The Limitations and Extent of Category Generalization

Within a Partially Learned Hierarchical Structure

by

Matthew E. Lancaster

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

Approved July 2013 by the
Graduate Supervisory Committee:

Donald Homa, Chair
Arthur Glenberg
Michelene Chi
Gene Brewer

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

Most people are experts in some area of information; however, they may not be knowledgeable about other closely related areas. How knowledge is generalized to hierarchically related categories was explored. Past work has found little to no generalization to categories closely related to learned categories. These results do not fit well with other work focusing on attention during and after category learning. The current work attempted to merge these two areas of by creating a category structure with the best chance to detect generalization. Participants learned order level bird categories and family level wading bird categories. Then participants completed multiple measures to test generalization to old wading bird categories, new wading bird categories, owl and raptor categories, and lizard categories. As expected, the generalization measures converged on a single overall pattern of generalization. No generalization was found, except for already learned categories. This pattern fits well with past work on generalization within a hierarchy, but do not fit well with theories of dimensional attention. Reasons why these findings do not match are discussed, as well as directions for future research.

ACKNOWLEDGEMENTS

I would like to thank my chair and mentor, Don Homa, for his guidance and mentorship in countless projects, including this one, and for his encouragement for me to always do my best work. I would also like to thank the members of my committee, Micki Chi, Art Glenberg, and Gene Brewer and for their insightful comments on theory and methodology. I would like to thank Michael Hout for his help with programming. Lastly, I would also like to thank Jennifer Lancaster for her help in creating an extraordinary number of complex stimuli to help create a large and complete hierarchy.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

The Limitations and Extent of Category Generalization

Within a Partially Learned Hierarchical Structure

An expert is typically a person who has acquired remarkable skill or knowledge in a particular domain, usually acquired only following extensive training. While most people tend to think of experts as rare individuals, the truth is that most people are experts in several areas. For example, all humans become experts at identifying and recognizing human faces (Pascalis, de Haan, & Nelson, 2002; Richler, Wong, & Gauthier, 2011; Tanaka & Gauthier, 1997). Most occupations also require some degree of expertise, having complex decision trees and requiring detailed knowledge (Schraagen, 2006). For example, contractors and construction workers must be knowledgeable about the nature of materials, methods of assembly, building fabrication, and delivery systems. Medical professionals become experts in diagnosing diseases or recognizing problems on radiological images (Brooks, Norman, & Allen, 1991; Patel, Kaufman, & Magder, 1996). Even domains like chick sexing require impressive amounts of expertise (Biederman & Shiffrar, 1987). Expertise can also describe physical skills such as athletic performance, musical skill, or dancing (e.g. Hodges, Starkes, & MacMahon, 2006). Despite these wide ranging domains of expertise, there is at least one common aspect to all areas of expertise. Experts must make fine-grained distinctions between very similar stimuli that non-experts would not differentiate. The ability to categorize at these lower levels of a hierarchy, instead of at the higher levels where non-experts typically categorize, is one of the key defining features of an expert.

We know an impressive amount of how expertise develops (e.g. Ericsson, 1996; Ericsson, 2006) and how experts differ from novices (e.g. Chi, Glaser, & Feltovitch, 1981; Tanaka & Taylor, 1991). For example, Ericsson (2006) notes that the best musicians had practiced for over 10,000 hours, and did not just practice, but deliberately practiced, using the best techniques for improving performance. Chess experts differ from novices in a number of ways, one of which is that experts can remember the placement of more pieces on a chess board after a brief exposure (see Gobet & Charness, 2006, for a review). One of the primary conclusions about expertise is that it is domain limited (Chi, 2006), although the reasons behind strict domain limitations remain unclear.

While there has been a large amount of research focusing on expertise, one aspect of expertise and category learning has been under explored. This area is how much and how people extend their knowledge of learned categories to related, but unlearned categories. Imagine a radiologist who has become an expert at identifying bone cancer tumors in x-rays. Would that radiologist then be better, than a novice, at identifying or faster to learn to identify other types of cancer, e.g. lung cancer; or better at identifying other non cancerous growths such as ovarian cysts.

