5-12-2024

The Effects of Classification versus Observation During Category Learning

Enoch Sarakpo
Syracuse University

Abstract

We report a category-learning experiment that examines the learning outcomes of classification and observational training methods across three category structures. Thus, we crossed training (classification vs. observation) and category type (natural vs. featural vs. relational). Some subjects classified the stimuli (side-by-side bird pairs) and received corrective feedback after each response, whereas others studied these stimuli, wherein they were presented with the corresponding category label. The posttest was an endorsement task made up of repeated and novel items. We did find an observation training advantage, as subjects in the observation training were better in the natural and relational categories for the repeated items, compared to subjects in the classification training. However, this advantage disappeared with feature-based categories, as subjects in both training conditions achieved comparable performance on the endorsement task. For the transfer to novel scenario, there were no differences between the training conditions across all categories. This implies that observation training can be leveraged as a training approach to the existing methods.

THE EFFECTS OF CLASSIFICATION VERSUS OBSERVATION DURING CATEGORY LEARNING

by

Enoch Sarakpo

B.A., University of Ghana, 2011

M.S., Sapienza Università di Roma, 2020

Thesis

Submitted in partial fulfillment of the requirements for the degree of

Master of Science in Psychology.

Syracuse University

May 2024

Table of Contents

1.1 The Testing Effect in Category Learning

1.2 Introduction

Testing has been one of the most used methods to assess learners' ability to transfer their knowledge to a novel context. Understanding the extent to which the benefit of testing extends beyond the classroom is of utmost importance, not only to educators and policy makers but to the public as a whole. With the growing interest in testing, researchers across diverse fields such as cognitive psychology and the learning sciences, have sought answers to investigate the benefits of testing and how testing impacts the transfer of learning. Studies on retrieval practice are not new and can be traced back centuries. Some earlier sources of the phenomenon were reported by Spitzer (1939) and Roediger and Karpicke (2006).

Previous studies have explored the benefits of testing within a category-learning paradigm, wherein subjects either learn by classification or observation (Patterson and Kurtz, 2019; Ashby et al., 2002) with both studies revealing contradictory findings. Testing in this context corresponds to classification, whereas observation corresponds to studying. It is also conceivable that these contradictory outcomes could vary as a function of the type of category structure that is being learned. Understanding how a category structure impacts its learnability and transfer will help learners to strategize as to how to learn different category structures.

In this present work we explored how category structure and training type affects the transfer of learning. Category learning provides a good assessment of the transfer of learning because on each trial, subjects are presented a new exemplar, which requires them to transfer their knowledge about previous exemplars to novel cases.

This work would help provide an insight into how learning occurs across different category structures and how classification versus observation affect the transfer of learning. In the sections that follow, we will review work that examines the relationship between the principles of retrieval practice, classification versus observation training in category learning, and the transfer of learning.

1.3 Testing Effect

One of the most potent methods of memory retention is having learners practice retrieving the to-be learned material. When subjects are asked to retrieve the to-be learned materials rather than restudying it, the retrieval of the to-be-learned material produces better retention of this material. This phenomenon is formally known as the *testing effect* (Roediger, 2013).

The testing effect usually involves a series of stages. Specifically, subjects are first presented with the to-be learned materials to study, after studying, some of the subjects are asked to retrieve the encoded information whilst others are asked to restudy it. The posttest usually involves materials that are identical to the studied material. Results generally show better performance for retrieved materials as compared to restudied materials (Carpenter, 2012).

Retrieval practice has been robust both in and outside the laboratory, and an efficient learning tool (Roediger, 2013). Studies on retrieval practice has been conducted in both the laboratory and in applied environments (Butler, 2010; Butler et al., 2017; Carpenter, 2009, 2011. 2012; Carpenter et al., 2008; Carpenter & Yeung, 2017; Carrier & Pashler, 1992; Corral et al., 2020; Dunlosky et al., 2013; Eglington & Kang, 2018; Kang et al., 2013; Karpicke, 2012; Kornell & Bjork, 2008; Larsen et al., 2013; Lee & Ahn, 2018; McDaniel et al., 2015; Pyc & Rawson, 2009; Roediger, 2013; Roediger & Butler, 2011).

Researchers over the years have placed greater emphasis on stimuli that could easily be learned through memorization. For instance, stimuli used by earlier researchers include word lists (e.g., Wheeler, Ewers, & Buonanno, 2003) and facts (e.g., Carpenter, Pashler Wixted, &

Vul, 2008; Karpicke & Roediger, 2008). For instance, Carpenter et al. (2008) tasked their subjects to study Swahili-English word pairs. Though these earlier studies show the advantages of retrieval practice, the stimuli used in these experiments could be learned by just memorization. Kang et al. (2007) asked a section of their subjects to study some selected articles. Other subjects were asked to answer brief questions while others were given multiple-choice questions on the articles. The results revealed that subjects learn better from testing than restudying.

Though there has been substantial number of investigations on the testing effect and its benefits on memory, less is known about the transfer of learning from testing, especially involving meaningful stimuli that does not require rote memorization.

Learning is a lifelong phenomenon, policy makers expect that whatever learners are taught, they should be able to be transfer to novel scenarios. The essence of transfer is thus vital and as put forward by Schmidt and Bjork (1992), that the end product of learning is to aid transfer to similar or novel situations and scenarios.

1.4 Transfer of Learning

Transfer entails how learners apply the knowledge learnt from one context to another (Alexander & Murphy, 1999). For transfer to take place, learners must be able to apply the learned concept from one context to a novel context or similar context (Corral & Carpenter, 2020). Investigating how learners are able to move from one phase of learning into another through transfer is thus important, as is it reduces the cognitive resources that learners would have to spend making sense of encountered and unencountered scenarios.

A substantial amount of student’s life is spent in school with the hope that once they complete, they would be able apply the knowledge in novel areas. Transfer of knowledge is thus of essence. According to Druckman and Bjork (1994), a substantial amount of our time would have been wasted if knowledge and skills do not transfer. Being able to transfer knowledge learnt into new environments or scenarios is thus vital.

Though there are different forms of training in categorization studies, in the present study, we focus on classification versus observation. In classification (retrieval) training, subjects are presented with the to-be learned material and asked to retrieve, wherein they are given corrective feedback after each response. Classifying the to-be-learned category and then getting corrective feedback afterwards, involves retrieval because learners must retrieve their hypotheses about the categories and prior exemplars.

On the contrary, in observation training, subjects are presented with the to-be-learned materials with the correct answer and are not required to retrieve prior exemplars or hypotheses. Observation thus corresponds to re-study, as subjects must simply study each exemplar and are not required to actively engage in retrieval of prior information.

Corral and Carpenter (2020; also see Corral et al., 2023) described what learners must do in order to aid the transfer of learning to novel scenarios. Corral and Carpenter postulated that for transfer to be successful, subjects must (a) learn and retain the concepts that are being taught, (b) understand those concepts deeply enough to realize that they apply to new scenarios, and (c) translate those concepts into new contexts appropriately based on the specifics of the novel scenario. For example, when a student is taught a physics problem (e.g., acceleration) in class and the student encounters a novel acceleration problem in an exam or a different context, the student should be able to recognize that the novel scenario is an acceleration problem. To transfer their knowledge from previous exemplars to the novel scenarios, the student must recognize that the concepts from previous acceleration problem applies to the novel scenario.

Thus, by recognizing the category of a problem type, it can help learners figure out what solution strategy to use (Corral et al., 2020).

