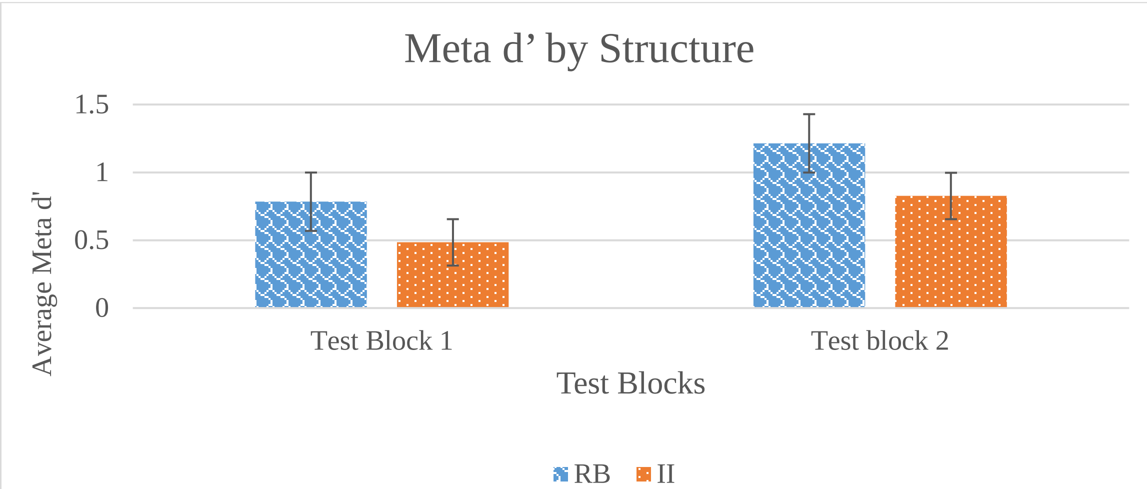


Figure 17 Meta d’ across blocks for each of the category structures.

Blind Sight

I examined variations in awareness at the individual level by searching for incidents of blind sight within the data. As in experiment 1 and 2, a participant qualified as having blind sight by being a learner (>55% test accuracy) in conjunction with a low Meta d’ score (Meta d’ < 1). Results indicated that there were significantly greater proportion of participants displaying blind sight across the II structure (.44) compared to the RB structure (.12), $z= -2.99$ $p<.01$.

Blind Insight

Because the experiment was within subjects the breakdown based on accuracy resulted in comparison groups that did not consistently contain the same subjects for RB and II structures. As a result, I decided to compare groups using a nonparametric statistical test. A Mann-Whitney U test indicated that the Meta d’ distributions between RB non-learners (M = -0.07) and II non-learners (M = 0.16) were not significantly different (Mann–Whitney U = 2, n1 = 10 n2 = 6, P = 0.183 two-tailed).

Modeling

I fit all individual response data using the same computational models utilized for experiment 1 and 2. Within each structure, participants were grouped as optimal-learner, sub-optimal learner, or non-learner. A break-down of accuracy by model group (Figure 18) demonstrates the effect using a sub-optimal rule has on performance within each category structure.