The present work addresses the nature of expertise for natural categories and whether its knowledge can be applied to related areas that have not been previously learned. Natural categories have typically been classified into a hierarchy where the basic level (e.g. trees or dogs) is the level where non-experts categorize. Subordinate and superordinate categories are below and above the basic level (Rosch, Mervis, Gray, Johnson, & Boyes-Bream, 1976**).** Most researchers (e.g. Tanaka & Taylor, 1991) have defined experts as masters of a basic level domain, e.g. dog experts know a great deal about dogs and about different kinds of dogs. However, work by Tanaka, Curran, and Sheinberg (2005) has suggested that generalization is limited to within what is typically considered a subordinate level category. For example when learning about wading bird species, the knowledge attained is only applied to other wading bird species, but not other types of bird species, such as owl species. The reason for this limitation is not known, nor is it known if other measures of generalization would converge to the same conclusion. The current experiments attempt to answer these questions.

The literature review below provides an overview of relevant research in three inter-related areas relevant to the current proposal - categorical expertise, hierarchical representation and perceptual categorization. This is followed by a concise synthesis of relevant work and rationalization for the current study.

Categorical Expertise

Hierarchical structure. Most natural categories can be organized into hierarchical trees where smaller categorical distinctions are combined into larger and larger categories (see Figure 2 for an example). Classically, levels in hierarchies have been labeled as basic, subordinate, or superordinate levels (Rosch et al., 1976; Corter & Gluck, 1992). The basic level is usually defined as the entry level of a hierarchy for non-experts, typically associated with a general shape, e.g., a generic bird or chair. Superordinate levels are more inclusive categories higher in a hierarchy. Subordinate levels are subdivisions of basic categories and are lower in the hierarchy than the basic level. However, as Coley, Hayes, Lawson, and Moloney (2004**)** note, the basic level of different hierarchies are not necessarily at the same taxonomic level. For example, the dog and bird categories are considered basic categories, but they are at very different levels of scientific taxonomies (Coley, et al., 2004**;** Tanaka & Taylor, 1991). Because of this a number of researchers have explored why the basic level is the typical entry level for non-experts.

Using a category verification task, Rosch et al. (1976) showed that people first classify items at the basic level. In a category verification task, a category is named and then an item is shown. The participant must respond as quickly as possible about whether the item is a member of the named category or not. Rosch et al. (1976) found that the level they considered basic produced the fastest verification response times, suggesting that participants first classified items at the basic level.

To explain why basic level categories are the entry level categories, Rosch et al. (1976) proposed the cue validity hypothesis. According to Rosch et al., the basic level is the most inclusive level in a hierarchy that still maintains "the correlational nature of the environment" (pp. 385). To provide evidence for this account they had participants list attributes of categories at different levels of different hierarchies. For example, one nonbiological category used, was the superordinate group musical instruments with basic level categories like guitar, and subordinate level categories like folk guitar. For nonbiological categories the basic and subordinate levels had approximately the same number of common features, but the superordinate group had fewer common attributes, an outcome consistent with the cue validity hypothesis, which states that the basic level is the most inclusive category maximizes the number of common features. However, for the biological hierarchies, a similar number of common features were found for all the hierarchical levels. Rosch et al. (1976) argued that this pattern was found because the level they listed as superordinate was, in fact, the actual basic level for those hierarchies. Evidence from Coley et al. (2004) and Tversky & Hemenway (1984) support this claim.

Murphy and Smith (1982) noted that in addition to containing the most distinctive attributes, basic level categories also are more common, learned earlier, and have shorter and more distinctive names than other level categories. They taught people an artificial category hierarchy, and found that it was the distinctive attributes that drove the basic level distinction, not those other factors.

One complication is that different types of categories may vary in the depth and size of the hierarchy. For example furniture has relatively few levels, while, as Medin and colleagues (Bailenson, Shum, Atran, Medin, & Coley, 2002;**,** Medin, Lynch, Coley, & Atran, 1997) show, other categories such as birds or trees can have a high number of levels clearly divided into folk hierarchies. While folk taxonomies can vary in structure (e.g. see Bailenson et al. (2002) which are discussed in more detail in the hierarchical representation section below) even non-expert classifications roughly correspond to classic levels in taxonomies (e.g. genera, families, and orders).

