August 2019

The Effects of Feature Verbalizablity and Indirect Feedback on Implicit Category Learning

Bailey N. Brashears
The University of Western Ontario

Supervisor
Minda, John P
The University of Western Ontario

Graduate Program in Psychology

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

Abstract

This study consisted of two experiments intended to investigate the effects of varying factors on the use of verbal and implicit classification systems when learning novel categories in an interactive video game environment. Experiment 1 measured the effects of feature type (easy vs difficult to describe verbally), and Experiment 2 measured the effect of direct vs indirect feedback. Verbal and implicit classification were operationalized by measuring rule-based and family resemblance strategy use respectively. Experiment 1 found that participants presented with stimuli that were easy to describe verbally were more likely to use rule-based classification, while participants presented with stimuli that were difficult to describe verbally showed no preference for one form of classification. Experiment 2 found that participants favoured rule-based classification regardless of whether they received direct or indirect feedback. The results of this study open up a novel field of research within category learning, further exploring the effects of feature verbalizablity.

Keywords: Category learning, COVIS theory, feature verbalizablity

Summary for Lay Audience

This study was interested in how changes in the environment can affect how people learn how to sort things into groups. People can learn how to sort things with different strategies: you can come up with a rule based on one aspect of the item you're trying to sort, or you can learn what each group looks like by paying attention to the overall appearance of the item. In order to do this, we had everyone who participated play a video game we designed where they had to sort cartoon monsters into one of two groups and/or feed them a certain amount of food. In Experiment 1, the game only involved sorting the monsters. In this experiment, some players saw monsters that had features that were easy to describe verbally, while others saw monsters with features that were difficult to describe verbally. We predicted that having features that were easy to describe would make it easier for players to come up with a rule to describe which monsters belonged to each group. The player's performance matched this prediction. In Experiment 2, players would either sort the monster into a group and then guess how much to feed them, or would only guess how much to feed them. This was based on a previous study where players would do this on paper, but we transferred it to a video game environment. This study found that when players were not asked to sort the monsters into groups, they were less likely to come up with a simple rule for what groups the monsters belonged to; instead they based their decisions on the overall look of the monsters, not any one aspect of their appearance. This is because asking players to sort the monsters into groups made them realize there were groups to learn and they were actively trying to use a rule to learn them. The performance of our players did not match this prediction. We found that players were more likely to use a rule than base their decision on the overall look whether or not they were asked to sort them into groups first.

Table of Contents

The Effects of Feature Verbalizablity and Indirect Feedback on Implicit Category Learning

Category learning is a part of everyday life. People form, update, and use a variety of categories to make classification decisions about a variety of things in daily life, from animals to food to vehicles. Most category learning and use takes place without an individual's conscious awareness and - in the average adult - with a high level of accuracy. However, there is a wide variety of factors that can influence how we determine which things belong to what category, and how our brains process the categories themselves.

Review of Literature

The COVIS (**Co**mpetition between **V**erbal and **I**mplicit **S**ystems) theory of category learning posits that category learning is accomplished by two separate but competing systems: the verbal system, which deals with learning explicit category rules, and the implicit system, which involves learning more complex categories through the procedural learning of multiple exemplars (Ashby & Maddox, 2011). The *verbal* system operates using active hypothesis testing, with individuals making, testing, and revising rules on a conscious level. This kind of category learning is effective when learning categories with a feature-based rule or series of rules determining category membership. However, individuals are capable of learning much more complex categories through the use of the implicit system. The *implicit* system can operate outside of conscious awareness and uses dopamine-mediated learning in order to gradually acquire categories based on covariation of features, family resemblance, and other more complex and subtle distinctions that can broadly be described as non-rule-described categories (Minda & Miles, 2010). Generally, the verbal category learning system can be considered to learn categories with rules that are easy to verbalize, while the implicit category learning system can be considered to learn categories with rules that are difficult to verbalize (Ashby & Maddox, 2011). This dual-route theory has basis in neurobiological findings, as well as in behavioral findings (Ashby & Casale, 2003; Ashby & Maddox, 2011).

According to COVIS theory, both of these systems work simultaneously, but there is a bias towards the verbal system that results in implicit learning being initially not expressed, as the implicit process is slower. In general, an individual will attempt to consciously find a rule-based solution during categorization tasks, and implicit categorization will occur if conscious verbal categorization is unsuccessful (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Maddox & Ashby, 2004). When classifying ambiguous stimuli that could be classified according to either rule-based (verbal) or family resemblance (implicit) strategies, both European Americans and Asian Americans as well as East Asians showed significant preference for using rule-based categorizations strategies (Norenzayan, Smith, Kim, & Nisbett, 2002). However, there has been recent work calling this bias into question; a study modelled on the work of Norenzayan et al failed to replicate the results and instead found a preference for family resemblance categorization strategies in their sample of European American participants (Murphy, Bosch, & Kim, 2017).

While the effects of language are easiest to see in verbal category learning, language can be an important factor for both category learning routes. Research on the role of language in category learning has found that merely introducing novel words as labels can highlight the item categories (Balaban & Waxman, 1997). The inclusion of existing category labels in a categorization task can also affect the perception of individual category member features (Lupyan, 2008). One study found that embedding a simple two-dimensional stimuli (a Gabor patch with varying orientation and spacing of lines) within category irrelevant features and labelling the stimuli holistically as a “flower” decreased reliance on individual features of the stimulus (Perry & Lupyan, 2014). This study compared Gabor patches with no additional features- referred to as “minerals”- with Gabor patches with the additional feature of flower petals; participants attended to orientation more in the mineral condition than in the flower condition. Perry and Lupyan concluded that placing a stimulus in a richer visual context distracted participants from the normally visually salient feature of orientation.

Language can affect many facets of category learning; however, one specific factor that can affect which category learning strategies are favored is the verbalizablity of the item being categorized or of the item’s features. Feature verbalizablity refers to whether the individual features of a stimuli are easy or difficult to describe verbally. While the effect on the verbalizablity of stimulus features specifically has not been explored in the current literature, the findings of previous research on language and category learning as well as the theory of COVIS allow us to formulate predictions for this study. As verbal category descriptions facilitate category learning and verbal route categorization relies on easy to verbalize rules, we hypothesize that when presented with stimuli with features that are difficult to describe verbally, individuals will be less likely to use rule-based categorization. Lupyan’s work was largely concerned with assigning verbal labels to categories to increase holistic classification, while we expect the emphasis on verbalizablity of the individual features to include rule-based classification.

The nature of the items being classified is not the only factor in which categorization strategies are used; another important factor is the nature of the feedback given during category learning. Many category learning experiments deal in supervised category learning. In a supervised category learning task, an individual might be randomly presented instances from two categories, make a classification decision, and then receive corrective feedback on this decision. It is through this repeated process of deciding and correcting that an individual learns to associate items with the correct category (Pothos & Chater, 2005).

In contrast, many concepts and categories are learned without supervision in real-world situations. When an individual learns in an unsupervised manner, there are no predefined categories, nor is there feedback on category membership; all learning in a novel environment is unsupervised and organic (Clapper & Bower, 1991). Unsupervised learning, however, is not just a guessing process; research has found that prior knowledge can greatly facilitate the unsupervised category learning process, indicating that there *is* a learning aspect to unsupervised categorization (Clapper, 2007). Since individuals are usually aware of both the hypothesis-testing process and the rules it generates, most unsupervised learning could also be considered verbal-route category learning (Love, 2002; Nosofsky, Palmeri, & McKinley, 1994).

Another aspect of category learning that can affect acquisition is if the category is being learned in service of another task. (Ross, 1997) found a category use effect in tasks where subjects were asked to diagnose a disease based on a list of symptoms and then prescribe a treatment for the patient. Symptoms were all equally and perfectly predictive of a disease, while only two of the symptoms were predictive for the treatment. This experiment found that the symptoms that were predictive of the treatment came to be viewed as more predictive than the non-treatment predictive symptoms of the diagnosis itself, despite the fact that all the symptoms were equally predictive of disease. The study also found that this category use effect, where treatment-predictive symptoms were incorrectly viewed as more predictive of diagnosis than diagnostic-only symptoms, still occurs when symptoms are not perfectly predictive and when the second task is an arbitrary category use task. The category use effect demonstrates how category learning and one's perception of a stimulus' features can be influenced by knowledge gained from other tasks - even after learning the relevant category. Further research has suggested that this original category representation is retroactively changed by this kind of task (See: Ross, 2000).

In Minda and Ross' 2004 study, the researchers aimed to explore an indirect kind of category learning that contrasts with classification learning. Indirect category learning tasks, as defined by Minda and Ross, are similar to unsupervised category learning tasks in that subjects are not told that there are underlying categories in the task and learning categories is not presented as the goal of the task. However, like supervised category learning, feedback is provided and knowledge of the category structure will improve performance. While this kind of learning has not been fully explored in the current category learning research climate, it is common in many day-to-day situations where category decisions are made. An example of this kind of indirect category learning can be found in the world of gardening. An individual pulling weeds in a garden is likely to learn, through trial and error, that some weeds can be removed easily while some break off at the roots if not handled correctly. They can then learn to identify easy-to-pull versus difficult-to-pull weeds without any kind of supervision or feedback on plant category membership. The individuals would be unaware of the actual natural category of the weeds (e.g., their species) aside from any features that are in aid of the easy/difficult distinction. This is the kind of indirect learning that Minda and Ross replicated in a controlled experimental setting in this study.

The study itself consisted of participants being presented with fictitious animals comprised of five binary features. Participants were asked to determine how much food they thought each animal should receive. While there was an underlying grouping that affected the correct feeding choices, only some participants were asked to explicitly categorize them as part of the experiment. The other participants would have to learn the categories indirectly through the feedback on the feeding decision itself.