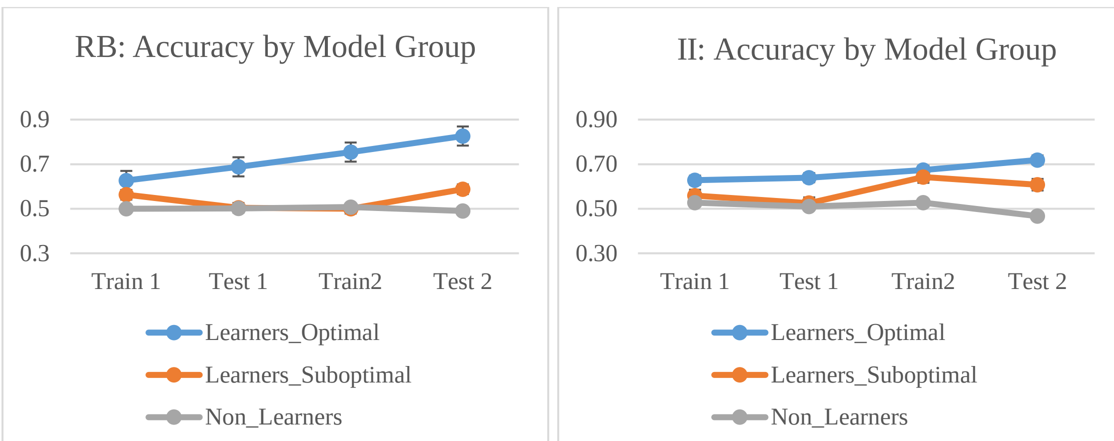


Figure 18 Accuracy by model group for RB and II category structures

Meta d' also varies as a function of model group, consistent with the results shown in experiment 1 and 2. The break-down of Meta d' across the two test blocks within the RB and II category structures can be seen in Figure 19. Because the experiment was within subjects the breakdown based on model fits resulted in comparison groups that did not consistently contain the same subjects for RB and II structures. As a result, I decided to compare groups using a nonparametric statistical test. A Mann-Whitney U test indicated that the Meta d' distributions of the RB Optimal rule users differed significantly from that of the II optimal rule users (Mann–Whitney $U = 2$, n1 = 23 n2 = 2, $P < 0.01$ two-tailed). There was not a significant difference between structures at the sub-optimal level. This is further support of an implicit disruption being evident at the metacognitive level.

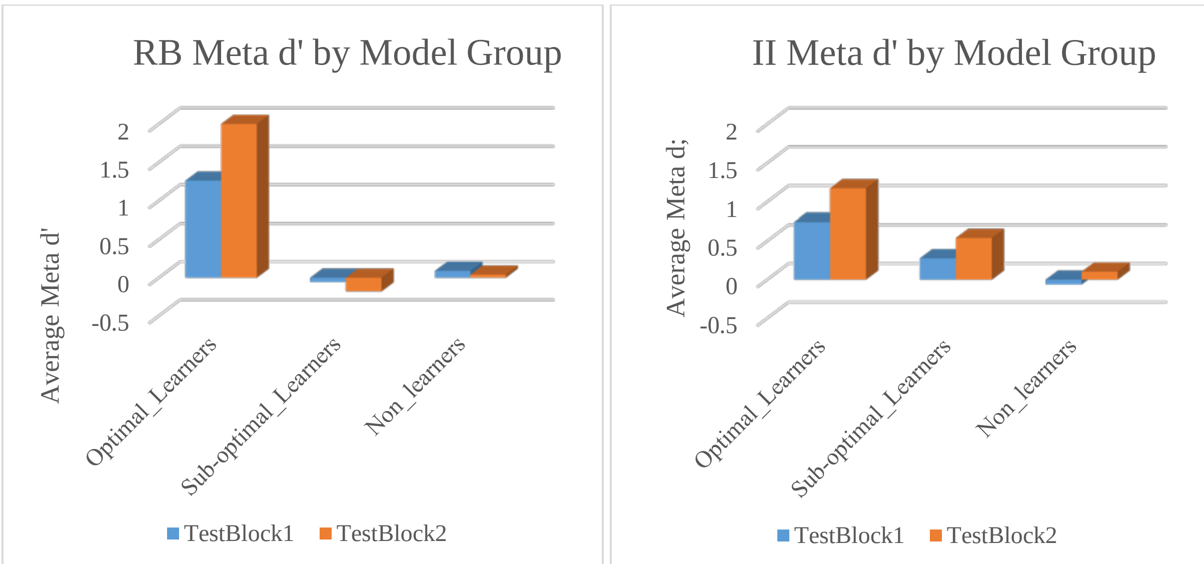


Figure 19 Meta d' by Model Group in both the RB and II category structures.

fNIRS

Based on general linear model analysis described above, ß values estimated during level 1 SPM at the individual participant level were extracted to be used for statistical comparisons at the group level in SPSS. A within subjects repeated measure 2 (structure) x 4 (block) x 20 (nirs channel) ANOVA was conducted on the ß values of optimal and sub-optimal rule users. There was not a main effect of structure, $F(1, 32) = .370$, $p = .934$, *partial* $\eta^2 = .011$. There was not a significant main effect for block, $F(3, 96) = .161$, $p = .922$, *partial* $\eta^2 = .005$. There was not a significant main effect of channel, $F(19, 608) = .585$, $p < .052$, *partial* $\eta^2 = .018$.

There was a significant interaction between structure and block, $F(3, 96) = 3.079$, $p<.05$, *partial* $\eta^2 = .088$. Pairwise comparisons revealed that in the RB structure, there was a significant decrease in activation occurring across blocks. Differences between the RB and II structures were most pronounced during the training blocks, with activation for the II structure being higher than the RB structure when collapsed across the two training blocks (M_{II}-M_{RB}= 3.50E-5).