Corter and Gluck (1992) argue that cue validity is only part of the reason that determines the entry level of categorization. They also argue that category validity plays an important role in determining the basic level. Category validity states that people will use the hierarchical level that provides the highest confidence in feature inference. For example given the feature, can fly, people would choose the category robin instead of bird because all robins can fly while some birds cannot. Cue validity pushes category use towards higher levels of the hierarchy, because higher levels are inclusive of lower levels there are at least as many members of the higher level category with a feature as the lower level category (e.g. there are more animals that have wings than birds that have wings because animals include birds), while category validity encourages use of the lowest levels of a hierarchy. Corter and Gluck (1992) argue the basic category level is optimal compromise between these two opposing forces. Coley et al. (2004) provide results that show that these two operations do, in fact, provide different best levels in a hierarchy. When participants listed features, the largest gain in features was at the life form level (e.g. bird); however, when completing feature induction tasks, the largest increase in induction power was at the folk generic level (e.g. eagle). Coley et al. (2004) argue that cue and category validity do not compete against each other to create a single basic level, but that the goal of the categorization plays a role in what hierarchical level is used.

This idea, and, more generally, all the work focusing on categorization level, illuminates two questions about expertise. The first is how do experts, and their goals in categorization, differ from non-experts? The second is how do experts represent category hierarchies? These two questions are addressed in each of the next three sections.

Categorization. One of the primary distinctions between experts' and non-experts' categorization, is the knowledge used in making category decisions. Chi et al. (1981) analyzed the differences between physics experts and novices in how they classify physics problems. In their first study, participants were asked to sort 24 physics problems into however many groups they thought were appropriate. While experts took longer to make their sorting decisions, both novices and experts created an equal number of groups. The important difference, however, was the groups there were created and why they were created. Novices tended to group problems based on surface features (e.g. problems that included an inclined plane were grouped together), while experts tended to sort based on deeper structures to the problems, such as the physics principle necessary to solve the problem. A second study conducted by Chi et al. (1981), replicated this finding. They created problems that specifically had similar surface features, but differed in the law that was to be used to solve the problem. They found the same pattern of sorting found in the first study.

In another experiment, Chi et al. (1981) had experts and novices elaborate on category labels that were previously identified. They found that when given a category label, novices described surface features that could be useful in solving the problem (e.g. given the label “inclined plane” novices might identify aspects such as length or incline angle). When an expert is given the same label, they begin by identifying mechanical principles that may be useful in solving the problem. They then continued on to identify surface features that may be important. These studies indicate that experts’ knowledge is both greater and of a different kind than novices’ knowledge.

More evidence of this can be seen by exploring children who are becoming experts. Johnson and Mervis (1994) taught young children about shorebirds over several sessions of playing a fun game. They compared how children made grouping decisions before and after learning about the birds. They found that even though the number of decisions based on attributional properties increased after learning, the vast majority of decisions were still based on morphological similarity. While this differs from the work of Chi et al. (1981), there are several reasons this could be so. First, the children in the Johnson and Mervis (1994) study had relatively little learning. Second, as Gobbo and Chi (1986) note, morphological features and more abstract features are often highly correlated. When looking at children who are more experienced in their expertise domain, a different picture emerges.

Gobbo and Chi (1986) showed young children, who were either relative experts or novices about dinosaurs, dinosaur pictures and asked them to identify facts about the dinosaurs. Novices and experts identified approximately the same number of explicit propositions (facts observable in the pictures), but experts identified many more implicit propositions (facts that were not directly observable from the pictures). This held true even for the dinosaurs that were unfamiliar to the experts. This shows that the experts could use explicit cues to infer implicit cues. Gobbo and Chi's (1986) analysis of the connections between the children's propositions point to exactly this fact. Novices' often listed features with no transition between them, while experts often transitioned from perceptual features to explain what that feature's purpose was. Gobbo and Chi (1986) argue that these findings indicate that not only do the experts have more knowledge than the novices, but that the knowledge is also more integrated.