Minda and Ross (2004) found that the subjects who learned the categories through the indirect feedback were more likely to use family resemblance category learning strategies than those who received direct classification feedback. Because verbal category learning is largely comprised of active hypothesis testing, we can conclude that this hypothesis testing and deliberate categorization is less likely to occur when individuals are receiving indirect feedback.

Another notable exploration of indirect learning was in a 2007 study by Brooks, Squire-Graydon, & Wood, where the authors argued that people often express a belief that natural categories have simple classifications rules, even if those rules do not exists. They refer to this as the “simpler-than-it-is belief” and if people have this belief, they should seek simple, rule-based solutions even if there is no simple rule. Brooks et al. also predicted that people would be less likely to show this pattern if attention were diverted from classification. Consistent with Minda and Ross (2004), they found that when individuals has their attention diverted away from the category membership of the novel stimulus, they were more likely to rely on family resemblance categorization rather than rule-based categorization.

Indirect learning is not unique to category learning. This phenomenon has also been studied in the context of language learning and the learning of artificial grammar. The concept of statistical learning in artificial grammar has much in common with the paradigms of unsupervised learning as well as indirect learning. In these statistical learning paradigms, a participant is presented with a constant stream of a constructed language, with no further instructions or feedback. Despite this, individuals are still able to learn to distinguish words from nonwords in this artificial language (Saffran, Aslin, & Newport, 1996). While the research is not conclusive, indirect learning strategies for vocabulary may enhance the proficiency of language learners (Naeimi & Foo, 2015; Schmitt & McCarthy, 1997).

The simpler-than-it-is belief can be demonstrated by asking individuals a question such as, “What makes a dog a dog?” Most people would answer something simple like “it has fur” or “it has four legs,” but it stands to reason that any such simple rule would have exceptions and include incorrect items as well. When individuals are aware that category learning must occur in order to proceed, the default reaction is to find and test a simple rule, regardless of the fact that the vast majority of natural categories cannot be divided as such. It is because of this simpler-than-it-is belief - and the findings of Brooks et al - that the current study was designed from the beginning in such a way that it would not appear as a standard categorization task. By framing the experiment as a video game, we created a level of diverted attention in the design itself, before the task was even implemented.

Current Study

The current study consists of two experiments with the overarching purpose of examining factors that influence category learning strategy through the lens of COVIS theory. The purpose of Experiment 1 was to investigate the effects of feature type (easy to describe verbally vs difficult to describe verbally) on category learning preference. The experiment was conducted in a novel and interactive video game environment designed by the researcher. Feature verbalizablity is a possible factor in the selection of a category learning strategy that has not been addressed by the current literature. This experiment is intended as both an investigation of the effects of feature verbalizablity and a relatively simple proof-of-concept task built in the novel video game environment.

Our hypothesis for Experiment 1 was that individuals who learn categories with features that are easy to describe verbally will show a preference for rule-based categorization, while individuals who learn categories with features that are difficult to describe verbally may show a preference for family resemblance categorization. Our reasoning for this is that when individuals are learning stimuli with features that are easy to describe verbally, it may be easier to identify a single feature rule for classification, since verbal descriptions are important for this kind of category learning. Furthermore, when individuals cannot easily describe a feature in simple verbal terms, this may bias them towards using an implicit category learning strategy instead of trying to find a verbally described rule.

The purpose of Experiment 2 was to theoretically replicate Minda and Ross' study (2004) in the same video game environment as Experiment 1. This is not a direct replication, but rather used most of the original study procedures with novel stimulus in a more interactive environment. Experiment 2 was intended to investigate the effect of direct vs indirect feedback on category learning preference. Our hypothesis for Experiment 2 - consistent with the original study - is that individuals will be more likely to favor family resemblance strategies when they receive indirect feedback than individuals who receive direct classification feedback.

Experiment 1

Methods

Participants

Due to the simultaneous nature of the data collection, the participant demographics for Experiment 1 cannot be separated from those of Experiment 2.[1] Therefore, the demographics of the entire sample for both experiments will be discussed. A total of 185 students participated in both studies; all subjects were recruited through the University of Western Ontario's SONA system. The data of eight participants was lost due to errors from the gaming platform being used to conduct the experiments. There were 121 female participants, 62 male participants, and two participants who chose not to answer. 177 of the participants were between 17 and 22, while the remaining 8 participants were over 23 or chose not to report their age. There were 60 Chinese participants, 53 white participants, 26 South Asian participants, and the remaining 46 participants belonged to other races/ethnicities.

Seventy undergraduate students attending the University of Western Ontario participated in this experiment. Participants were compensated with course credit in an undergraduate psychology course. The seventy participants in Experiment 1 were drawn randomly from this subject pool. Two of the seventy participants (one easily-verbalizable, one not-easily verbalizable) were not included in the final sample as they failed to learn the training stimulus within our criterion.

[1] The collection of demographic information as well as participant recruitment was handled by the University of Western Ontario's SONA system. Individuals for both Experiment 1 and Experiment 2 were collected in this way so experiment participation would be mutually exclusive (i.e., participants signed up once and participated either in Experiment 1 or Experiment 2). However, the SONA system assigns a unique ID number that did not correspond to our internal subject numbers that indicated which experiment they participated in. For this reason, only demographics for the entire subject pool between Experiment 1 and Experiment 2 are available to us.

Materials

Stimuli. The stimuli were digital pixel images of fictional monsters constructed of five binary features. The first set consisted of features selected to be easy to describe: spots (two vs. four), eyes (two vs. three), ears (teal vs. orange), tail shape (square vs. triangle), and back colour (red vs. green). This easily verbalizable stimulus set is included in Appendix A. The second set also consisted of digital pixel images of fictional monsters constructed of five binary features, this time selected to be difficult to describe verbally: spots (uneven stripes vs. uniform polka dots), eyes (narrow and vertical vs. wide and horizontal), ears (short and pointed vs. long and floppy), nose shape (long and pointed vs. short and blunt), and back shape (short bumpy "spines" vs. tall "sail"). This not-easily verbalizable stimulus set is included in Appendix B.

In order to verify that the novel stimuli intended to have feature sets that were easily and not-easily verablizable were verifiably such, a norming study was conducted. The norming study was designed to collect descriptions of the features from one group of individuals and determine if another group of individuals could then identify the same feature based on those descriptions. This study consisted of two phases. In the first phase, 63 individuals participated in a Qualtrics study through the University of Western Ontario's SONA system. In this study, participants were shown one of the four stimulus prototypes followed by one of the features in isolation. They were then asked to describe the individual features in the provided text box. Then, for each feature, the two most common descriptors from the 63 responses were selected to be used in Phase Two.

In the second phase, 30 individuals participated in a Qualtrics study through Amazon's mTurk system. They were shown the top two descriptors from Phase One and asked, out of the two paired features, which was best fit by the descriptors. The accuracy from Phase Two was then calculated compared to the results from Phase One. A Paired Sample Student's t-test found that accuracy was significantly higher for the easily-verbalizable items (M = 0.95, SD = 0.09) than for the not-easily verbalizable items (M = 0.63, SD = 0.14), t(29) = 11.55, $p < .001$.

The participants in the second phase were better able to differentiate the feature pairs from descriptions generated in the first phase for the feature pairs in the easily-verbalizable category. This means that, when describing features in the easily-verbalizable category, the generated descriptions were more distinct from each other when compared to the description of the paired feature. To illustrate this, the description for the spot feature pair in the easily-verbalizable category set included “two yellow hearts” and “four yellow hearts” with “yellow spots” included in each description, a clear and easy to describe distinction. However, for the spot feature pair in the not-easily verbaizable category set, the descriptions collected included “green spot pattern” and “beige markings” with “green spots” included in each description; in this case, any of the descriptions used could apply to either of the members of the feature pair, so accuracy was lower in matching the description to the correct stimulus. From these results we can conclude that the easily verbalizable stimulus features are objectively easier to verbally describe than the not-easily verbalizable stimulus features.

Table 1

Training Stimulus															
Stimulus	1	2	3	4	5	Size	Food	Stimulus	1	2	3	4	5	Size	Food
Category A								**Category B**							
A1	0	0	0	0	0	0	4	B1	1	1	1	1	1	0	8
A2	0	1	0	0	0	0	4	B2	1	0	1	1	1	0	8
A3	0	0	1	0	0	0	4	B3	1	1	0	1	1	0	8
A4	0	0	0	1	0	0	4	B4	1	1	1	0	1	0	8
A5	0	0	0	0	1	0	4	B5	1	1	1	1	0	0	8
A6	0	0	0	0	0	1	7	B6	1	1	1	1	1	1	11
A7	0	1	0	0	0	1	7	B7	1	0	1	1	1	1	11
A8	0	0	1	0	0	1	7	B8	1	1	0	1	1	1	11
A9	0	0	0	1	0	1	7	B9	1	1	1	0	1	1	11
A10	0	0	0	0	1	1	7	B10	1	1	1	1	0	1	11
A11	0	0	0	0	0	2	11	B11	1	1	1	1	1	2	14
A12	0	1	0	0	0	2	11	B12	1	0	1	1	1	2	14
A13	0	0	1	0	0	2	11	B13	1	1	0	1	1	2	14
A14	0	0	0	1	0	2	11	B14	1	1	1	0	1	2	14
A15	0	0	0	0	1	2	11	B15	1	1	1	1	0	2	14

The binary notation for the category sets was taken from Minda and Ross (2004) and is shown in Table 1. Note that the food amounts listed in Table 1 were used only for Experiment 2, these values were not used in Experiment 1. Like the original study, each training category contained fifteen stimuli with five exemplars presented in three sizes. Small stimuli (Size 0) were presented at 75 x 51 pixels, medium stimuli (Size 1) were presented at 150 x 102 pixels, and large stimuli (Size 2) were presented at 225 x 153 pixels. Each category set generated for the training phase included the prototype and four stimuli that varied by a single feature; one random feature would always be present on all stimuli in the category. This was designed such that the category could be learned either by a family resemblance (FR) or a criterial attribute (CA) strategy. A family resemblance strategy could be used due to the similar appearance of each stimulus to the prototype while a criterial attribute strategy could be used due to the single feature which perfectly predicts category membership.