To assess my hypotheses on the differences in Optimal and Sub-Optimal activation across RB and II structures, I conducted two separate 2 (Optimal vs. Sub-Optimal) x 4 (Block) ANOVAs. Within the RB structure, there was a marginally significant main effect for model group, $F(1, 31) = 3.957$, $p = .056$, *partial* $\eta^2 = .056$ with the activation of Optimal rule users being significantly lower than that of Sub-Optimal rule users. Within the II structure, there was not a significant main effect for model group, $F(1, 31) = .851$, $p = .363$, *partial* $\eta^2 = .027$. Across both structures there was generally higher PFC activation for sub-optimal rule users who are still trying to learn the rule. However, within the II condition, the difference between optimal and sub-optimal activation was much less pronounced, with more channels maintaining high activation across task duration compared to RB (Figure 20). These results support my hypothesis that there would be higher sustained activation throughout the task duration for the II structure due to the implicit learning disruption.

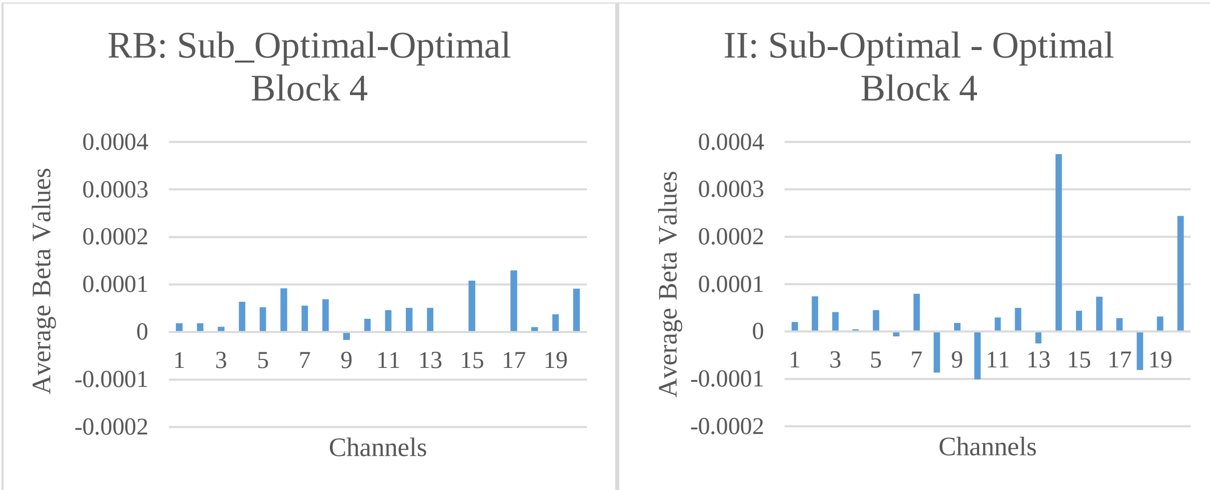


Figure 20 Difference scores for Optimal and Sub-optimal ß values in each category structure

Discussion

The primary goal of experiment 3 was to replicate the implicit learning decrement shown by Maddox and colleagues (2004), along with observing the associated relationship between performance, subjective awareness (Meta d') and neurophysiological data. This was supported, with performance and Meta d' in the RB structures being significantly higher than that of the disrupted II condition. Meta d' proved, as in Experiment 1 and 2, to be a valid predictor of both task performance and decision bound rule use. There were no significant differences in the number of cases of blind insight across category structures. This is in line with my predictions, as I expected to see Meta d' suffer decrements along with performance due to the variable response mapping disruption.

fNIRS comparisons between the ß values of each structure indicate higher PFC activation for II structure compared to RB, particularly within the training blocks with traditional categorization design. This supports my hypothesis of higher sustained activation across blocks due to an implicit learning decrement. Further comparisons on modeling groups indicated that there were significant differences in the RB structure between the ß values of Optimal and Sub-optimal rule users, with Optimal rule users displaying a decrease in activation across blocks. This difference was not significant between Optimal and Sub-Optimal rule users in the II category structure. This supports the hypothesis that the variable response mapping method disrupted implicit learning structures. Overall, these neurophysiological results support the hypothesis that there is higher PFC activation associated with sub-optimal strategy use. This corresponds to the effect shown with Meta d’, with Sub-Optimal rule users having lower Meta d’ than Optimal rule users.