Chi and Koeske (1983) provide more evidence of this by having a young dinosaur expert play a game that elicited his knowledge about dinosaurs. The researchers then used this knowledge to construct a representation of the child's dinosaur knowledge. They made two mappings; one which consisted of dinosaurs the child experienced more often and a second mapping that consisted of dinosaurs the child experienced less often. The key difference between the two types of dinosaurs was not the number of knowledge nodes, but the number of links to those nodes and more importantly the number of links between dinosaurs. There were more links in the mapping of the more experienced dinosaurs, indicating a more cohesive and integrated representation.

It appears clear that experts have more knowledge, more abstract knowledge, and a more integrated representation of that knowledge than do novices. It is important to discuss how these differences affect experts' hierarchical categorization. Tanaka and Taylor (1991) explored the question of whether experts categorize basic and subordinate level categories differently than novices. They looked at dog and bird experts and had them classify both dogs and birds. Thus, the bird experts acted as novices for dog categorization and vice versa. In Tanaka and Taylor's first study, they had the experts list features for the subordinate, basic, and superordinate levels of both their expert and novice domain. Similar to Rosch et al.'s (1976) finding, the number of new features increased from the superordinate to basic level, however, the differences between the basic and subordinate level was based on the domain. For the expert domain, there were approximately an equal number of new features at the subordinate level as there were at the basic level. In the novice domain there was a drop in the number of new features identified at the subordinate level. Johnson and Mervis (1997), using a similar methodology, found the same pattern of results, but also found that most of the new features at the basic level were physical, while most of the new features at the subordinate level were behavioral in nature.

In a second study by Johnson and Mervis (1997), bird experts typically name birds at a subordinate level rather than at the basic level. Tanaka and Taylor (1991) also found the same pattern of results for bird experts, however, it is important to note that this pattern did not hold true for dog experts. Tanaka and Taylor (1991) and Johnson and Mervis (1997) both found, using a category verification task similar to Rosch et al. (1976), that for the experts' domain the subordinate level category was verified as quickly as the basic level category (although the superordinate level category was still slower). These results indicate that the knowledge gained through expertise acquisition shifts the entry level for categorization to a lower level on the hierarchical tree. It also seems that the same does not hold true for moving up the hierarchical tree. Despite these changes to hierarchical classification and enhancements of subordinate level categories, there are some limitations to how this increase in knowledge can be utilized.

Hierarchical generalization. There appears to be a limit to what the knowledge of expertise can be applied to. Chi (1997) argued that learning should not generalize across an ontological barrier. Most would also agree that learning would not generalize across superordinate categories (e.g. animals and artifacts), and transfer across basic level categories (e.g. birds and reptiles) may be unlikely. However; there is evidence that generalization is even more limited than that.

Although Gobbo and Chi (1986) found that child dinosaur experts inferred properties about unknown dinosaurs from the knowledge they had about familiar dinosaurs, this was not strictly the case in the child studied by Chi and Koeske (1983). Despite forcing an equal number of knowledge nodes during analysis, the child's conceptual map for the less well-known dinosaurs was less developed. Memory for dinosaurs was also tested at the time of the conceptual mapping and a year later. While memory for the less well known dinosaurs was lower at the first time of test, a year later the child had retained the memory for the well know dinosaurs, but memory for the unknown dinosaurs had decreased.

Diamond and Carey (1986) also found limited generalization for dog categories unfamiliar to dog experts. To do this, they used effects found in facial recognition experts to explore dog expertise. Recognition for faces is decremented when the faces are presented upside down, inverted. They hypothesized that dog experts would show the same effect for inverted pictures of dogs. They failed to find an expected strong inversion effect for dog pictures. However, when the experiment was replicated with the limitation that only breeds that the experts had personal experience with were used, the expected inversion effect was found; recognition was slower when the images were upside down. In a similar finding, Johnson (2001) studied what affected bird experts ratings of category typicality. For novices, the central tendency (the average values for the relevant dimensions) was key to identifying which birds were considered the best exemplars of the category. For experts, subject familiarity was more important in determining category typicality. However, when the experts were given unfamiliar species of birds, they acted in much the same way as novices, and central tendency played a more important role in how typicality was rated.

Brooks, et al. (1991) were interested in medical experts' ability to classify diseases that had a similar or different presentation than training. Family practitioners and first year residents were shown skin lesions and diagnoses. During a second phase, they were shown skin lesions that were the same as those shown before, similar to those shown before (and in the same category), different from the old exemplars (but in the same category), and unrelated lesions. Participants were asked to categorize these. Brooks et al.'s (1991) key finding was that both experts and intermediate experts were worse at categorizing the different exemplars than the similar exemplars of the same category. This indicates that the learning of those lesion categories did not transfer throughout the breadth of the group.