In studying the transfer of learning, most researchers utilize research paradigms that include scenarios that requires subjects to transfer the knowledge learned from the to-be learned scenario to novel scenarios. For example, McDaniel., et al (2007) explored out how retrieval practice aids transfer to novel scenarios in college students. In this study, participants were presented with a quiz in three presentation formats. Some of the participants learned the quiz through multiple choice (MC), whilst others learned through short answers (SA), and the third group studied the quiz by reading only (RO). The posttest was made up of repeated materials and new materials. The results revealed a retrieval practice advantage for the quizzed participants as compared to the participants that were asked to read only. Though there was retrieval advantage, material used in the study were not made up of complex material that entails the transfer of of the learned materials(Leahy et al., 2015; Pan & Rickard, 2018; van Gog & Sweller, 2015), but rather materials that could be learned through rote memorization, thus rendering such a study inconclusive.

Rohrer et al. (2010) investigated the extent to which the amount of transfer impacts the strength of the effect of retrieval practice. Across two experiments, 4th- and 5th-grade pupils were tasked to learn to designate regions or cities to mapped positions. Participants in the experiment either engaged in retrieval (test group) or study (study only). Prior to being asked to either retrieve or study, all participants went through a training session on both retrieval (testing) and studying. The posttest was made up of studied questions and novel questions. The outcome of the study reveals a greater retrieval practice advantage on the transfer items for participants that were tested compared to the participants that restudied.

Though greater transfer was found for participants that were asked to retrieve, the prior exposure to study and testing and the subsequent testing in the retrieval condition might have provided the retrieval group a learning advantage compared to participants that only studied. The alternation between study-test-re-study-test is a confound. Firstly, study-test-re-study-test allows the testing group to alternate between testing and re-studying, whereas the re-study group only gets to re-study. For this reason, the re-study group might become bored whereas the testing group might be more engaged as a result of alternating between tasks (Healy et al.,2017). Secondly, task alternation between testing and re-studying allows the testing group to space their study, whereas the re-study group has to engage in massed study (Carpenter et al., 2005).

1.5 The testing effect and transfer

In a related review, Pan, and Rickard (2018) explored the impact of the retrieval practice effect on transfer. The analysis shows that the effect of transfer tends to disappear when the posttest is made up of material that have not been encountered during the initial encoding phase. Critically, Pan and Rickard noted that theories on the testing effect seem to mostly revolve around explaining memory benefits, but do not explain why retrieval practice would aid transfer.

1.6 Categorization, Testing, Transfer and Cognitive Load

The hallmark of categorization is to be able to transfer knowledge learnt from one environment to novel scenario. The two main types of training basically employed in categorization are classification and observation training. In classification training, subjects are presented with a category without the label, and they are asked to provide the label. Classification requires subjects to retrieve previous exemplars and hypotheses about the category and thus corresponds to engaging in retrieval. Classification training thus entails a greater amount of cognitive load because it requires more mental effort and attention from the learner.

This is so because it involves more complex and abstract cognitive processes, such as hypothesis testing, and feedback processing. Classification training encourage learners to focus on getting correct answers by hypothesizing what determine category membership as opposed to learning about the overall nature of the categories (Levering & Kurtz, 2015).

In contrast, in observational learning, subjects are presented with the correct answer and are not required to retrieve the aforementioned information, and so this corresponds to the restudy version of the testing effect paradigm. Observation learning thus involves less cognitive load, because subjects are not required to commit cognitive resources on hypothesizing about what determines a category membership (Levering & Kurtz, 2015).

Yang and Shanks (2018) studied retrieval practice in the painting styles of various artist. Yang and Shanks asked their subjects to learn the painting styles of artists across four lists. The initial stage of the study was made up of a painting style of an individual artist. Yang and Shanks assigned participants to three conditions. Participants in the study group were given the painting styles of artist to study, followed by math test, and then followed by study again in that order for the four different lists. For the testing group, subjects initially studied the painting styles of the artists, took a math test, and then were given a classification task, in which they classified which artist was the one who drew the painting that subjects were previously presented. Participants in the math condition initially studied the painting styles, were then given and tested on math questions, in that order. The outcome of Yang and Shanks study revealed a testing effect for the testing group.

On the other hand, Lee, and Ahn (2018) failed to show the retrieval practice advantage in a categorization task. In their experiment, participants learn the painting styles of artist divided into two distinct parts (A and B). Participants in the study group were assessed initially or not on the learned painting style of part A prior to continuing on to study the painting styles of part B. Results of their study failed to find a testing effect.

Both Yang and Shanks (2018) and Lee and Ahn (2018) papers suffers from the same methodological limitations, as participants in all conditions initially engaged in an initial study. Thus, subjects in the classification conditions were allowed to switch between study and classification, whereas this was not the case for subjects in the study conditions. Engaging in study and then classification affords participants to switch between tasks, which can offset boredom and reorient attention (known as the *cognitive antidote* to boredom; see Healy et al.,2017). On the other hand, the study group simply re-studies the stimuli again and so they do not get the benefits of task switching. This methodological limitation thus makes drawing firm conclusions from these studies questionable.

Levering & Kurtz (2015) compared classification training to observational training task wherein participants were provided labeled examples and not required to make a classification judgement. The stimuli for the experiment were made up of line drawings of made-up cartoon animals. The stimuli differed along some feature dimensions: beak, antenna, wing, tail, and feet. The results from their study, shows an observational training advantage compared to classification training. The use of made-up cartoon animals may limit generalizability to more complex or naturalistic categories. Research involving an ecologically valid categories would provide a better understanding of how people learn categories in their daily lives.

Across three experiments, Patterson, and Kurtz (2020) investigated the effect learning type in relational category learning. The stimuli for their experiment consisted of three relational categories. The study was made up of comparison learning conditions, wherein two side-by-side stimuli were shown and remained until the trial elapses. The comparison learning groups were matching-category classification and varied-pairs classification. The other condition was the one item per trial conditions, wherein a single stimulus was shown. The conditions for the one item were one-item classification and one-item observational. Participants in Patterson and Kurtz experiments were asked to either learn the relational category by classification, whereby they were provided with feedback or to learn by observation wherein the subjects were provided with the category label to study. Thus, classification and observation were crossed with single versus two-item learning; one-item learning conditions, only a single item was presented on each trial, whereas two-items from the same category were presented to subjects in the two-item trials.

The posttest was made up of an endorsement task whereby participants were shown a given category with a label, and the participants were asked to indicate if it was correct. During the posttest, a single item was presented on each trial, in which subjects determined if it was a member of a given category. Results from the Patterson and Kurtz studies shows an observational learning advantage compared to the classification learning for repeated items whilst for the novel items this advantage disappeared for single items.

Ashby et al. (2002) explored the effects of classification and observation training categorization task. In their experiment, participants were either provided with the name of a category and then later shown an exemplar belonging to the very category. Participants assigned to this condition were the observation condition. Participants assigned to the classification training were shown the category without the label, and thus were tasked to retrieve the category label. After retrieving the category label, the participant was given corrective feedback. The result from Ashby et al. shows that participants who were asked to classify performed better during the posttest compared to participants in the observation condition.

It is worth noting that Ashby et al. (2002) utilized dot patterns and dashes to represent their category structures. Their stimuli consisted of four pairs of category structures. Two pairs belonged to category A, represented by dashes, and the other two pairs belonged to category B, represented by dot patterns. The categories were differed by the location of the means. The dot patterns and dashes varied in length and orientation. In contrast, the stimuli used by Patterson and Kurtz (2020) were made up of Stonehenge like manner arranged rocks, that varied in color, size, and shape, with the category type defined by the relative position of the shapes. Also, the stimuli were presented either as same-category pairs, mixed pairs, or single-item. Subjects in Ashby et al studies were informed about the two categories: “A” and “B”, and that each had an equal and probable chance of occurrence. On the other hand, subjects in Patterson and Kurtz study were not informed about such an equal and probable chance of occurrence. These differences might explain the differences in the results between the two papers.

Sarakpo and Corral (2022) investigated retrieval practice using featural and relational categories. Stimuli for the experiment were Shepard circles. Participants in the classification training were asked to classify the stimuli, whereby they were provided with corrective feedback after each response. For the observational condition, participants were not asked to classify but were shown the stimuli and asked to study it and then find a way to determine its category membership.