General Discussion

COVIS gains empirical validation through the demonstration of dissociations between explicit and implicit learning. Understanding the conditions in which these two learning systems are different and being able to consistently and reliably replicate these behavioral differences is critical to supporting the separate systems theory of category learning. Despite the evidence in support of the separate systems theory to learning, there was a clear gap in the literature when it comes to subjective experience. There were very few examples of subjective measures of awareness being integrated within a COVIS based categorization learning task. The goal of this series of experiments was to rectify this gap by conceptualizing and applying a measure of metacognitive accuracy to a two-structure categorization framework. I believed that the addition of Meta d’ would allow a greater range of predictions to be made on the differences between explicit and implicit learning.

Combined results from experiment 1, 2 and 3, are supportive of Meta d’ being useful within a COVIS framework. Meta d’ corresponded highly with accuracy on both Rule-Based and

Information Integration categorization structures across all three experiments. In addition to its strong relationship with task accuracy, the Meta d' measure was sensitive enough to show significant changes based on a participant's use of Optimal or Sub-Optimal decision bounds. This data supports the use of a subjective awareness scale as a predictor of overall task performance.

There is also reasonable evidence to suggest that participants possess some awareness of the success/failure of their own strategies, given the relationship between decisions bound models and Meta d'. This was most evident across the II conditions, with there being a greater dissociation between a participant's task performance and their metacognitive accuracy. It is possible that implicit learning does not occur as linearly as explicit learning, with implicit category knowledge influencing confidence before it appears in response accuracy. This was evident through the significant difference in proportion of blind sight within RB and II structures. Blind sight, or the phenomenon of high performance despite low metacognitive accuracy, was seen significantly more within the II condition, compared to RB in experiments 1, 2, and 3. Overall the occurrence of blind sight and blind insight within the II category structures, and not the RB, supports my hypothesis that there would be a greater dissociation between awareness and performance for II structures. RB category structures rely on explicit category rule knowledge and therefore result in a more consistent and positive relationship between awareness and performance.

I was unable to replicate a disruption to the implicit system via removing motor response mapping within experiment 2. One possible explanation for this could be that adding the subjective measure immediately following each trial allowed for explicit processing to occur due to the stimuli being held in working memory during the confidence rating period. Confidence scales are not frequently used within the COVIS framework and therefore the full implications of such a manipulation are not well known. This explanation is supported through the fNIRS data in experiment 3 which showed interruptions to the expected activation patterns during the test blocks with confidence ratings. Predicted differences between RB and II activation were observed within the training blocks only. I believe this is further evidence that the confidence ratings interfered with the categorization task to some degree. Despite this, there were some unique interactions when comparing the II conditions within both experiments 1 and 2 with Meta d' scores, i.e., several of the predicted patterns held in subjective data across structures and model groupings.

Experiment 3 allowed me to examine the relationship between performance, Meta d', and prefrontal cortex activation. There were promising trends that corresponded to my theoretical predictions, particularly within the differences between Optimal and Sub-Optimal rule users that encourages the use of these two measures together within a categorization learning framework.

Conclusions

Across the 3 experiments, I was able to use Meta d' in conjunction with decision bound model fits to make predictions of performance throughout implicit and explicit categorization tasks. These collective results indicate that metacognitive accuracy, an implicit structure, was sensitive to whether a person is using the correct decision bound. This provides useful information about the strategies employed through an implicit task that have previously been described as difficult to verbalize and inaccessible. Meta d' provides previously unavailable information on a participant's awareness and strategy use throughout a task. The fact that Meta d’ displayed variations unique to each category structure in response to various manipulations indicates that it could be a useful measure to integrate into COVIS tasks upon further validation.

The COVIS framework allowed me to evaluate Meta d’ within both explicit and implicit structures. Previous studies on the validity of Meta d’ as a measure of implicit learning did not possess this duality, and therefore the measure had not been fully tested within an explicit learning framework. Despite its usefulness in assessing my predictions, Meta d’ proved problematic in its current operational definition as a measure of implicit learning. Specifically, the ability of Meta d’ to be sensitive to variations within a rule-based task, in addition to the task conditions with implicit disruptions, indicate that Meta d’ can’t exclusively be considered an indication of implicit learning, as it is also apparently sensitive to explicit learning. Future research should focus on exploring the robustness of Meta d’ as a predictor of implicit and explicit learning across various manipulations, such as category discriminability and variations in working memory demands. Relying on behavioral performance data alone would fail to explore the nuances of this type of category learning.