The work described above indicates that experts often fail at generalizing their knowledge to even closely related categories and instances. Researchers have explored this failure to generalize. Tanaka et al. (2005) taught novices two levels of a bird hierarchy. Using several training tasks over several sessions, participants learned to distinguish wading birds and owls (the family level) and learned to distinguish one of the families at the species level (e.g. eastern screech owl). The training tasks included naming tasks, where a picture was presented and the correct response had to be entered, a category verification task, and an object classification task, where a label was given and the participant chose the correct bird from two choices. Both before and after the training sessions, participants completed a species discrimination task where participants were shown exemplars sequentially and then asked if they were the same or different species. In the discrimination task there were three types of items: old exemplars seen during training, new exemplars from species that had been seen during training, and new exemplars from new species. Sensitivity to species distinctions increased based on training type and item type. Discrimination for new exemplars (regardless of species) increased with species training as compared to family training; however, the exemplars from old species had higher sensitivity than those from the new species. Tanaka et al. (2005) also found that post training discrimination was more sensitive at the species level than pre-training discrimination, even for the species that were only trained at the family level (although sensitivity for the species trained level were much higher). However, the pre-training discrimination test consisted only of exemplars that were used in training and not of any new species that were untrained, so it is unclear whether sensitivity would have increased for the new species tested during the post training discrimination test.

Scott, Tanaka, Sheinberg, and Curran (2008) conducted a conceptual replication of Tanaka et al. (2005), using modern and classic cars as the basic level and models as the subordinate level categories. They found that from pre- to post training there was a small increase in sensitivity for new exemplars from trained models, but no increase for exemplars from untrained models, even when the untrained models were from the basic category that had received subordinate training. Using a similar procedure, except using experts, Bukach, Phillips, and Gauthier (2010) found that modern car experts do not generalize their knowledge to classic cars. Novices who were the same age as the modern car experts were also more sensitive to modern cars; however, older novices did not show a difference between modern and classic cars. The authors hypothesize that this is the case because older novices have had some experience with classic cars.

The work described above indicates that little of the knowledge gained through learning is transferred to new categories and that there must be some exposure to those categories prior to transfer. However, Lancaster and Homa (2012) showed that at least some knowledge can be used to categorize related but unexperienced categories. In one experiment subjects were taught to differentiate between bird families (e.g. herons and cranes). They then tested categorization performance for trained exemplars, new exemplars from species that were used during training, and exemplars from new species that were still within the learned families (e.g. a blue crane). They found that while performance for the new species was the lowest of the three types of items, participants were still well over chance in classifying the new species. A second study taught participants to differentiate between three species of a single family (e.g. a sandhill crane and a whooping crane). The participants were then tested using training exemplars, new exemplars of the training species, a new species within the learned family, and new species outside the learned family. The relevant finding here is that for the new species in the learned family items performance was poor, but better than chance. It is also important to note that the errors made on those items tended to err toward the new family response. This indicates that there was a strong representation of the learned species, but a weaker representation of the family level category; however, it is remarkable that there was any representation of the family category given that no family level categorization occurred.

Quinn and Tanaka (2007) explored the development of subordinate level representations in infants. Their procedure consisted of two phases; in the first phase infants were familiarized with a breed of dog (or cat); in the second phase the infants were familiarized with either another breed of dog or were switched to a breed of cat. This was followed by a looking preference test consisting of the breed just seen and the corresponding same basic category breed. The results indicated that when the infants had been familiarized with both subordinate level categories from within the same basic level category, they differentiated the breeds. This was not the case when the infants had been familiarized with both a breed of cat and dog. The authors concluded that to have a firm representation of a subordinate level category, there must be a clear comparison group. This appears to be the case, since the infants that only learned one subordinate level category generalized that learning to a basic level category. However, there is one question that this work does not clearly answer. Since the preference test consisted of two breeds that had been experienced before, it is unclear if that was a condition for the differentiation or if a third breed had been used as a novelty preference would the preference had remained. Or to put it more succinctly, was the differentiation dependent on a comparison to a previously learned category, or once a category had been differentiated, would it remain so for any subordinate level comparison. This question is key to the larger issue of whether specific experience is required for any learning to transfer.