Table 2

							Testing Stimulus													
Stimulus	**1**	**2**	**3**	**4**	**5**	**Size**	**Stimulus**	**1**	**2**	**3**	**4**	**5**	**Size**	**Stimulus**	**1**	**2**	**3**	**4**	**5**	**Size**
Training Items							**Exception Items**							**Single Feature**						
TA1	0	0	0	0	0	1	T1	1	0	0	0	0	1	T11	0	X	X	X	X	1
TA2	0	1	0	0	0	1	T2	1	1	0	0	0	1	T12	X	0	X	X	X	1
TA3	0	0	1	0	0	1	T3	1	0	1	0	0	1	T13	X	X	0	X	X	1
TA4	0	0	0	1	0	1	T4	1	0	0	1	0	1	T14	X	X	X	0	X	1
TA5	0	0	0	0	1	1	T5	1	0	0	0	1	1	T15	X	X	X	X	0	1
TB1	1	1	1	1	1	1	T6	0	1	1	1	1	1	T16	1	X	X	X	X	1
TB2	1	0	1	1	1	1	T7	0	0	1	1	1	1	T17	X	1	X	X	X	1
TB3	1	1	0	1	1	1	T8	0	1	0	1	1	1	T18	X	X	1	X	X	1
TB4	1	1	1	0	1	1	T9	0	1	1	0	1	1	T19	X	X	X	1	X	1
TB5	1	1	1	1	0	1	T10	0	1	1	1	0	1	T20	X	X	X	X	1	1

In addition to the training stimuli, a group of test stimuli was also generated to be used in the study and is shown in Table 2. The test stimulus consisted of three subsets of ten items each; all test stimuli were presented at the medium (1) size. The first subset- items TA1-5 and TB1-5- were repetitions of the training stimulus used to gage the participant’s accuracy. The second subset- items T1-10- were exception items taken from the original Minda and Ross (2004) study;

these items had conflicting category membership based on whether the participant used a FR or CA strategy and were used to differentiate these strategies. The final subset- items T11-20- were each of the single features presented in isolation used to gage the participant's attention towards the individual features that comprised the two category prototypes.

Procedure

The experiment was conducted using a video game programmed in GameMaker Studio 2. All responses were recorded by the program and saved to a plain text document on the testing machine. Randomization of condition and CA feature was also handled by the program.

Figure 1. Screenshot of Experiment 1 Gameplay (Easily-Verbalizable Condition)

Participants were randomly assigned to one of two stimulus sets: easily-verbalizable features and not-easily-verbalizable features. This affected only the visual appearance of the stimuli used and the experiment proceeded identically regardless of condition. Participants in both conditions were introduced to the game "Monster Farm" where they would play as a farmer and their job was to figure out which group each of the “monsters” living on a farm belonged to. They were also told that some features or aspects of the monsters might help them determine the correct group, either group A or group B. On each trial, they would need to select a collar for the monster with the group letter using either the A or B key. If the classification was correct, they were shown a check mark; if the classification was incorrect, they were shown an X. This feedback was displayed with the stimulus still visible until the next trial began automatically after a few seconds.

The training phase would continue until a subject had completed at least four blocks (60 trials) and completed at least twelve trials in a row correctly. If a subject did not complete this criteria within 400 trials or one hour, the game would end without the testing phase and the participant’s data would not be further analyzed. After reaching the criterion, both groups entered the testing phase. In this phase, subjects were shown each of the test stimuli in Table 2 and sorted the monster, as in the training phase, into the group they thought it belonged to. In this phase, there was no feedback on their decisions and each stimulus was shown once in a randomized order.

Results

Learning Rate Analysis

The first analysis focused on the number of trials it took a participant to reach criterion in each condition. This was defined as making a correct categorization decision on 12 trials in a row, after a minimum of 60 trials. A Welch’s Two Sample t-test found that learning rate for the easily-verbalizable condition ($M = 82.68$, $SD = 50.23$) and the not-easily verbalizable condition ($M = 89.50$, $SD = 43.2$) did not differ significantly from one another, $t(64.56) = -0.60$, $p = .550$.

These results are visualized in Figure 2. Note that the minimum number of trials a participant could complete the training phase in was 60, regardless of if the criteria was met before 60 trials.

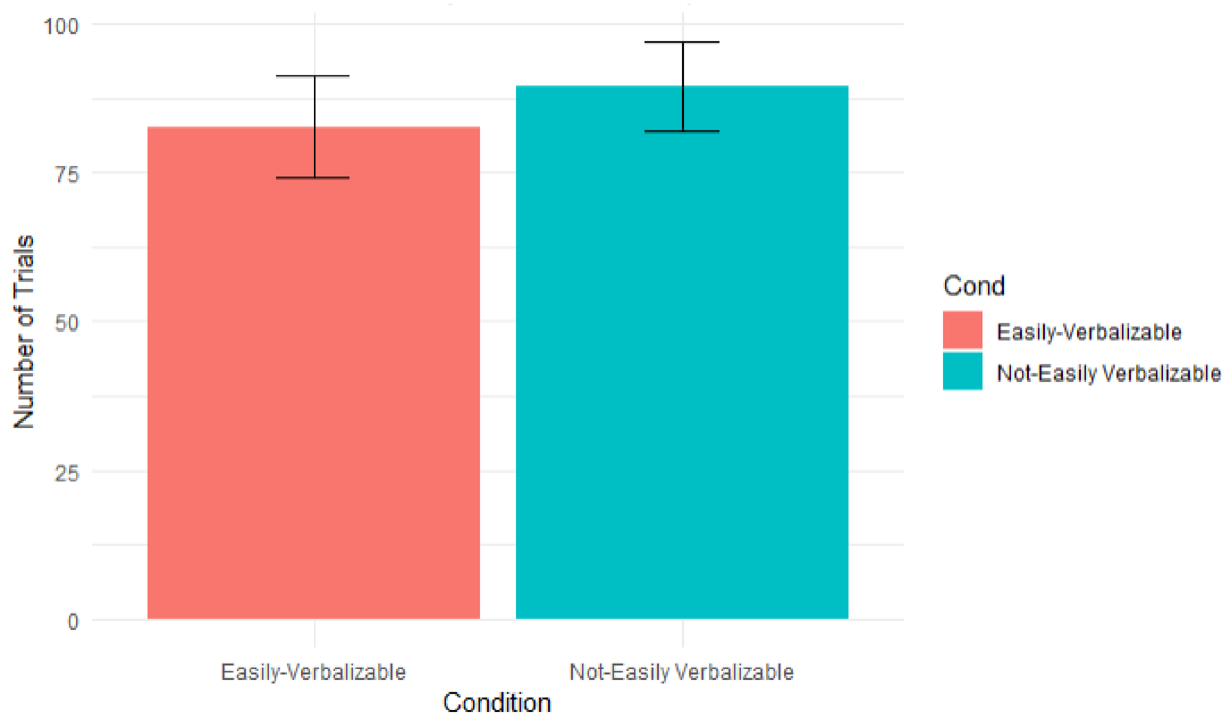


Figure 2. Mean Number of Trials to Criterion by Condition in Experiment One

Accuracy Analysis

The second analysis focused on participant's accuracy on the training items as presented during the testing phase in each condition. All ten training items were presented, randomized with the other testing stimuli and presented only in the medium size. A Welch's Two Sample t-test found that accuracy was significantly higher in the easily-verbalizable condition (M = 0.93, SD = 0.12) than in the not-easily verbalizable condition (M = 0.80, SD = 0.18), t(58.47) = 3.50, p = .001. These results are visualized in Figure 3.

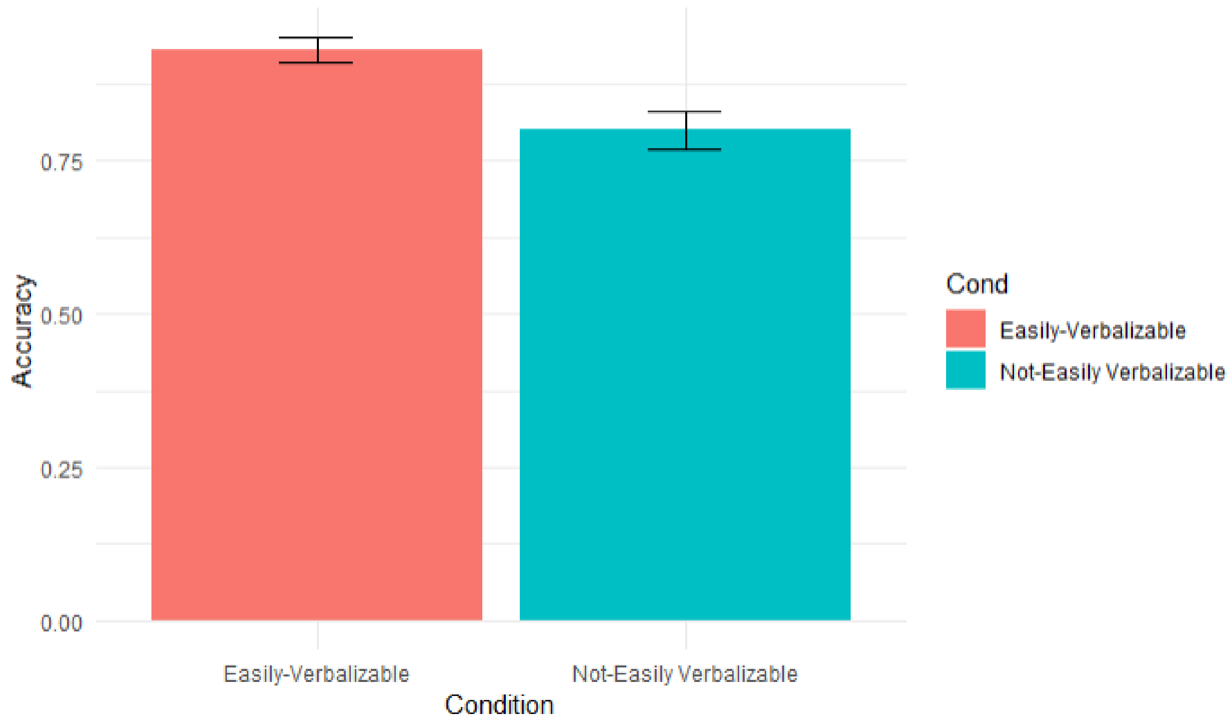


Figure 3. Mean Accuracy in the Testing Phase by Condition in Experiment One

Categorization Strategy Modelling Analysis

The final analysis of this data set focused on the categorization strategies used by the participants. Modelling was conducted using participants' responses to the ten exception items presented in the testing phase, randomized with the other testing stimuli and presented only in the medium size. The exception items were designed such that a participant using a FR rule and a participant using a CA rule would categorize the stimulus in opposite categories. Therefore, based on the participants' responses, we can determine what strategy each participant was most likely relying on.