The posttest was made up of an endorsement task, and our results revealed a statistically significant interaction. Specifically, participants assigned the classification training, better learned the relational categories than participants assigned to the observational training. On the other hand, there were no differences in subjects ‘mean performance in the featural categories between the observation training and the classification training. However, it is important to mention that our stimuli consisted of two side-by-side objects, and defined by the brightness, radius tilt or size of the dimension. The identification of a stimulus for a participant was contingent on the values of the stimulus on a single dimension, which was counterbalanced between conditions. We did not test for memory (repeated items) in the above experiment. Additionally, performance was generally low in this study, and as such, it is important to replicate this finding by testing for memory and with stimuli that are more learnable, because the results might differ with more learnable stimuli. Table 1 shows studies showing classification and observational training advantages.

Table 1: *Classification and Observational training advantages.*

Study	ST	CTA	OTA
Patterson and Kurtz (2020)	rocks	No	Yes
Levering and Kurtz (2015)	cartoon animals	No	Yes
Sarakpo and Corral (2023)	Shepherd circles	Yes	No
Ashby et al., (2002)	dots and dashes	Yes	No
Lee and Ahn (2018)	Painting Styles	No	Yes
Yang ang Shanks (2018)	Painting Styles	Yes	No
Jacoby et al., (2010)	Birds	Yes	No

Note. ST = Stimulus Type, CTA = Classification Training Advantage, OTA = Observational Training Advantage.

Various theories make different predictions about which type of training should lead to better learning. In the section that follows, we review some predictions that might follow from the extant theories.

1.7 Transfer and Cognitive Load Theories

Sweller (1988, 2011) proposed cognitive load theory, which holds that because working memory has a limited capacity, overburdening it impairs learning. With high cognitive overload subjects are left with little capacity for other aspects of the task at hand. Classification tasks require subjects to retrieve the category label, which might put demands on working memory during training and thus might lead to lower performance than observation training.

Cognitive load theory might predict that for easier to learn categories, learning by classification is better, but for harder to learn categories learning by observation is better, because classification might place a heavier load on the learner than observation. Thus, for cognitive load theory, this interaction is contingent upon how difficult it is to learn the category type.

Stimuli that involve more interconnections require learners to hold more information in working memory. For this reason, more interconnected items might be more difficult to learn. This interaction thus seems to follow for featural, and relational categories based on element interactivity account.

For instance, based on cognitive load theory, Leahy et al. (2015), it might be predicted that the retrieval practice advantage may disappear with to be-learned material that are high in element interconnectivity. The degree of element interconnectedness refers to how well the elements of a task can be well learned without the need to learn the relations involving any other elements (Sweller,1994).

This is so because concepts that are high in interconnectivity strain working memory more than concepts that are low in element interconnectivity. Leahy et al. tested this prediction in primary school students aged 8 to 9 across three different experiments. Primary school students were presented with materials related to bus schedule with high element interconnectivity. The results of their study failed to show evidence of the testing effect and were thus in line with their predictions.

In a related review, Sweller and Van Gog (2015) concluded that the interconnectivity of the to be learned concepts reduces the testing effect. The cognitive demands in classifying highly interconnected elements might therefore lead to the disappearance of the testing effect.

Featural concepts are concepts defined by a given feature or set of features. Being a member in feature-based category is defined by visible characteristics like size, shape, or color (Tomlinson & Bradley, 2010). An example of a feature category is a stimulus defined by size (e.g., all objects are small vs. all objects are large).

On the other hand, relational concepts are concepts defined by the way its elements are bound to specific role-filler bindings by shared relations (Corral & Jones, 2012, 2014). For example, John threw a pen at Mark. This scenario is defined by the specific way in which the throw relation binds John and Mark to specific roles, such that it is John who is doing the throwing and Mark who is at the receiving end.

Relational concepts generally involve a relatively high amount of element interconnectivity. Feature categories thus involve fewer interconnections among its elements than relational categories.

With the fewer element interconnectivity of featural concepts and the high element interconnectivity of relational concepts, we expect an interaction between classification/observation and relational/featural categories. We expect observational training to benefit relational categories more so than classification, whereas for featural categories, we expect classification training to better benefit learning than observational training.

Though these studies used educational materials, the findings by Leahy et al. (2015) and the review by Sweller and Van Gog (2015) suggest that there is an interaction between training and category type.

In conclusion, cognitive load theory predicts that for easier to learn categories, learning by classification is better, but for harder to learn categories learning by observation is better (because classification might place a heavier load on the learner than observation). Because of the high cognitive demand as predicted by cognitive load theory, subjects would be left with little capacity for other aspects of the task at hand. Thus, classification should benefit the learning of feature-based categories, but this advantage should disappear with relational categories.

Current theories of the testing effect predict testing improves memory of the material that is retrieved and thus by having better memory of that information, they might predict a benefit of classification over observational learning, because subjects who engage in classification should have better memory of the to-be-learned concepts and previous exemplars than subjects who engage in observation (Bjork, 1975; Carpenter, 2009; Karpicke &Roediger, 2007).

The motivation of the present thesis is thus to examine which method of training produces superior transfer and whether this effect might depend on the type of category that is being learned. We would like to extend the testing effect to category learning by using stimuli that are ecologically valid and relatively learnable, and to also assess whether the results differ for repeated versus novel items (i.e., whether the effects hold on items that involve memory and transfer). To have a better understanding of this question, we will investigate classification training versus observational training across different category structures and training type, and whether it aids memory for studied items and the transfer to novel items.

To assess memory and transfer, we included repeated and novel items in the posttest. We included that because our results on the studied and the novel items might vary based on memory versus transfer, where we might see a benefit of classification for repeated items, but the effect might be weaker for novel items.

From Bjork (1975; also see Carpenter, 2009) elaborative retrieval hypothesis, we might predict a main effect of training condition, with classification doing better regardless of the category and item type. In contrast, the work on cognitive load (Sweller ,1988, 2011) and element interconnectivity (Leahy et al., 2015) might lead to the prediction that classification will benefit the learning of feature-based categories, but this advantage should disappear or become weaker for relational categories.

1.8 Present research

In this present experiment, we explored whether the principles of retrieval practice can be extended to the learning of different category structures. The three different category structures used in the study were natural categories, feature-based categories, and relational categories. To test these questions, we used natural stimuli. The stimulus for our experiment were made up of bird pairs. The natural category is the natural family/category to which the bird pairs belong to. Bird pairs were from the same family and were both from the Estrildidae family or both were from the Fringillidae family.

For the feature-based bird pair categories, the bird pairs were characterized by the presence of a given feature, wherein both bird pairs were colorful for one category, whereas for another category, both bird pairs were darker. For the relational categories, the categories were characterized by the shared relations binding elements to fill specific roles (Corral & Jones, 2012, 2014), thus for our stimulus the left bird was brighter than the right bird for a given category or vice versa for the other category.

We crossed the category type (natural category vs. featural category vs. relational category) with training type (classification vs. observation). Participants who receive classification training were asked to classify the presented bird pairs, whereby they were provided with corrective feedback after each response. By being asked to classify, participants have the opportunity of retrieving the to-be-learned category hypotheses or exemplars. For the observational training, participants were not required to classify but rather were shown a bird pair with a category label and asked to study it, and then find a way of determining its category membership. We thus investigated (a) which training type (classification vs observation) aids better performance/learning of a category structure (natural category vs. feature-based category vs. relational-based category), (b) which training type (classification vs observation) leads to better performance for old and novel items(transfer), and (c) does the effect of training format vary as a function of category type.

Based on theories of the testing effect (Carpenter, 2009, 2012; Pan & Rickard, 2018), one might predict that there will be a main effect of training where classification training would lead to better performance irrespective of the type of item.