Given this work and the other research described in this section it seems clear that some limited knowledge is transferred to some closely related subordinate categories. It is unclear what drives this generalization and what its limits are. Understanding how experts represent hierarchical categories may give some insight into these questions.

Hierarchical representation. Despite the fact that aspects of expertise have been carefully investigated, relatively little research has been conducted explicitly exploring experts hierarchical representations. Chi, Glaser, and Rees (1982) looked at the sorting solutions of physics experts and novices. They found that experts had well defined hierarchical representations that consisted of multiple levels and clear reasoning for the experts' grouping. Novices did not have well defined levels, and some failed to define any levels, refusing to combine or separate any of the categories created during the initial sort. Interestingly, the novices' initial sorting appeared to conform to the subordinate level sorting of the experts. There could be two possible explanations for this. The first is that previous work (e.g. Rosch et al., 1976) is wrong that the entry level for novices would always be the basic level. The alternative explanation is that as the physics experts became experts, their entry level for categorization went up a level in the hierarchy and was, in fact, a superordinate level. Deneault and Richard (2005) provide evidence that children are more successful at class inclusion questions that involve superordinate level categories than problems that do not. This implies that superordinate categories may play a more important role than previously thought. However, this idea is in contrast to other work (e.g. Tanaka & Taylor, 1991 and Johnson & Mervis, 1997) that indicates that the entry level for experts becomes a subordinate level. Work conducted by Griffee and Dougher (2002) and Saiki (1998) seems to indicate that knowledge learned at higher levels of a hierarchy generalize down the hierarchy, but knowledge does not seem to generalize up the hierarchy. However, the type of information used to make decisions may play a role in creating experts' psychological representations. As previously described, work by Chi and colleagues (Chi et al., 1981 and Gobbo & Chi, 1986) identified that novices use primarily surface or perceptual features, whereas, experts use both perceptual features and more abstract features.

Work by Medin and colleagues (Bailenson et al., 2002; Medin et al., 1997) have explored the differences in hierarchical representations between different types of experts. Bailenson et al. (2002) explored the representation of bird hierarchies in American experts and novices and members of a Central American tribe (the Itza' Maya). The authors tested the participants' representations using a number of methods; sorting, "goodness of example" typicality ratings, and category based induction. The results indicated that for the two types of experts, U.S. and Itza, the hierarchical representations largely conformed to bird taxonomies. However, there were two key cultural differences. For U.S. participants, experts and novices, passerines (i.e. songbirds) were central to the category birds. This was not true for the Itza, indicating that non-perceptual dimensions can change the organization of representations. Bailenson et al. (2002) also found that while both types of experts' representations resembled taxonomies, the reasons for the distinctions often differed. For example, the U.S. experts often gave geographical range or evolutionary age reasons to justify decisions, while the Itza often gave ecological reasons (e.g. diet) to justify the same decisions. This finding is consistent with other research discussed above where features are often highly inter-correlated.

Medin et al. (1997) also explored how the goals of expects influenced perceptual and non-perceptual similarity. Three types of tree experts, taxonomists, landscapers, and maintenance workers, completed a sorting procedure. Results showed that landscapers' and maintenance workers' hierarchies had moderate correlations with scientific taxonomic classifications. The only moderate correlation indicates that they had other, non-taxonomic reasons for organizing categories. This assertion is further supported by the justifications they gave for their initial sort. Taxonomists' reasons were almost exclusively taxonomic in nature, while landscapers and maintenance workers gave a wide variety of other reasons, such as its utility, its nativeness to the area, or a weed category. These differences led to differences in the overall hierarchical structure. Maintenance workers had more categories in their initial sort with relatively fewer subordinate and superordinate categories, which is similar to the novices from Chi et al. (1982). This may indicate that superordinate or subordinate grouping may not be particularly useful for maintenance workers goals. For the landscapers, the folk distance between items in the lower scientific ranks (e.g. genus level) were similar to the taxonomists. The taxonomists' distances continue to increase at each higher level of scientific rank, while the landscapers' distances leveled off midway through the scientific ranks. This indicates that the more subordinate levels of scientific taxonomy were useful to the landscapers, but the higher subordinate levels stopped being of use and so were no further differentiated. This work by Medin and colleagues indicates that both perceptual features and non-perceptual features play a role in the hierarchical representations of experts, and, while often correlated, likely contribute their own unique variance to differences in organization. If this is true, then it is also likely the case that they both also contribute to generalization within a hierarchy.