These responses were then compared to the results predicted by twelve model responses and each participant was given a score out of one indicating how well their responses matched the predicted results of each model. For each matching response, the individual was given a 1. For each response that didn't match, the individual was given a 0. These scores were added together and divided by the number of items (ten) with a possible score of 1 indicating a perfect fit with the model.

The twelve models included responses based purely on a correct criterial attribute strategy or a family resemblance strategy; guessing models based on responding with only A or only B; responses based on a rule-based strategy using a feature other than the correct criterial attribute; responses based on a family resemblance strategy with a rule-based exception other than the correct criterial attribute.

For the correct criterial attribute strategy model, individuals' responses were compared to the expected responses if one was only attending to the criterial attribute. This would entail responding with A when the criteria attribute matched group A but three or four of the remaining features were consistent with group B. This indicates that the individual is ignoring most or all of the other features of the stimulus and only using the criterial feature to make their category decisions. The family resemblance strategy model compared the responses to the opposite pattern, one that was expected if one was attending to overall family resemblance regardless of the criterial attribute. This would entail responding B when three or four of the features were consistent with group B but the criterial attribute matched group A. This would indicate that the individual is only looking at the overall family resemblance to make their category decisions and is ignoring the criterial feature.

The A and B models were straightforward, with individual responses compared to all A responses or all B responses. The other rule (OR) set of models each assumed one was using rule-based categorization, but used an incorrect feature in their rule. These models were included because individuals could appear mostly accurate (80% accuracy, 0.8 rule-based model fit) but have not identified the correct criterial attribute. In contrast, the exception-rule (ER) set of models each assumed one was using family resemblance categorization, but had a rule-based exception for one of the features other than the ceriterial attribute. Like the OR models, the ER models were included to check for individuals that appeared mostly accurate to one of the main models, but actually failed to use the strategy assumed by those models. Therefore, all ten models- A, B, OR, and ER- served as safeguards in the data and allowed us to ensure that the strategies we theorized were actually being used by participants.

The model fits were calculated using Microsoft Excel for each participant and the results were visualized as a heat map (see Figure 4). Each row represents one participant, with their scores for each model included. In this heat map, green indicates good fit for the corresponding mode while red indicates a poor fit. These colours were generated in R, green was assigned to a score of 1, red to a score of 0, and yellow to a score of 0.5; colours for scores between these values were then automatically generated and a corresponding background colour was set for each value.

Experiment One, Easily-Verbalizable Condition Model Fits

CA	FR	A	B	OR1	OR2	OR3	OR4	ER1	ER2	ER3	ER4
0.7	0.3	0.4	0.6	0.5	0.7	0.5	0.5	0.5	0.3	0.5	0.5
0.8	0.2	0.5	0.5	0.6	0.6	1	0.6	0.4	0.4	0	0.4
0.3	0.7	0.4	0.6	0.5	0.1	0.3	0.5	0.5	0.9	0.7	0.5
0.2	0.8	0.5	0.5	0.4	0.4	0.4	0	0.6	0.6	0.6	1
0.1	0.9	0.4	0.6	0.3	0.1	0.3	0.3	0.7	0.9	0.7	0.7
0.2	0.8	0.5	0.5	0.4	0	0.4	0.4	0.6	1	0.6	0.6
0.3	0.7	0.4	0.6	0.3	0.1	0.5	0.5	0.7	0.9	0.5	0.5
0.7	0.3	0.8	0.2	0.5	0.5	0.7	0.7	0.5	0.5	0.3	0.3
0.9	0.1	0.6	0.4	0.7	0.9	0.7	0.7	0.3	0.1	0.3	0.3
0.9	0.1	0.6	0.4	0.7	0.9	0.7	0.7	0.3	0.1	0.3	0.3
0.3	0.7	0.4	0.6	0.5	0.1	0.3	0.5	0.5	0.9	0.7	0.5
0.9	0.1	0.4	0.6	0.7	0.9	0.7	0.7	0.3	0.1	0.3	0.3
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
0.8	0.2	0.3	0.7	0.6	0.6	0.8	0.8	0.4	0.4	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
0.9	0.1	0.4	0.6	0.9	0.7	0.7	0.7	0.1	0.3	0.3	0.3
0.9	0.1	0.4	0.6	0.9	0.7	0.7	0.7	0.1	0.3	0.3	0.3
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2

Experiment One, Not-Easily Verbalizable Condition Model Fits

CA	FR	A	B	OR1	OR2	OR3	OR4	ER1	ER2	ER3	ER4
0.6	0.4	0.3	0.7	0.4	0.4	0.8	0.6	0.6	0.6	0.2	0.4
0.5	0.5	1	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.4	0.6	0.1	0.9	0.4	0.4	0.6	0.4	0.6	0.6	0.4	0.6
0.3	0.7	0.4	0.6	0.5	0.1	0.3	0.5	0.5	0.9	0.7	0.5
0.9	0.1	0.4	0.6	0.7	0.7	0.9	0.7	0.3	0.3	0.1	0.3
0.8	0.2	0.5	0.5	0.6	0.8	0.6	0.8	0.4	0.2	0.4	0.2
0.2	0.8	0.7	0.3	0.2	0.4	0.2	0.4	0.8	0.6	0.8	0.6
0.9	0.1	0.4	0.6	0.7	0.7	0.7	0.7	0.3	0.3	0.3	0.3
0.1	0.9	0.6	0.4	0.3	0.3	0.3	0.1	0.7	0.7	0.7	0.9
0.4	0.6	0.9	0.1	0.4	0.4	0.4	0.4	0.6	0.6	0.6	0.6
0.2	0.8	0.5	0.5	0	0.4	0.4	0.4	1	0.6	0.6	0.6
0.9	0.1	0.6	0.4	0.7	0.7	0.7	0.9	0.3	0.3	0.3	0.1
0.3	0.7	0.6	0.4	0.5	0.3	0.5	0.1	0.5	0.7	0.5	0.9
0.1	0.9	0.6	0.4	0.3	0.3	0.1	0.3	0.7	0.7	0.9	0.7
0.3	0.7	0.4	0.6	0.1	0.3	0.5	0.5	0.9	0.7	0.5	0.5
0.2	0.8	0.5	0.5	0	0.4	0.4	0.4	1	0.6	0.6	0.6
0.2	0.8	0.5	0.5	0.4	0.4	0	0.4	0.6	0.6	1	0.6
0.2	0.8	0.5	0.5	0.4	0.4	0.4	0	0.6	0.6	0.6	1
0.2	0.8	0.3	0.7	0.4	0.4	0.2	0.2	0.6	0.6	0.8	0.8
0.3	0.7	0.4	0.6	0.5	0.5	0.3	0.1	0.5	0.5	0.7	0.9
0.1	0.9	0.6	0.4	0.1	0.3	0.3	0.3	0.9	0.7	0.7	0.7
0.3	0.7	0.4	0.6	0.5	0.5	0.3	0.1	0.5	0.5	0.7	0.9
0	1	0.5	0.5	0.2	0.2	0.2	0.2	0.8	0.8	0.8	0.8
0.2	0.8	0.5	0.5	0.4	0.4	0	0.4	0.6	0.6	1	0.6
0.5	0.5	0.4	0.6	0.7	0.3	0.5	0.3	0.3	0.7	0.5	0.7
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
0.9	0.1	0.6	0.4	0.7	0.9	0.7	0.7	0.3	0.1	0.3	0.3
0.9	0.1	0.6	0.4	0.7	0.9	0.7	0.7	0.3	0.1	0.3	0.3
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
0.8	0.2	0.7	0.3	0.6	0.8	0.6	0.6	0.4	0.2	0.4	0.4
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2
1	0	0.5	0.5	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2

Figure 4. Heat Map Analysis of Model Fits in Experiment One

Following the heat map analysis, we decided to focus on the two main model fits, FR and CA. A mixed 2x2 ANOVA was conducted with model type as a within-subjects independent variable (CA vs. FR), stimulus set as a between-subjects independent variable (easily-verbalizable vs. not-easily verbalizable), and model fit as the dependent variable. There was no main effect of condition, $F(1, 66) < .01$, $p > .999$. There was a significant main effect of model type, indicating that the CA model ($M = 0.67$, $SD = 0.35$) was overall a better fit for participants than the FR model ($M = 0.33$, $SD = 0.35$) regardless of condition, $F(1, 66) = 19.12$, $p < .001$. There was also a significant interaction effect, $F(1, 66) = 14.78$, $p < .001$. These results are visualized in Figure 5. Due to the binary nature of the responses and the design of the exception items, the FR and CA model fits are inverse of one another as these model fits are mutually exclusive.