Based on cognitive load theory (Sweller, 1988, 2011) and element interactivity (Leahy et al.,2015) an interaction might be expected between training type and category type, wherein classification should benefit the learning of feature-based categories, but this advantage might disappear on relational categories.

In contrast, based on the findings from Ashby et al. (2002) and Patterson and Kurtz (2019), a different type of interaction might be expected between mode of training and category structure. Specifically, classification training might lead to better learning of the featural categories than observational training, whereas observational training might lead to better learning of the relational categories than classification training.

2.0 Methods

Our aim in the study was to examine if principles of the testing effect could lead to better performance in memory and also in the transfer of learning to novel items in a natural category learning task. The stimuli for the experiment were made up of bird pairs. The bird pairs were used to form three different category structures: (a) natural, (b) feature-based, and (c) relational categories.

2.1 Experiment

2.2 *Subjects*

Participants were 293 Introductory Psychology students at Syracuse University, who served as participants in exchange for course credit. The participants were given partial course credit for taking part in the study.

2.3 Design and Materials

A 3 × 2 × 2 mixed design was used for the experiment. Thus, category type (natural vs. feature vs. relational categories) was crossed with training type (classification vs. observation) and test type (repeated vs. novel). The category type and training type were between-subject factors whilst the test type was a within-subjects factor. All images contained a single bird. We used a total of 120 images for this experiment: 60 images were from the Estrildid family whilst the remaining 60 were from the Fringillidae family. Each stimulus consisted of a bird pair. We used bird pairs to denote a structured relationship between the two birds in relational category condition, and we wanted to keep the stimuli that subjects were presented consistent across all category types.

From the order *Passeriformes*, we took birds from the Estrildid (tiny seed-eating birds of the passerine family found in the Old-World tropics and Australasia). The other family was made up of the Fringillidae (true finches of small to medium-sized passerine birds). For the natural categories, Category A consisted of bird pairs from the Fringillidae family, whereas category B was made up of the Estrildid family. Bird pairs for a given category were thus always from the same bird family.

For the feature-based condition, Category A consisted of colorful bird pairs, whereas for Category B it consisted of darker bird pairs. The feature categories were partitioned such that 24 images of each family were ambiguous, 18 were clearly colorful and 18 were clearly darker (less colorful images). These images were closer to their corresponding category rule (i.e., for category A these were more colorful than dark, and for category B these were darker than colorful).

For the relational conditions, the stimuli in Category A had colorful pictures on the left side and darker pictures on the right side for a bird pair, whereas for Category B, it had darker pictures on the left side and colorful pictures on the right side for a bird pair. For each subject, the bird pairs that were selected were randomized, subject to the constraint that they cohered to the corresponding category rule.

Stimuli for natural categories were from two bird families (Estrildidae and Fringillidae). Categories A consisted of bird pairs from Estrildidae family whilst for category B, the bird pairs were from the Fringillidae family. For the natural category, this meant that the two bird families were never paired together, but the feature and relational categories involved pairs that were and were not from the same family. Also, the featural and relational categories were characterized by how colorful and less colorful the bird pair were, but that was the same for the natural categories.

Images of bird pairs used in this study were taken from www.whatbird.com and Google Images. In line with Jacoby et al. (2010), the birds were taken from a single taxonomic order called *Passeriformes* (perching birds). For access to the stimuli, please visit https://osf.io/uvtr7/

2.4 Procedure

Participants were provided a cover story at the beginning of the experiment. Participants were provided a cover story in which they were being asked to join a task force to help the government decode bird pair patterns that were being used by an alien species to communicate with one another.

One hundred birds were used for the training phase (i.e., 50 bird pairs), and 20 bird pairs were included in the posttest.

All stimuli were presented on a standard 24-inch LCD computer monitor on a black background and all responses were entered using a computer keyboard.

The training phase was made up two blocks with 50 bird pairs in each block. The two training blocks included the same bird pairs. We included two blocks to give subjects a chance to better learn the categories, thereby avoiding floor effects. This was necessary after running a pilot study where we realised that a single training block was not sufficient for the learning of the relational and natural categories. There was a rest break after every 25 trials, which was self-paced.

The sequence in which the bird pairs appeared for each participant was randomized. For each participant, the category label that was used for each category was randomized. On each trial, the bird image that appeared on the left side and right side of the screen was randomized. For each subject, the bird pairs that were selected was randomized, subject to the constraint that they cohered to the corresponding category rule. For the natural category, this meant that the two bird families were never paired together, but the feature and relational categories involved pairs that were and were not from the same family.

For the relational category conditions, this was constrained by the category rule, such that the image that was shown on the left and the right sides of the screen needed to cohere to the category rule of being more/less colorful. By doing this, it would allow for a natural randomization of which birds would be presented on the right and left sides of the screen for each subject in the relational category conditions.

Participants assigned to the classification training were to classify the bird pairs that were presented one pair at a time. After making a classification judgment, participants were provided with correct-answer feedback, whereby they were presented the corresponding category label and were asked to study the bird pair carefully and think about how they are related and their similarities and differences. Also, the correct answer appeared in green during the learning phase and was presented directly beneath the stimulus.

For participants in the observation condition, the bird pairs were shown with one pair at a time on the center of the screen, with the corresponding category label beneath it. Subjects were presented the corresponding category label and were asked to study the bird pair carefully and think about how they are related and their similarities and differences. Thus, with the exception of the classification condition making a classification judgment before seeing the correct category label, the classification and observational conditions were identical.

Once subjects in both the classification and observation conditions were shown the correct category label for the stimulus, subjects in both conditions were required to enter the key press that corresponded to the category label in order to move on to the next trial. The screen was cleared after each response. All participants went through the training phase, and then completed the posttest. The two category labels that we used were Alkins and Bafsters, and the category that they corresponded to (i.e., Category A and Category B) was counterbalanced across conditions.

The intertrial interval was 300ms. Once the category label on each trial was presented, there was a waiting time of 5s.

The posttest was made up of an endorsement task and consisted of repeated and novel items, in which participants determined if the label of a bird pair was correct or wrong. We employed an endorsement task instead of a classification task on the posttest in order to reduce the potential effects of transfer appropriate processing (Morris et al.,1977), since subjects in the classification condition engaged in classification during training.

In all, 20 bird pairs were used for the endorsement task with 10 bird pairs sampled from the training (repeated) set and the remaining 10 pairs constituted novel bird pairs. The 10 bird pairs that made up the studied items were randomly selected from the bird pairs used during the training and thus, were encountered by the participants beforehand, with the novel bird pairs only encountered during the posttest.

In the endorsement task, the bird pairs were randomly selected (subject to the constraint of being novel and repeated bird pairs), with the order of presentation randomized for each participant. For both repeated and novel bird pairs, half of the pairs were from Category A with the other from Category B. For each bird pair, a category label was displayed beneath the stimulus and the subjects were asked to “Press "Y" if this label is correct or press "N" if it is incorrect”. No feedback was presented. Figure 1 illustrates this general procedure used throughout.

A.


Press "A" if these bird pairs are Alkins'
'Press "B" if these bird pairs are Bafsters

B.


Please study these bird pairs carefully and try to figure out how to identify Bafsters bird pair'
Try to think about how they are related and about their similarities and differences
Bafsters
Press B when you are ready to move on

C.

D.

Figure 1. An example of a/an (a) classification training for natural category A (b)observation training for feature category A (c) observation training for feature category B, and (d) endorsement task for the posttest.