Most of the works described here have used sorting procedures to identify hierarchical representations; however, there are other methods that may reveal this structure as well. While not explicitly testing for this, Homa and Silver (1976) revealed that multidimensional scaling could reveal the organization of subordinate divisions within basic level categories. They showed participants word triads and varied the psychological distance between the words; short within category distance, long within category distance, or between category distance. Participants had to respond to whether all the words in the triad were in the same category or not. Although not specifically analyzed, it appears that within a category subordinate level category members are clustered together; for example, in the birds category all the songbirds are grouped together and all the birds of prey are grouped together. Thus, the multidimensional scaling reveals the hierarchical structure of the psychological space. Homa and Silver's (1976) results also corroborate this conclusion. They found that responses to the trials where all the words were within the same subordinate category were faster than when the second word was in a different subordinate category, which were in turn faster than items where one of the words was from a second basic level category. This indicates that when the first word occurred the basic and subordinate levels were activated. If all the words were within that subordinate category, then no further activation was required; however, if the second word was in another subordinate category, then it needed to be activated, which required more time; but not as much time as when a different basic level category had to be activated.

In addition to being able to identify subordinate level categories, multidimensional scaling can show the changes in participants' psychological structure due to learning. Homa, Rhoads, and Chambliss (1979) showed that as more learning occurs, category representations changed in two ways. First, categories become more differentiated, becoming further apart from each other in psychological space. Second, members of each category clustered closely around the category's centroid. While Homa et al. (1979) explored perceptual categories that lacked hierarchical properties, the same changes should occur in hierarchical categories. First, the basic categories should begin to differentiate, followed by the subordinate categories beginning to differentiate from each other within the basic level cluster.

Perceptual Categorization

Categorical variables. There is not much work on what category variables lead to category learning and generalization in hierarchical structures (as mentioned in the sections above). There is, however, a large amount of research focusing on what aspects of categories lead to learning and abstraction of core essences of categories.

While most research (e.g. Ashby & Gott, 1988; Minda & Smith, 2001) on categorization has used only two categories for learning, there is some work on the effect of learning a larger number of categories. Homa and Chambliss (1975) explored the effect on transfer performance after learning two, four, or six different categories. They found that performance for the exemplars shown during learning remained high, regardless of the number of categories learned, but absolute performance for the new exemplars decreased as the number of categories learned. However, it is clear that people know more than six categories. It may be that both the number of categories learned and the timing of when the categories are learned affect generalization performance. Some work has shown that people are able to alter their representation of a learned category, with no further exposure to that category (Homa & Rogers, 2011). Thus, people modified their categorical representation without learning them at the same time. Since the work here focuses on learning multiple categories within a hierarchy, the number of categories learned was a concern; however, the categories were learned a few at a time and there was never a time when there are six response options.

One of the more obvious variables that affects category learning is the number of different exemplars that are shown during learning, often called category size. As work by Homa and colleagues (Homa & Chambliss, 1975; Homa & Cultice, 1984) show, as the number of exemplars shown during learning increases, the ability to generalize learning to other category members increases. While this seem like a rather uninteresting effect, it is important to factor the effect into designs involving hierarchical categories, because subordinate level categories have fewer possible exemplars by definition.

A final key aspect of category structure that determines a person's representation of a category is the variability of the exemplars seen during learning. Homa and Cultice (1984) showed that the amount and type of variability of exemplars has a strong effect on a person's ability to transfer that knowledge. They found that if all of the learning stimuli were low distortions of a prototype, then that knowledge transferred to new low distortions, but it did not transfer to medium or high distortions. Similarly, if a participant was exposed to only high level distortions during learning, then the participant had poor transfer to all types of distortions. The best transfer occurred when medium distortions or distortions of mixed levels were used during learning. The level of exemplar variability is intimately tied to hierarchical structures. Higher level categories inherently have more variability than lower level categories, and the lowest levels of a hierarchy may have very little variation in them.