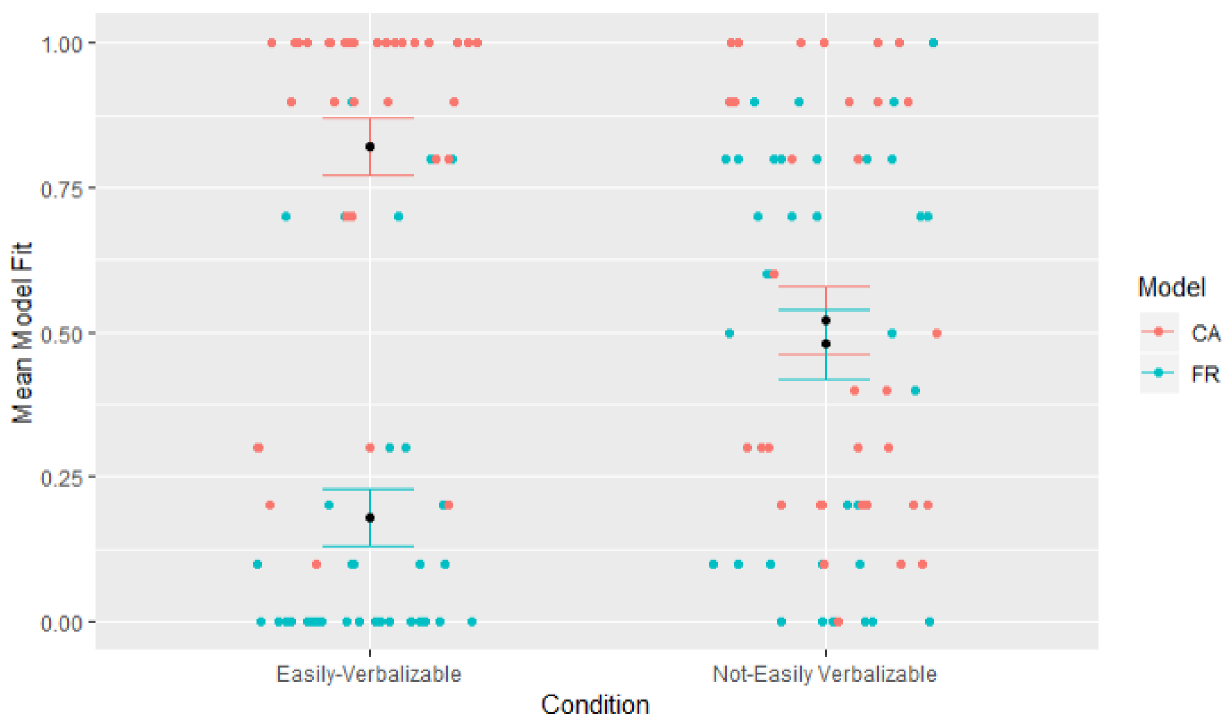


Figure 5. Model Fit (CA and FR) by Condition for Experiment One

Discussion

We found in this experiment that the easily-verbalizable stimulus set and not-easily verbalizable stimulus set were both equally easy to learn, as seen in the learning rate analysis, but participants were better at retaining the category membership of the easily-verbalizable stimulus set into the training phase than the not-easily verbalizable stimulus set, as seen in the accuracy analysis.

Additionally, our modelling analysis found that individuals in the easily-verbalizable condition were more likely to use CA strategies, while individuals in the not-easily verbalizable condition did not show a strong preference for either strategy. This result is consistent with our hypothesis that a stimulus with not-easily verbalizable features would facilitate implicit category learning strategies; individuals in the not-easily verbalizable condition were more likely to use family resemblance strategies than those in the easily-verbalizable condition even if there was not a strong preference for them. Without much, if any, previous research on the effects of feature verbalizablity on category learning, this experiment was largely exploratory.

This preference for rule-based learning is further seen by the finding that individuals were more likely to retain the category membership knowledge into the testing phase in the easily-verbalizable condition, as individuals were able to use the previously discovered rules to easily sort the stimulus again without having to rely on knowledge of the prototype.

Consistent with the findings of Perry and Lupyan (2014), we expected labelling the stimulus as “monsters” and placing them in a rich visual context to cause participants to view the stimuli more holistically, rather than attending only to the discrete features they contained. Similar to Perry and Lupyan’s (2014) flowers, our stimuli had a variety of visual features irrelevant to their category membership. Furthermore, we attempted to extend this concept by varying the verbalizablity of the individual features. From these results, we can conclude that feature verbalizablity does play a role in category learning strategy selection and features that are easy to describe verbally are conducive to rule-based strategy use and verbal route category learning.

Experiment 2

Methods

Participants

Due to the simultaneous nature of the data collection, the participant demographics for Experiment 2 cannot be separated from those of Experiment 1. 107 undergraduate students attending the University of Western Ontario participated in this experiment. The 107 students were drawn randomly from the previously discussed sample. Participants were compensated with course credit in an undergraduate psychology course. 30 of the 107 participants (PO = 20, CP = 10) were not included in the final sample as they failed to learn the training stimulus within our criterion. This was expected, as our pilot testing found that this task was more difficult than the one used in Experiment 1 and a larger sample was collected to compensate for this.

Materials

The “easily verbalizable” stimulus from Experiment 1 were again used in Experiment 2. The underlying representation, including the generation of training and test stimuli, were identical to Experiment 1.

Procedure

Like Experiment 1, this experiment was conducted using a video game programmed in GameMaker Studio 2. All responses were recorded by the program and saved to a plain text document on the testing machine. Randomization of condition and CA feature was also handled by the program.

Figure 6. Screenshot of Experiment 2 Gameplay (CP Condition)

Participants were randomly assigned to one of two learning conditions: prediction-only (PO) or classification-and-prediction (CP). Participants in the PO condition were introduced to the game “Monster Farm” where they would play as a farmer and their job was to figure out how much food each of the “monsters” living on a farm should receive. They were also told that some features or aspects of the monsters might help then determine the correct amount, which ranged from four to fourteen pounds. On each trial, the participant saw the monster and used the 2 and 1 keys to add and subtract pounds of food from a bowl on screen (the program only allowed between four and fourteen pounds to be in the bowl). They then used the arrow keys to move their character towards the monster and “feed” it by touching it with the bowl. A small word balloon with a heart would appear if they were correct. If they were incorrect, the correct amount of food the monster should have received would appear instead.

Participants in the CP conditions were given prediction instructions as above but were also told that before feeding the monsters they would need to first sort them into two categories- group A or group B. On each trial, they would need to select a collar for the monster with the group letter using either the A or B key. If the classification was correct, they were shown a check mark; if the classification was incorrect, they were shown an X. The trial would then continue with the feeding task as in the PO condition.

The criterion was the same in Experiment 1; once the criterion was reached, the testing phase began. The testing phase was identical in methods and logic to Experiment 1 except a feeding task was used instead of an A/B categorization task.

Results

Learning Rate Analysis

The first analysis focused on the number of trials it took a participant to reach criterion in each condition. This was defined as making a correct feeding decision on 12 trials in a row, after a minimum of 60 trials. A Welch’s Two Sample t-test found that learning rate for the CP condition ($M = 111.95$, $SD = 40.64$) was significantly faster than in the PO condition ($M = 169.03$, $SD = 80.84$), $t(54.26) = 3.90$, $p < .001$. These results are visualized in Figure 7. Note that the minimum number of trials a participant could complete the training phase in was 60.

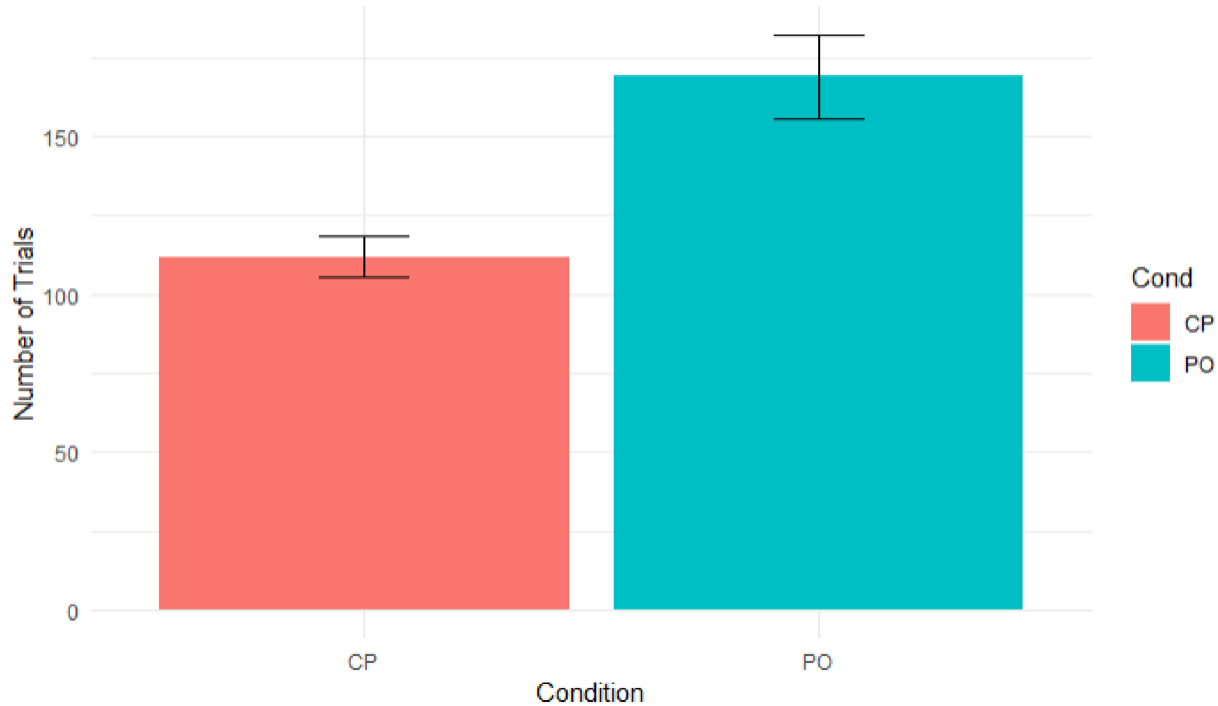


Figure 7. Mean Number of Trials to Criterion by Condition in Experiment Two

Accuracy Analysis

The second analysis focused on participant's accuracy on training items presented during the testing phase in each condition. All ten training items were presented, randomized with the other testing stimulus and presented only in the medium size. A Welch's Two Sample t-test found that accuracy in the PO condition (M = 0.74, SD = 0.26) and the CP condition (M = 0.78, SD = 0.23) did not differ significantly from one another, t(73.64) = -0.62, p = .539. These results are visualized in Figure 8.