3.0 Results

3.1 Primary Results

We examined subjects' performance on the posttest using a 3 (Category type: natural vs. featural vs. relational) × 2 (Training: classification vs. observational) × 2 (posttest: repeated vs. novel) mixed analysis of variance (ANOVA), with category and training type as between-subject factors and posttest as a within-subject factor. This mixed ANOVA revealed a main effect of training type, F (1,291) = 3.87, p = .050, MSE = 0. 091. That is, participants in the observational training performed better in the posttest compared to participants in the classification training condition. There was no interaction between training type and category type, F (2, 287) = 2.44, p =.089, MSE = .058. Table 2 illustrates participants mean posttest performance for each category type across training types.

Table 2

Mean performance on the posttest for repeated and novel items.

	NC	NO	FC	FO	RC	RO
Repeated Items	0.569 (0.028)	0.657 (0.023)	0.856 (0.028)	0.823 (0.029)	0.539 (0.029)	0.622 (0.023)
Novel Items	0.502 (0.022)	0.546 (0.022)	0.815 (0.029)	0.806 (0.031)	0.487 (0.027)	0.525 (0.026)
All Items	0.535 (0.018)	0.602 (0.017)	0.835 (0.026)	0.815 (0.028)	0.513 (0.023)	0.574 (0.019)

Note. Standard errors of the mean are shown in parentheses. NC = Natural Classification, NO = Natural observation, FC = Featural Classification, FO = Featural Observation, RC = Relational Classification, and RO = Relational Observation.

3.2 Results for Repeated Items

Figure 2 illustrates participants mean performance on the posttest for repeated items for the category types by training types. The mean performance on the posttest for the repeated items for the observational training was better for the natural category (M = 0.657, SE = .023) as compared to the classification training (*M* = .569, *SE* = .028), *p* =.029. The mean performance on the posttest on repeated items for observational training for the relational categories was better (*M* = .622, *SE* =.023) than the classification training for the repeated items for the relational category (*M* = 0.539, *SE* = .029), *p* =.029. This demonstrate that observational training led to better posttest performance than classification training for natural and relational repeated categories. The mean performance on the posttest for the repeated items for the classification training on the featural categories (*M* = .856, *SE* = .028) was not different from the observational training condition (*M* = .823, *SE* = .029), *p* = .414.

Taken together, these results suggest that there is an observational training advantage for the repeated items for the natural and relational categories as compared to the classification training. We did not observe this observational training advantage in the featural categories.

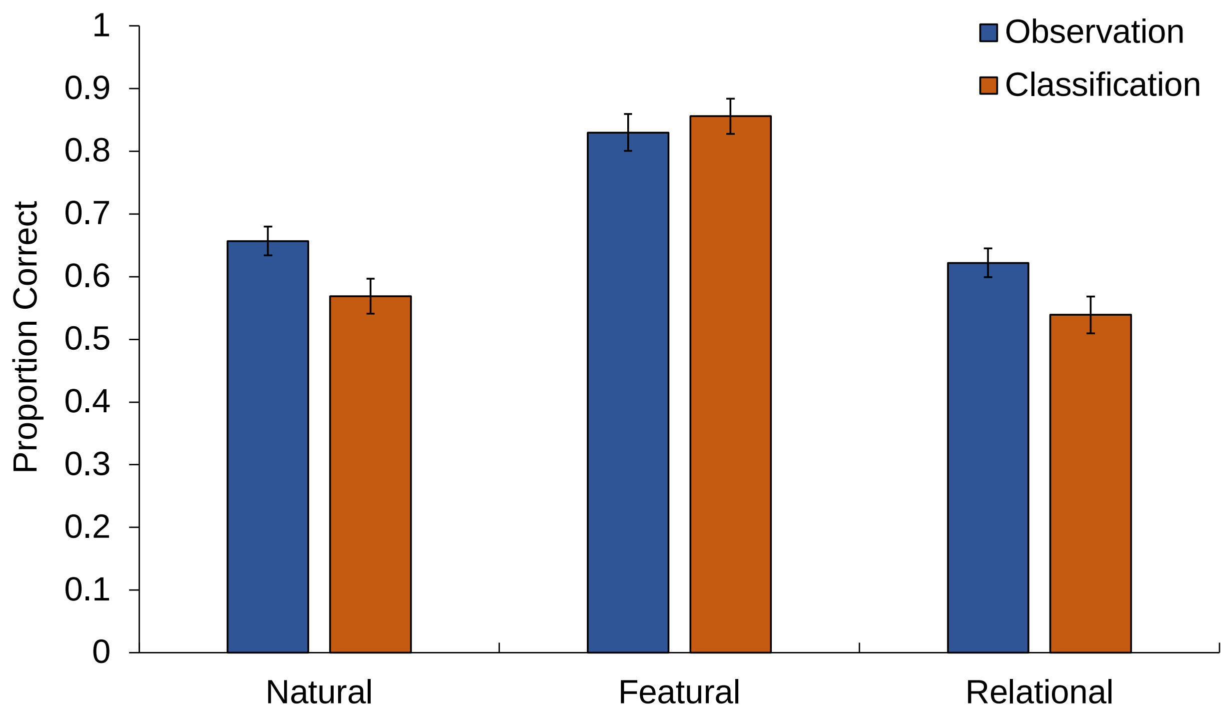

Figure 2. Mean performance and standard errors of the mean on the posttest for each category type by training type on the repeated items.

3.3 Results for Transfer Items

Figure 3 shows mean performance on the posttest for novel items by training type. For the novel items in the natural categories, there were no differences in mean performance for subjects in the observational/natural category training (*M* = 0.546, *SE* = .022) and subjects who received classification/natural category training (*M* = 0.502, *SE* = .022), p =.258. For the novel items in the featural categories, there were no differences in mean performance between subjects in the observational/featural category training (*M* = 0.815, *SE* = .029) and subjects in the classification/featural category training (*M* = 0.806, *SE* = .031), p = .258. For the novel items in the relational categories, there were no differences in mean performance between subjects in the observational/relational category training (*M* = 0.525, *SE* = .026) and subjects who received classification/relational category training (*M* = 0.487, *SE* =.027), p>.258. The subjects mean performance in the novel items follows the same trend as in the repeated items but did not reach statistical significance.

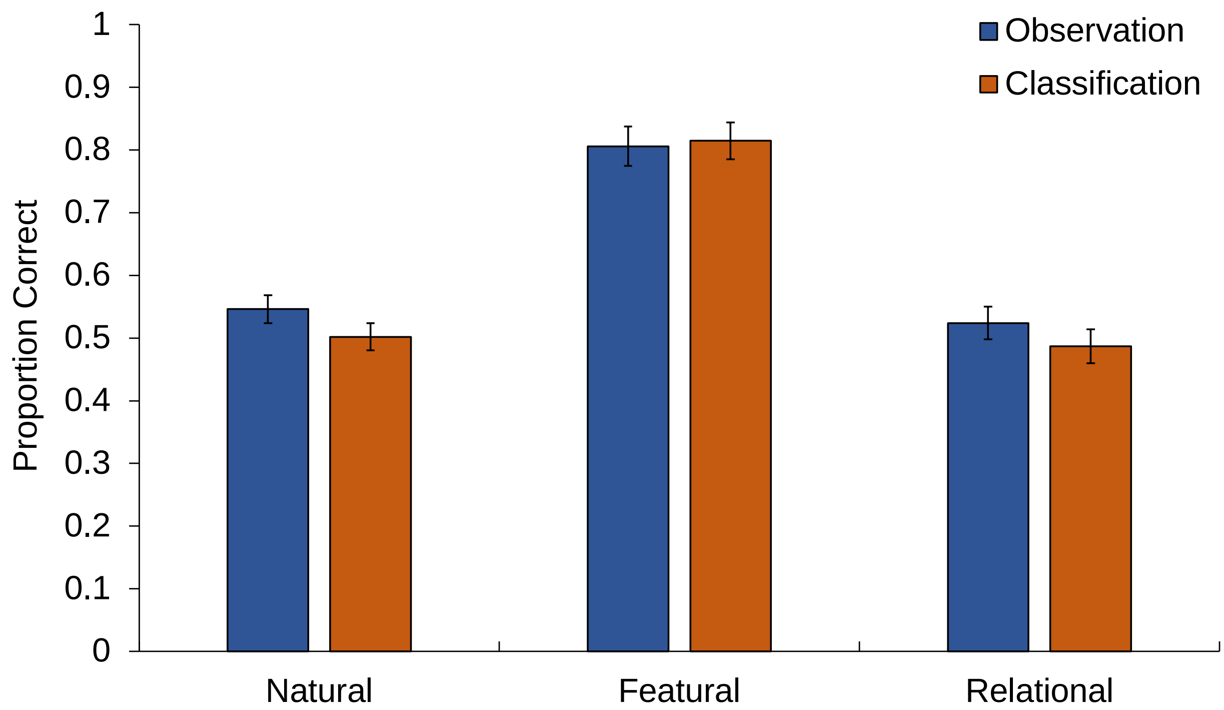

Figure 3. Mean performance and standard errors of the mean on the posttest for each category type by training type on the novel items.