APPENDIX: APPROVAL LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Corey J. Bohil**

Date: **July 30, 2018**

Dear Researcher:

On 07/30/2018 the IRB approved the following modifications to human participant research until 07/29/2019 inclusive:

Type of Review:	IRB Continuing Review Application Form Expedited Review Category #4 & 7
Modification Type:	Audrey Hill's name is changed to Audrey Zlatkin. Andrew Wismer is removed and Clay Killingsworth is added as Sub-PI. Revised Study Application, version 1.4, is attached. Revised wording on the fNIRS. Revised Consent was uploaded and approved for use.
Project Title:	Brain activity during category learning
Investigator:	Corey J. Bohil
IRB Number:	SBE-12-08996
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 07/29/2019, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

This letter is signed by:

Kamille Chap

Signature applied by Kamille Chaparro on 07/30/2018 11:26:46 AM EDT

Designated Reviewer

REFERENCES

Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., & Waldron, E.M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442-481.

Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14* (1), 33-55.

Ashby, F.G., & Maddox, W.T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149-178.

Ashby, F.G., & Maddox, W.T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224 (1)*, 149-178.

Ashby, F.G., Maddox, W.T., & Bohil, C.J. (2002) Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*(5), 666-677.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P., McKendrick, R., & Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and technological development. *Frontiers in Human Neuroscience, 7*, 1-13.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology/Revue canadienne de psychologie, 40*(4), 343-367

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage, 54* (4), 2808–2821.

Dienes, Z., Altmann, G.T.M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21* (5), 1322-1338.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgement knowledge. *Psychological Research, 69,* 338-351.

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage*, *63*(2), 921-935.

Haider, H., Eichler, A., & Lange, T. (2011). An old problem: How can we distinguish between conscious and unconscious knowledge acquired in an implicit learning task. *Consciousness and Cognition, 20*, 658–672.

Maddox, W.T., Bohil, C.J., & Ing, A.D. (2004), Evidence for a procedural-learning based system in category learning. *Psychonomic Bulletin & Review, 11*(5), 945-952.

Worthy, D.A., Markman, A.B., & Maddox, W.T. (2009). What is pressure? Evidence for social pressure as a type of regulatory focus. *Psychonmic Bulletin & Review, 16* (2) 344-349.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cognition, 21* (1), 422-430.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification Learning. *Psychological Review, 85* (3), 207-238.

Pasquali, A., Timmermans, B., & Cleeremans, A. (2010). Know thyself: metacognitive networks and measures of consciousness. *Cognition, 117* (2), 182-190.

Persaud, N., McLeod, P., and Cowey, A. (2007) Post-decision wagering objectively measures awareness. *Nature Neuroscience, 10*(2), 257–261.

Ramsøy, T.Z., and Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3,* 1-23.

Rausch, M., & Zehetleitner, M. (2016). Visibility Is Not Equivalent to Confidence in a Low Contrast Orientation Discrimination Task. *Frontiers in Psychology, 7*, 591.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6* (6), 855-863.

Rips, L.J., Shoben, E.J., and Smith, E.E. (1973). Semantic distance and the verification of semantic distance. *Journal of Verbal Learning and Verbal Behavior, 12* (1), 1-20.

Rosch, E. (1973). Natural Categories. Cognitive Psychology, 4 (3), 228-250

Sandberg, K., Timmermans, B., Overgaard, M., and Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition, 19,* 1069-1078.

Scott, R.B., Dienes, Z., Barret, A.B., Bor, D., and Seth, A.K. (2014). Blind Insight: Metacognitive discrimination despite chance performance. *Psychological Science, 25* (12), 2199-2208.

Smith, E.E. and Medin, D.L. (1981). Categories and Concepts. Cambridge, MA: Harvard University Press

Wismer, A.J., Zlatkin, A.H., & Bohil, C.J. (2016, November). *A subjective measure of explicit and implicit category rule learning*. Poster presented at the 58th annual meeting of the Psychonomic Society, Boston, MA.

Zlatkin, A. H., Patel, P., Wismer, A. J., & Bohil, C.J. (2016, November). *Awareness of explicit and implicit category rule learning*. Poster presented at the 58th annual meeting of the Psychonomic Society, Boston, MA.