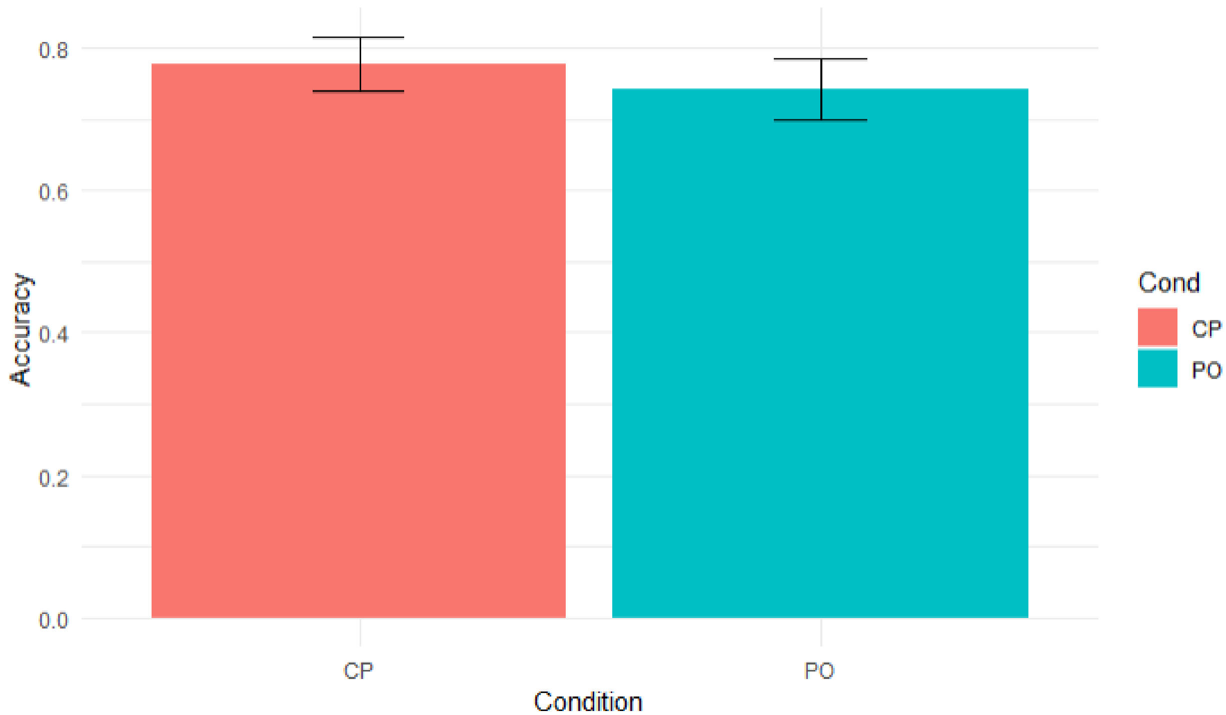


Figure 8. Mean Accuracy in the Testing Phase by Condition in Experiment Two

Categorization Strategy Modelling Analysis

The final analysis of this data set focused on the categorization strategies used by the participants. Modelling was conducted using participants' responses to the ten exception items presented in the testing phase, randomized with the other testing stimuli and presented only in the medium size. The exception items were designed such that a participant using a FR rule and a participant using a CA rule would categorize the stimulus in opposite categories. Therefore, based on the participants' responses, we can determine what strategy each participant was most likely relying on. The modelling analysis differed slightly from Experiment 1 due to the responses being a feeding decision instead of a direct category decision. Instead, we interpreted a response of 7 as a group A category decision and a response of 11 as a group B category decision, as these were the correct feeding responses for medium-sized stimulus in their respective groups.

These responses were then compared to the results predicted by twelve model responses and each participant was given a score out of one indicating how well their responses matched the predicted results of each model. For each matching response, the individual was given a 1. For each response that didn't match, the individual was given a 0. These scores were added together and divided by the number of items (ten) with a possible score of 1 indicating a perfect fit with the model.

The twelve models included responses based purely on a family resemblance strategy or a correct criterial attribute strategy; guessing models based on responding with only 7 or only 11; responses based on a family resemblance strategy with a rule-based exception other than the correct criterial attribute; and responses based on a rule-based strategy using a feature other than the correct criterial attribute. Numerical responses other than 7 or 11 were not considered; an alternative model was attempted that granted partial points for responses one point higher or lower than the desired response to account for participant mistakes, but this model did not produce significantly different results and was not used.

For the correct criterial attribute strategy model, individuals' responses were compared to the expected responses if one was only attending to the criterial attribute. This would entail responding A when the criterial attribute matched group A but three or four of the remaining features were consistent with group B. This indicates that the individual is ignoring most or all of the other features of the stimulus and only using the criterial feature to make their category decisions. The family resemblance strategy model compared the responses to the opposite pattern, one that was expected if one was attending to overall family resemblance regardless of the criterial attribute. This would entail responding B when three or four of the features were consistent with group B but the criterial attribute matched group A. This would indicate that the individual is only looking at the overall family resemblance to make their category decisions and is ignoring the criterial feature.

The 7 and 11 models were straightforward, with individual responses compared to all 7 responses or all 11 responses. The other rule (OR) set of models each assumed one was using rule-based categorization, but used an incorrect feature in their rule. These models were included because individuals could appear mostly accurate (80% accuracy, 0.8 rule-based model fit) but have not identified the correct criterial attribute. In contrast, the exception-rule (ER) set of models each assumed one was using family resemblance categorization, but had a rule-based exception for one of the features other than the criterial attribute. Like the OR models, the ER models were included to check for individuals that appeared mostly accurate to one of the main models, but actually failed to use the strategy assumed by those models. Therefore, all ten models- 7, 11, OR, and ER- served as safeguards in the data and allowed us to ensure that the strategies we theorized were actually being used by participants.

The model fits were calculated using Microsoft Excel for each participant and the results were visualized as a heat map (see Figure 9). Each row represents one participant, with their scores for each model included. In this heat map, green indicates good fit for the corresponding mode while red indicates a poor fit. These colours were generated in R, green was assigned to a score of 1, red to a score of 0, and yellow to a score of 0.5; colours for scores between these values were then automatically generated and a corresponding background colour was set for each value.