3.4 Planned Analyses

Our analysis plan stated that we would conduct 3 × 2 × 2 mixed ANOVA (category type: natural vs featural vs relational × training: classification vs observation × posttest: novel vs repeated). We followed up with this approach. For the repeated items, there was a main effect of training condition, F (1,292) = 4.43, p = .036, MSE = 0.156. That is, participants in the observation training condition performed better in the posttest for repeated items compared to subjects in the classification training condition. For the repeated items in the posttest, there was an interaction between training mode and category type, F (2,292) = 3.27, p = .040, MSE = 0.115. Observational training led to better posttest performance than classification training for natural and relational repeated categories (both *ps* < .029). There were no differences in posttest performance between observation and classification for featural categories (p = .414). For the novel items, there was no main effect of training mode, F (2,287) = 1.29, p = .258, *MSE* = .044, nor was there an interaction between training type and category type, F (2,287) = 0.59, p = .558, *MSE* = 0.02. All analyses for the experiment can be accessed at https://osf.io/uvtr7/.

4.0 Discussion

In this study, we used classification and observational training methods to test if retrieval practice could improve memory for repeated items and the transfer of learning to novel scenarios in a category learning task. We crossed training type (classification vs. observation) and category type (natural vs. featural vs. relational). Some subjects learned these categories by classifying the stimuli (side-by-side bird pairs) and receiving corrective feedback, whereas others learned them through observation, wherein the stimuli were presented with the corresponding category label.

The type of training was statistically significant and there was no interaction between training and category type. For the repeated items, the results show a main effect of training type and an interaction between training type and category type. The results show an observational training advantage in the repeated items for the natural categories and in the relational categories, but this effect seems to disappear for the featural categories and is not present at all on the novel items for any of the categories. Thus, there is a memorial benefit of learning of the to-be-learned categories, particularly for subjects in the observational training condition in the natural and relational categories, but this benefit was weaker for the transfer to the novel materials.

This result is in line with prior studies that indicated that learners in the observational training conditions performed better than learners in the classification training condition (Levering & Kurtz, 2015), and for repeated items (Patterson & Kurtz 2020). We found a similar pattern, whereby subjects in the observational training condition performed better than subjects in the classification training condition for the repeated items in the relational and natural categories, with the benefit going away in the feature-based categories.

The outcome of our study could be explained with the cognitive load theory (Sweller 1988, 2011), particularly for the relational categories, and in part (for the natural categories) where participants demonstrated an observational training advantage over classification training, especially when assessed on repeated items. This is so because the act of retrieving the to-be-learned categories (classification training) involves more mental effort and exhausts working memory resources than restudying (observational training). Classification requires subjects to retrieve previous exemplars and hypotheses about the category which entails retrieving information from long-term memory, a very laborious and demanding process. On the contrary, observational training subjects are presented with the correct category answer and are not required to retrieve or hypothesize about the category, and thus entails an inactive encounter with the to-be-learned category, a less laborious task. Retrieving and hypothesizing might have led to depletion in working memory (Chen et al, 2018; Schmeichel, 2007) likely leading to the lower mean performance in the classification training compared to subjects given observational training.

Our results are also in line with the predictions of Leahy et al. (2015), and Sweller and Van Gog (2015), that the testing effect may disappear with to be-leaned material that are high in element interconnectivity. It seems the relational categories and the natural categories were highly interconnected leading to the observational training advantage in the repeated items compared to the classification training.

The results of this study are in conflict with studies showing advantages of classification training over observational training (see Ashby et al. 2002; Sarakpo & Corral, 2021). One plausible explanation for these discrepancies lies in the stimuli used in these studies. Specifically, Ashby et al. used stimuli made up of dots and dashes, varying in length and orientation. In the study by Sarakpo and Corral, they utilized two-side-by-side objects defined by values across three dimensions: brightness, size, and radius tilt.

For our present work, we used bird pairs to denote three category structures: For the natural categories, category A comprised of bird pairs from the Fringillidae family, whereas category B consisted of bird pairs from the Estrildid family. For the feature-based condition, category A, were made up of colorful bird pairs, whereas category B included darker bird pairs. In the relational categories, we manipulated the arrangement of the bird pairs. For category A, colorful pictures appeared on the left side and darker pictures on the right side for each bird pair. Conversely, in Category B, the arrangement was reversed.

These variations in stimuli likely contributed to the observed differences across these studies. Also, how the researchers structured the categories across the different studies might have led to participants, requiring distinct learning strategies. We also tested for memory in the present study whereas Ashby et al, and Sarakpo and Corral did not.

For the transfer to novel items, there were no differences in subjects' performance given the training type. In the retrieval practice literature, not much work is done on tasks involving transfer of learning to novel scenarios (Carpenter et al., 2020). Current theories on retrieval practice investigates how retrieval strengthens memory of the to-be-learned materials (Carpenter et al., 2022; Pan & Rickard, 2018). On the other hand, these extant theories do not clearly state how this strengthening in memory might aid in the transfer of learning to novel scenarios.

This study shows that to transfer to novel scenarios in a category learning task, memory of the to-be-learned materials is not sufficient to translate to the novel scenario, but rather subjects might need to be able to recognize what makes a category type distinct from the other and then be able to transfer that learned information to novel situations.

The few studies on category learning that have examined classification versus observational training have not examined whether their effects vary as a function of the category type. We build on this work by investigating whether the effects these different types of training vary as a function of the type of category that subjects are learning.

Even though we did not find a reliable interaction between training type and category, we did find there is some evidence that the outcome varies based on the type of training and the categories (when we collapsed across natural and relational categories) for the repeated items. These category types are usually not the focus of researchers when looking at retrieval practice and transfer.

Beyond the theoretical implications, the result of this study has implication in cognitive psychology and the learning sciences. With scientists looking for better ways to teach students, this study has shown that observational training method can be leveraged as a powerful training approach to add to the existing methods of teaching. Thus, giving learners correct answers during learning helps improve their memory and performance when assessed on repeated items as compared to when learners are asked to actively discover the answers themselves (classification). Also, in the learning of natural and relational categories, this study demonstrate that observation training may be a more effective and efficient way to teach students. For instance, when teaching students natural categories like birds, students might benefit more from the teaching by being given examples from these categories to observe the similarities and differences among them.

4.1 Limitations and Future Directions

The present study represents a first attempt to investigate if the principles of retrieval practice could be extended to category learning, and whether the benefits of classification and observation training types vary as a function of the type of category that subjects are learning. Although our results favored observational training in the repeated items for relational and natural categories, it is worth noting some limitations.

The relational categories were difficult for subjects to learn. The relational categories were characterized by the shared relations binding elements to fill specific roles (Corral & Jones, 2012, 2014), thus for our stimulus the left bird was brighter than the right bird for a given category or vice versa for the other category. Being able to determine a category membership is contingent on subjects understanding the shared relations binding elements to fill specific roles. Because of the shared relations, relational categories are usually more abstract, and complex compared to featural categories, requiring higher-order cognitive processes such as analogical reasoning (Gentner & Markman 1997) and comparison (Gentner & Markman 1997; Hammer et al, 2008). The process of abstracting and comparing the bird pairs might be challenging for learners in the relational categories, leading to the subjects' lower performance in the relational categories.

Also, with the shared relations, surface similarity of relational categories might also influence how subjects form wrong associations leading to the difficulty in subjects learning the relational categories. For instance, subjects might fail to notice that in the relational category, the stimuli in Category A had colorful birds on the left side and darker birds on the right side for a bird pair, whereas for Category B, it had darker birds on the left side and colorful birds on the right side for a bird pair, as subjects might be focusing on some seemingly obvious relations such as the beak or tail length of the birds: they might be focusing on elements of the stimuli that are not part of the relational category. Future studies could address this issue by using relations that are more salient to learners.

For instance, if the relational stimuli were very easy to learn, participants in our study might have been able to learn at comparable rate just like the featural categories in the repeated items, and in the transfer to novel items. For example, if the participants were to learn that the relations of the bird pairs were based on their colorfulness (e.g., left bird is colorful than right bird), they might have been able to apply this rule to the other category (e.g., right bird is colorful than left bird). Being able to learn such a rule, would have freed the participants from focusing on irrelevant features such as the beak or tail length of the bird pairs. Participants might have also been able to transfer the rule to the novel scenario, just as they did with the featural categories.

We manipulated the different category structures, and this manipulation might have imposed some difficulty and impacted our results.

For example, for the natural categories, category A consisted of bird pairs from the Fringillidae family, whereas category B was made up of the Estrildid family.

For the feature-based categories, the categories were defined by how colorful or darker the bird pairs are. These were so distinct and so apparent that participants were able to learn the categories so easy.

With the relational categories, the stimuli in category A had colorful birds on the left side and darker birds on the right side for a bird pair, and vice versa for category B. This might have imposed some difficulty in the participants learning.

The category manipulation was such that for each subject, the bird pairs that were selected was randomized, subject to the constraint that they cohered to the corresponding category rule. For the natural category, this meant that the two bird families were never paired together, but the feature and relational categories involved pairs that were and were not from the same family. This variation in category structures might introduce noise or variability in the data and might have impacted our results. In our future studies, we would want to manipulate the difficulty of categories to see how that would affect memory and the transfers of learning to novel scenarios.

The stimuli for our experiment were natural stimuli and as such were difficult to control. It was thus difficult to manipulate the natural category membership, and thus we could not reduce any form of noise that membership might impose in the natural categories. Our inability to control what determined a natural category might have impacted how subjects learn the natural categories and impacted our results. Also, creating relational and featural categories from natural categories might have also imposed some difficulties for participants to learn the categories. In terms of future work, we could replicate this study by utilizing other stimuli such as artificial rocks, as we could manipulate the shape and color.

It is worth noting that there was no delay testing in this study. Assessing retention after a delay, produces better performance for subjects who retrieved, compared to subjects that restudied. This is so because the delay might lead to a decrease in the retrieval strength of the restudied items compared to the retrieval strength of the tested items (Roediger &Karpicke, 2006; Roediger & Nestojko 2015). When assessed immediately after the training, subjects in the classification condition might not perform well as compared to subjects in the observational training condition as a result of a depletion in their working memory given the classification training (Chen et al, 2018; Schmeichel,2007).

Retrieval practice is not the same as classification, we are extending retrieval practice to category learning, and it is not clear to what extent this could be. It might be that retrieval practice and classification work in tandem, with retrieval practice playing a role in enhancing the long-term retention of the to-be-learned materials, and organization of knowledge (Roediger et al, 2011). Classification, on the other hand, helps us to organize the world and make sense of it (Hammer et al., 2008).

4.2 Conclusion

In this study, we investigated whether the principles of retrieval practice could be extended to category learning task. Some subjects classified the stimuli wherein they were provided corrective feedback after each response, whereas others studied these stimuli, with the corresponding category label. The posttest was made up on an endorsement task. In the endorsement task, a category label was displayed beneath the stimulus and the subjects were asked to “Press "Y" if this label is correct or press "N" if it is incorrect”.

We found a main effect of training where subjects with observational training performed better compared to subjects in the classification training. We found no support that classification training improves the subject’s performance in repeated categories, and also in the transfer to novel scenarios (relative to observation training).

5.0 References

Alexander, P. A., & Murphy, P. K. (1999). Nurturing the seeds of transfer: A domain-specific perspective. *International journal of educational research*, *31*(7), 561-576.

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.

Barenberg, J., Berse, T., Reimann, L., & Dutke, S. (2021). Testing and transfer: Retrieval practice effects across test formats in English vocabulary learning in school. *Applied Cognitive Psychology*, *35*(3), 700-710.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118–1133. http://dx.doi .org/10.1037/a0019902

Butler, A. C., Black-Maier, A. C., Raley, N. D., & Marsh, E. J. (2017). Retrieving and applying knowledge to different examples promotes transfer of learning. *Journal of Experimental Psychology: Applied*, *23,* 433–446. http://dx.doi.org/10.1037/xap0000142

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(5), 619-636.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563–1569. doi:10.1037/ a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributed to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552. doi:10.1037/ a0024140

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21*, 279–283.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36,* 438–448. doi:10.3758/MC.36.2.438

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory & Language, 92*, 128-141. https://dx.doi.org/10.1016/j.jml.2016.06.008.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & cognition, 20*, 633-642.

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: evidence from the spacing effect. *Educational Psychology Review, 30*, 483-501.

Corral, D., & Carpenter, S. K. (2020). Facilitating transfer through incorrect examples and explanatory feedback. *Quarterly Journal of Experimental Psychology, 73*, 1340-1359.

Corral, D., Carpenter, S. K., Perkins, K. M., & Gentile, D. A. (2020). Assessing students' use of optional online reviews. *Applied Cognitive Psychology*, *34*, 318-329.

Corral, D., Carpenter, S. K., & St. Hilaire, K. J. (2023). Testing versus studying during analogical problem solving. *Psychonomic Bulletin & Review*.

Corral, D. & Jones, M. (2012). Learning of relational categories as a function of higher-order structure. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th Annual*

Meeting of the Cognitive Science Society (pp. 1434-1439). Austin, TX: Cognitive Science Society.

Corral, D., & Jones, M. (2014). The effects of relational structure on analogical learning. *Cognition*, *132*(3), 280–300. https://doi.org/10.1016/j.cognition.2014.04.007

Corral, D., Quilici, J. L., & Rutchick, A. M. (2020). The effects of early schema acquisition on mathematical problem solving. *Psychological Research*, *6*, 1495–1506.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students’ learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4–58.

Eglington, L. G., & Kang, S. H. (2018). Retrieval practice benefits deductive inference. *Educational Psychology Review*, *30*, 215-228.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, *83*(1), 37.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American psychologist*, *52*(1), 45.

Google Images (2023). *Passerine images*. Retrieved from https://en.wikipedia.org/wiki/Passerine.

Karpicke, J. D. (2012). Retrieval-based learning: active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*, 157–163.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science, 19,* 585–592.

Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison processes in category learning: From theory to behavior. *Brain Research*, *1225*, 102-118.

Hoffman, A. B., & Murphy, G. L. (2006). Category dimensionality and feature knowledge: When more features are learned as easily as fewer. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(2), 301.

Hsu, A. S., & Griffiths, T. E. (2010, February). Effects of generative and discriminative learning on use of category variability. In *32nd annual conference of the cognitive science society*.