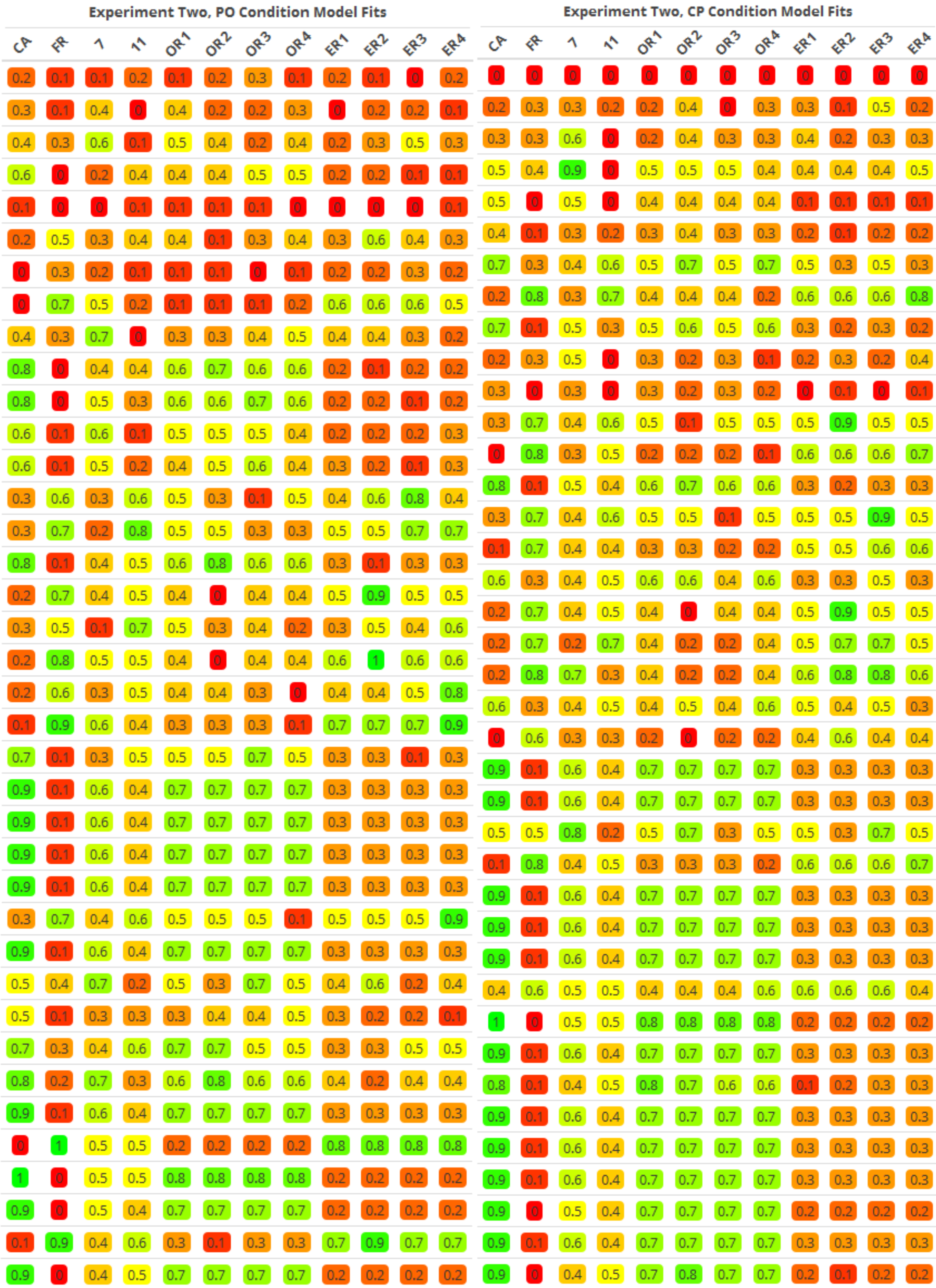


Figure 9. Heat Map Analysis of Model Fits in Experiment Two

Following the heat map analysis, we decided to focus on the two main model fits, FR and CA. A mixed 2x2 ANOVA was conducted with model type as a within-subjects independent variable (CA vs. FR), condition as a between-subjects independent variable (PO vs. CP), and model fit as the dependent variable. There was no main effect of condition, $F(1, 75) = 0.33$, $p = .566$. There was a significant main effect of model type, indicating that the CA model ($M = 0.52$, $SD = 0.33$) was overall a better fit for participants than the FR model ($M = 0.31$, $SD = 0.30$) regardless of condition, $F(1, 75) = 10.27$, $p = .002$. There was no significant interaction effect, $F(1, 75) = 0.05$, $p = .817$. These results are visualized in Figure 10.

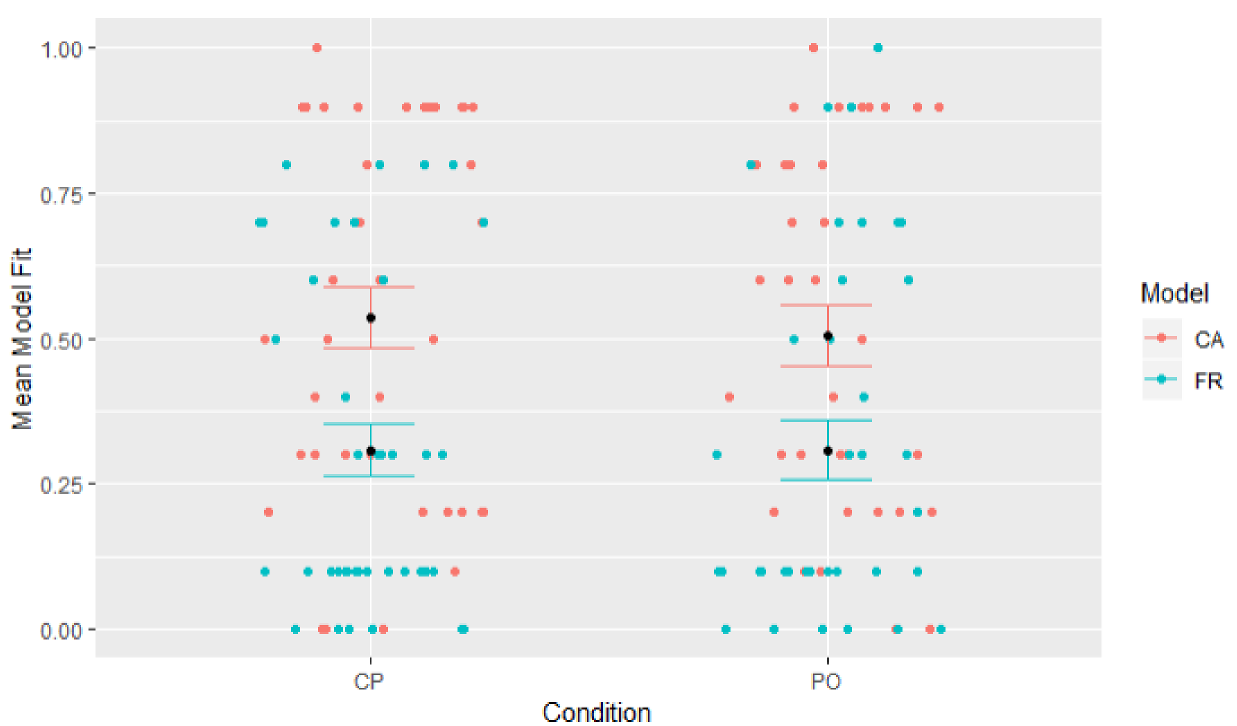


Figure 10. Model Fit (RB and FR) by Condition for Experiment Two

Discussion

We found in this experiment that individuals acquired the category knowledge faster in the CP condition than in the PO condition, as seen in the learning rate analysis, but participants seemed equally good at retaining the category membership into the training phase in both the PO and CP conditions, as seen in the accuracy analysis.

We failed to replicate the primary finding of Minda and Ross (2004) that indirect feedback biases individuals towards FR category learning strategies while individuals given direct categorization feedback were more likely to use CA category learning strategies. Instead, we found a preference for CA strategies regardless of condition. There was no difference in performance for direct vs indirect learning, the only demonstrated effect was that individuals given indirect feedback took longer to learn the category sets.

This finding is more unexpected than those of Experiment 1; as this was a theoretical replication of a previous study, we were anticipating similar results. The original study used the same underlying binary representation for their stimulus, but it's possible that these novel stimuli had less distinctive family resemblance between groups than in the Minda and Ross study.

Another possibility is the nature of the environment. The gaming environment could have been enough on its own to divert participants' attention, creating a level of indirect learning in both conditions. This could explain why we saw use of family resemblance category learning strategies in both conditions, though it does leave the question of why being directly asked to categorize the stimulus in the CP condition had no effect at all, unlike what was seen in the Minda and Ross study.

Conclusion

In Experiment 1 we demonstrated an interesting and novel effect of feature verbalizablity on preference for category learning strategies. Participants showed a strong and significant preference for rule based categorization when classifying the stimulus set with easily-verbalizable features compared to the stimulus set with not-easily verbalizable features. There was no strong preference for either rule-based or family resemblance strategies when classifying stimulus with not-easily verbalizable features. While the role of verbal labels and names has been of interest in categorization research, the effects of the verbalizablity of the individual features themselves is unexplored in the existing literature and this study demonstrates a need for further work.

In Experiment 2 we failed to replicate the findings of Minda and Ross (2004). Participants showed a preference for CA strategies regardless of feedback condition. There did not seem to be an effect of learning with a prediction task versus learning with a classification and prediction task on preference for category learning strategies. However, comparing the ratio of strategy use between Experiment 2 and the easily-verbalizable condition of Experiment 1, the preference for CA strategies seems less strong in Experiment 2 than in Experiment 1 despite using the same stimulus set and environment. In addition to the previously discussed possible effect of the environment and framing itself, there may have been an effect of just introducing the prediction task. Since participants in both conditions were presented with the prediction task as the ultimate goal, even if classification did occur, this may have been enough to divert attention from the classification task and caused indirect learning to occur in both conditions. While there was still a preference for CA learning in both conditions as well as Experiment 1, comparing the results does imply there is an effect of including a prediction task on preference for category learning strategy. In the future, another replication of Minda and Ross (2004) should be conducted, preferably a direct replication, in order to determine its validity. Our study varied in many minor aspects that may have influenced the results in unexpected ways.

In terms of possible limitations of this study, some researchers have proposed that, even when using an FR strategy, participants may only attend to one or two salient features (Smith, 2008). This would mean that a participant using an FR strategy may actually respond how we expected participants using a CA strategy would respond with attention paid only to the relevant categorization feature, which would complicate our model use in this study.

An important aspect of both Experiment 1 and Experiment 2 was the presentation of the stimulus as “monsters” within a video game, as opposed to collections of features to be classified. The findings of Perry and Lupyan (2014) indicate that this may have caused participants to attend less to the individual features of the stimuli. In the future, we would like to conduct a follow-up study with both stimulus sets in a traditional category learning experiment paradigm where they are not referred to holistically as “monsters.” This study would allow us to determine the effect of the overall category labelled compared the effect of the verbalizablity of the individual features.

The most interesting finding of this study came from Experiment 1. While our findings are not definitive as to the cause there are a few possible explanations. The first is that the easily-verbalizable features facilitated rule-based learning by encouraging use of verbal route categorization. When individuals are presented with features that they can easily describe to themself verbally it is easier for them to formulate and test a hypothesis (e.g., two spots = group A, four spots = group B). If an individual is presented with features that are more difficult to describe verbally it is more difficult to formulate a straightforward hypothesis; two different spot patterns that are difficult to describe distinctly can be harder to group into verbally described rules, since they are harder to describe. Another possibility is that the easily-verbalizable features appeared more distinct from one another compared to the not-easily verbalizable features. The easily-verbalizable features were designed with simple colour and number contrasts in order to be easy to describe verbally. However, this may have caused a larger visual contrast in this group compared to the not-easily verbalizable group, which had consistent colours and numbers, which means the results may have been due to the features appearing more distinct, so individuals attend to the individual features more for this stimulus group. Furthermore, the not-easily verbalizable stimulus set might have been viewed more holistically due to this decreased visual contrast, leading to more use of family resemblance categorization.