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1441.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European journal of cognitive psychology*, *19*(4-5), 528-558.

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical education, 47*(7), 674-682.

Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, *27*, 291-304.

Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology*, *110*(2), 203.

Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & cognition*, *43*(2), 266-282.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological bulletin*, *129*(4), 592.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, *13*(4), 329-358.

McDaniel, M. A., Bugg, J. M., Liu, Y., & Brick, J. (2015). When does the test-study-test sequence optimize learning and retention? *Journal of Experimental Psychology: Applied, 21*, 370–382. http://dx.doi.org/10 .1037/xap0000063

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European journal of cognitive psychology*, *19*(4-5), 494-513.

Morris, M. W., & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition*, *18*(4), 407-418.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, *16*(5), 519-533.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological bulletin*, *144*(7), 710.

Patterson, J. D., & Kurtz, K. J. (2020). Comparison-based learning of relational categories (you'll never guess). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 851.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*, 437– 447.

Roediger, H. L. III. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest, 14*, 1-3.

Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of learning and motivation*, *55*, 1-36.

Roediger, H. L., & Nestojko, J. F. (2015). The relative benefits of studying and testing on long-term retention. *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin*, 99-111.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 233.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382-439.Rowland, C. A. (2014). The effect of testing versus re-study on retention: a meta-analytic review of the testing effect. *Psychological bulletin*, *140*(6), 1432.

Sarakpo, E., & Corral, D. *Extending the testing effect to category learning: Observational versus classification training with feature- and relation-based*. Manuscript in preparation.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*(9), 641.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.

Van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*, 247-264.

*Whatbird (*2023). Retrieved from https://whatbird.com/.

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571-580.

Whitehead, P. S., Zamary, A., & Marsh, E. J. (2022). Transfer of category learning to impoverished contexts. *Psychonomic Bulletin & Review*, *29*(3), 1035-1044.

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 485.

ENOCH SARAKPO

enoch8hg@yahoo.com Skype: enoche.kofi

747 Sumner Avenue, Unit 1, Syracuse, NY 13210

EDUCATION

Syracuse University, Syracuse NY May 2024

Master of Science in Psychology

University of Rome, Rome, Italy March 2020

Masters in Cognitive Neuroscience

University of Ghana, Accra, Ghana May 2011

Bachelor of Arts Psychology

AWARDS AND HONOURS

- Graduate Teaching Assistantship, Syracuse University 2021-2024
- Summer Teaching Funding 2023
- Laziodisu Scholarship 2017- 2019
- Winner of the Travel Bursary of the International League 2019

Against Epilepsy (ILAE) McGill University, Montreal, Canada

- University of Rome Student Collaboration Scholarship 2018

RESEARCH EXPERIENCE

RESEARCH ASSISTANT: Corral Lab 2021 – to 2024

Syracuse University

Syracuse, USA

Supervisor: Dr. Daniel Corral

Responsibilities

- Designed experiments on categorization and the testing effect.
- Wrote MATLAB script and used it to analyse the categorization and the testing effect data.

RESEARCH ASSISTANT: Prof. Cornwell Lab 2021 – to 2023

Syracuse University

Syracuse, USA

Supervisor: Prof. Catherine Cornwell

Responsibilities

- Designed experiments on Odor Preference Test in Mice.

- Helped mentor summer pride students.

RESEARCH ASSISTANT: Motor Control and Cognition 2019 – 2020

Laboratory, University of Rome 'La Sapienza'

Rome, Italy

Supervisor: Prof. Stefano Ferraina

Responsibilities

- Designed a simple reaction time experiment using MATLAB (NIMH MonkeyLogic) under the Stop
  Signal Paradigm.
- Performed behavioural testing of participants in a Selective Stop Task.
- Wrote a MATLAB script and used it to analyse the behavioural data under the Selective Stop Signal Task.

RESEARCH ASSISTANT: Sport Psychology 2018

Laboratory (LPdS), University of Rome 'La Sapienza'

Rome, Italy

Supervisors: Prof. Fabio Lucidi and Dr Adrea Chirico

Responsibilities

- Collected field data on shooters for 'Quiet Eye' data analysis.
- Edited videos on 'Quiet Eye' using video Pad Video Editor.
- Searched for and summarized relevant bibliography on 'Quiet Eye'.
  (Basketball free throw).

POSTER PRESENTATIONS

- Enoch Sarakpo & Daniel Corral. (2023, November). The Effects of Classification versus Observation During Category Learning. Presented at the 64th Annual Psychonomic Society Conference, San Francisco, California, USA.
- Enoch Sarakpo. (2023, June). Selective stopping task in normal aging and Alzheimer's Disease. "Risk and Resilience to Alzheimer's Disease in African Americans"
- Enoch Sumakpoyaa & Daniel Corral. (2022, November). Partitioning Problem Solving into Its Component Parts and Its Effect on Learning. Presented at the 63rd Annual Psychonomic Society Conference, Boston, Massachusetts, USA.
- Daniel Corral & Enoch Sumakpoyaa. (2022, July). Testing the Testing Effect with Featural and Relational Categories presented at the 44th Annual Cognitive Science Conference, Toronto, Canada.

TEACHING EXPERIENCE

Teaching /Teaching Assistant, Syracuse University 2021-to Present

Syracuse, New York.

Responsibilities

- Taught and examined students in Cognitive Psychology (PSY 322)
- Held recitations in Foundations of Human Behaviour (PSY 205)
- Support in-class assignments and activities in Psychology of Childhood (PSY 335)
- Grading students' assignments and managing Blackboard page.

Teacher Seventh Day Adventist Junior High 2013- 2017

School, Chiraa Ghana.

Responsibilities

- Taught selected topics in Information and Communication Technology to high school students.
- Facilitated small group discussions on topics covered in class.

- Facilitated small group discussions on topics covered in class.

VOLUNTARY EXPERIENCE

Sostegno della Vita, Rome Italy 2018-2021

Responsibilities

- Helped distribute clothes, food, and the Word of God to the homeless at Rome Termini

Northern Empowerment Association (NEA), Ghana 2010- 2016

Responsibilities

- Volunteered with NEA teams (Health, education, food security, environmental protection, water sanitation, women empowerment, peace building) in deprived communities to meet their basic needs in a sustainable manner.
- Organised and administered the Electronic Medical Record and Reporting System of the eye laser service.
- Translated for a team of Canadian doctors.

TECHNICAL SKILLS

- **Basic Programming in MATLAB and R.**
- **Proficient user of Microsoft suite, SPSS.**
- **Basic user of Video Pad Video Editor.**
- **Animal Handling**

RESEARCH WORK

PARTITIONING PROBLEM SOLVING INTO ITS COMPONENT PARTS AND ITS EFFECT ON LEARNING

Department of Psychology, Syracuse University, New York.
(Ongoing research project under the supervision of Dr Daniel Corral)

TESTING THE TESTING EFFECT WITH FEATURAL AND RELATIONAL CATEGORIES
Department of Psychology, Syracuse University, New York.
(Ongoing research project under the supervision of Dr Daniel Corral)

March 2020
FORE-PERIOD DELAY EFFECT IN A STOP-SELECTIVE TASK
Faculty of Psychology and Medicine, University of Rome, La Sapienza
(*Master thesis project submitted under the supervision of Professors Stefano Ferraina and Fabrizio Doricchi)*

MANUSCRIPTS IN PREPARATION

Sarakpo E., & Corral, D. **EXTENDING THE TESTING EFFECT TO CATEGORY LEARNING: OBSERVATIONAL VERSUS CLASSIFICATION TRAINING WITH FEATURE AND RELATION-BASED**. Manuscript in preparation.

Sarakpo E., & Corral, D. **PARTITIONING PROBLEM SOLVING INTO ITS COMPONENT PARTS AND ITS EFFECT ON LEARNING.** Manuscript in preparation.