In the future, we would like to conduct additional research to further investigate these findings and identify the underlying cause. A possible study that could be used to identify if the results were the result of the features' verbalizablity or the visual distinctiveness could be to replicate Experiment 1 with two new stimulus sets. These stimulus sets would be balanced to control for their visual distinctiveness, possibly by using grey-scale stimuli or using less extreme colour contrasts than in the original study (our study used complimentary colours) or by adding colour contrasts to the not-easily verbalizable stimulus, with varying hues of the same colour to create a contrast that is visually distinctive but relatively difficult to describe verbally.

Overall, this study had novel findings that open up a new possible line of category learning research and demonstrated that a previous category learning study has results that may not be replicable. Additionally, we successfully created a new platform for conducting classification experiments that can be developed and adapted for future studies as we continue to better understand the effects of this video game environment on category learning. While our work on indirect feedback and diverted attention was inconclusive, we demonstrated that language is an important factor in category learning strategy use that should be explored further.

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., & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. *Advances in Consciousness Research*, *48*, 109–142.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*(1), 3–26.

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: identification in the service of use. *Memory & Cognition*, *35*(1), 1–14.

Clapper, J. P. (2007). Prior knowledge and correlational structure in unsupervised learning. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *61*(2), 109–127.

Clapper, J. P., & Bower, G. H. (1991). Learning and Applying Category Knowledge in Unsupervised Domains. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 27, pp. 65–108). Academic Press.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, *9*(4), 829–835.

Lupyan, G. (2008). From chair to“ chair”: A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology. General*, *137*(2), 348.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*(3), 309–332.

Minda, J. P., & Miles, S. J. (2010). Chapter 3 - The Influence of Verbal and Nonverbal Processing on Category Learning. In *Psychology of Learning and Motivation* (Vol. 52, pp. 117–162). Academic Press.

Minda, J. P., & Ross, B. H. (2004). Learning categories by making predictions: an investigation of indirect category learning. *Memory & Cognition*, *32*(8), 1355–1368.

Murphy, G. L., Bosch, D. A., & Kim, S. (2017). Do Americans Have a Preference for Rule-Based Classification? *Cognitive Science*, *41*(8), 2026–2052.

Naeimi, M., & Foo, T. C. V. (2015). Vocabulary Acquisition through Direct and Indirect Learning Strategies. *English Language Teaching*, *8*(10), 142.

Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, *26*(5), 653–684.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66–72.

Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *58*(4), 733–752.

Ross, B. H. (1997). The Use of Categories Affects Classification. *Journal of Memory and Language*, *37*(2), 240–267.

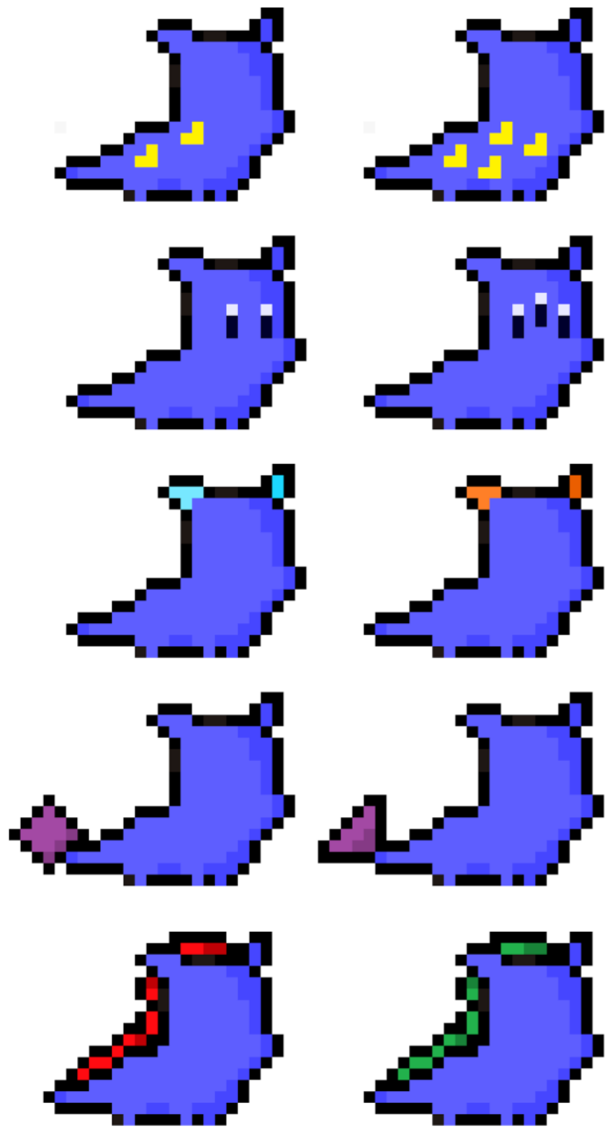
Ross, B. H. (2000). The effects of category use on learned categories. *Memory & Cognition*, *28*(1), 51–63.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Schmitt, N., & McCarthy, M. (1997). *Vocabulary: Description, acquisition and pedagogy*. researchgate.net.

Smith, E. E. (2008). The case for implicit category learning. *Cognitive, Affective & Behavioral Neuroscience*, *8*(1), 3–16.

Appendix A. Stimulus Material Used in Experiment 1 and 2 (Easily Verablizable Features)

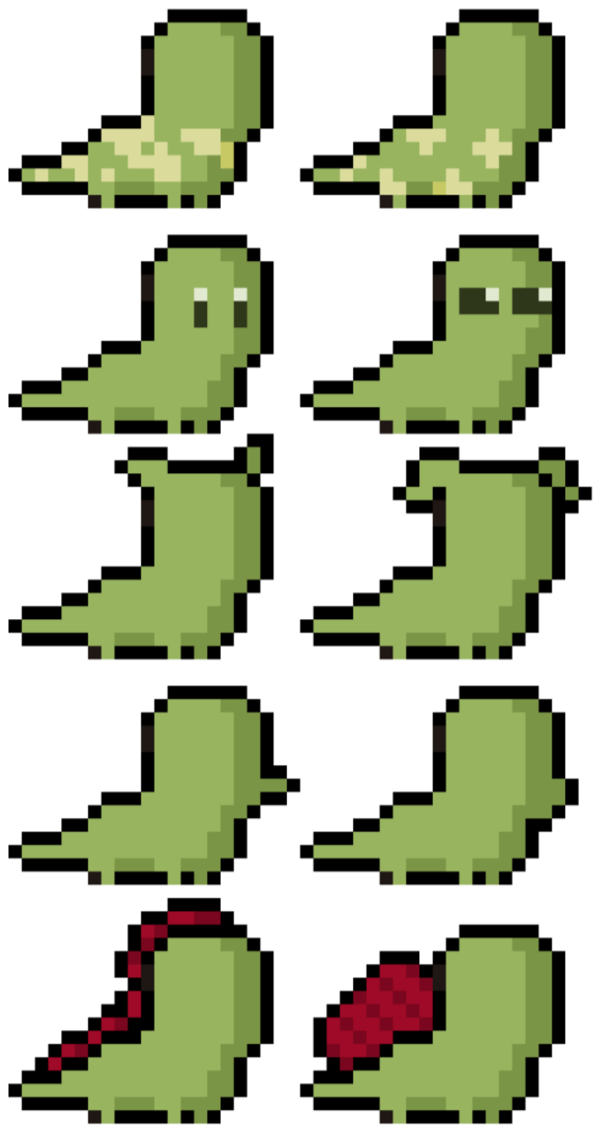


All five features of the easily verbalizable stimulus set. Group A (0) features are on the left, group B (1) features are on the right. Feature 1: spots, feature 2: eyes, feature 3: ears, feature 4: tail shape, and feature 5: back colour.

All possible members of group A for the easily verbalizable stimulus set; prototype followed by single feature variations. One feature was randomly removed from both groups in each trial so that one feature is always perfectly predictive of group membership.

All possible members of group B for the easily verbalizable stimulus set; prototype followed by single feature variations. One feature was randomly removed from both groups in each trial so that one feature is always perfectly predictive of group membership.

Appendix B. Stimulus Material Used in Experiment 1 (Not-Easily Verablizable Features)



All five features of the not-easily verbalizable stimulus set. Group A (0) features are on the left, group B (1) features are on the right. Feature 1: spots, feature 2: eyes, feature 3: ears, feature 4: nose shape, and feature 5: back shape.

All possible members of group A for the not-easily verbalizable stimulus set; prototype followed by single feature variations. One feature was randomly removed from both groups in each trial so that one feature is always perfectly predictive of group membership.

All possible members of group B for the not-easily verbalizable stimulus set; prototype followed by single feature variations. One feature was randomly removed from both groups in each trial so that one feature is always perfectly predictive of group membership.

Curriculum Vitae

Name:	Bailey Brashears
Post-secondary Education and Degrees:	Rice University Houston, Texas, USA 2013-2017 B.A. of Psychology, B.A. of Cognitive Science
Honours and Awards:	BMI International Student Scholarship 2017-2019
Related Work Experience	Teaching Assistant Computer Science Rice University 2015-2017
	Teaching Assistant Cognitive Psychology, Developmental Psychology The University of Western Ontario 2017-2019

Publications:

Brashears, B., Brashears, M., Miller, M., Ireland, M., & Davis, T. (2017, January). Is Ice Cream Less of an Animal Product than Hot Dogs?: The Role of Categorization in Judgments of Food Contamination Risk. *Manuscript in preparation.*

Minda, J. P., Brashears, B., & Hatherley, J. J. (2018, June 12). Prior Verbal Description of Novel Stimuli Improves the Learning of Information Integration Categories but not Rule-Described Categories. https://doi.org/10.31234/osf.io/mnefc. *Manuscript in preparation.*