It is important to note that the above research may not directly apply to the main focus of the current research, since the research here is focused on the transfer of learning from learned categories to new categories, while the above research discusses the transfer of learning to new exemplars within the learned categories. However, these effects are important to consider, because they are the best evidence available on how the category structure may affect transfer.

Another key aspect of perceptual categorization is attention allocation. A large amount of work (e.g. Minda & Smith, 2001; Nosofsky & Zaki, 2002) have shown that people selectively attend, or pay more attention, to some dimensions when learning categories. This weighting, of course, helps people identify new members of already learned categories, but it may also help people learn new categories that share diagnostic dimensions. Relatively little work has been done exploring how selective attention affects learning new categories. Some work by Goldstone (Goldstone & Steyvers, 2001; Kersten, Goldstone, & Schaffert, 1998) has explored this issue. Kersten et al. (1998) discuss an attentional mechanism they call attentional persistence. Attentional persistence is the tendency to continue to pay attention to features or dimensions that have been predictive in the past. Goldstone and Steyvers (2001) showed evidence that people do persist at attending dimensions that have been useful. They had participants learn two categories that varied on two dimensions, one was relevant and one was irrelevant. Then participants learned a second set of categories that also had one relevant and one irrelevant dimension. As predicted by attentional persistence, participants did best at learning the second category set when the relevant dimension was the same in the two category sets, and worst when the irrelevant dimension in the second set had been the relevant dimension in the first set. Thus, when relevant dimensions overlap, generalization should be facilitated.

Current Studies

The primary purpose of the proposed studies is to explore the learning and transfer within hierarchical categories, with a goal to assess the specificity and extent of generalization at increasing organizational distances. It is clear that expertise is domain limited (Chi, 2006); however, the extent of a domain is still unclear. The concept of domain specificity indicates that an expert's knowledge will transfer to nearby categories, but at some distance from previously learned categories, likely where there is another step up the hierarchical tree, transfer will abruptly stop. The evidence that is available points to a lack of or extremely limited transfer within a hierarchical structure (e.g. Scott et al., 2008; Tanaka et al., 2005). This is despite the fact that experts have an abundance of knowledge about the relevant basic level categories and similar subordinate level categories (Chi et al., 1981; Chi et al., 1982; Gobbo & Chi, 1986). However, those studies that show lack of generalization use the lowest subordinate levels (species and car models) and have very low variability of exemplars, which as Homa and Cultice (1984) show is not conducive to generalization or strong categorical representations. Their results indicate that some degree of variability is necessary for robust category representations. Although experts often make species level distinctions, they also make distinctions on multiple subordinate levels that have higher degrees of variability than the species level (Bailenson et al., 2002; Medin et al., 1997). These intermediate, subordinate levels may be key to experts' generalization.

Also, past work has focused on discriminability of categories, specifically the increase in sensitivity to differences after training; however, discriminability is not the only measure of generalization available. While immediate differentiation may not be evident, previous learning may have primed future learning (Schwartz, Bransford, & Sears, 2005). This would be the case if the to be generalized to categories were learned faster after previous learning than absent of previous learning. Another way to measure potential generalization would be to look at the psychological representation of the related subordinate categories using multidimensional scaling. Homa and Silver (1976) has shown that subordinate categories are evident within psychological space, and Homa et al. (1979) has shown that greater differentiation of categories occurs as increased learning occurs. Thus, if unlearned subordinate categories have begun to be differentiated from other subordinate categories, then this would indicate that some previous learning has transferred. These two measures, plus a potentially more sensitive discrimination test, due to a more robust category representation, should provide converging evidence for the extent of domain generalization.

Thus, the current series of experiments attempted to determine the limits of subordinate level category generalization using naturally occurring hierarchical categories (see Figure 2); using category structures and transfer measures that create the best opportunity for generalization. There are several theoretically interesting patterns of generalization that could plausibly occur. Figure 1 conceptually illustrates these outcomes based on the amount of generalization expected at different hierarchical distances from the learned categories.

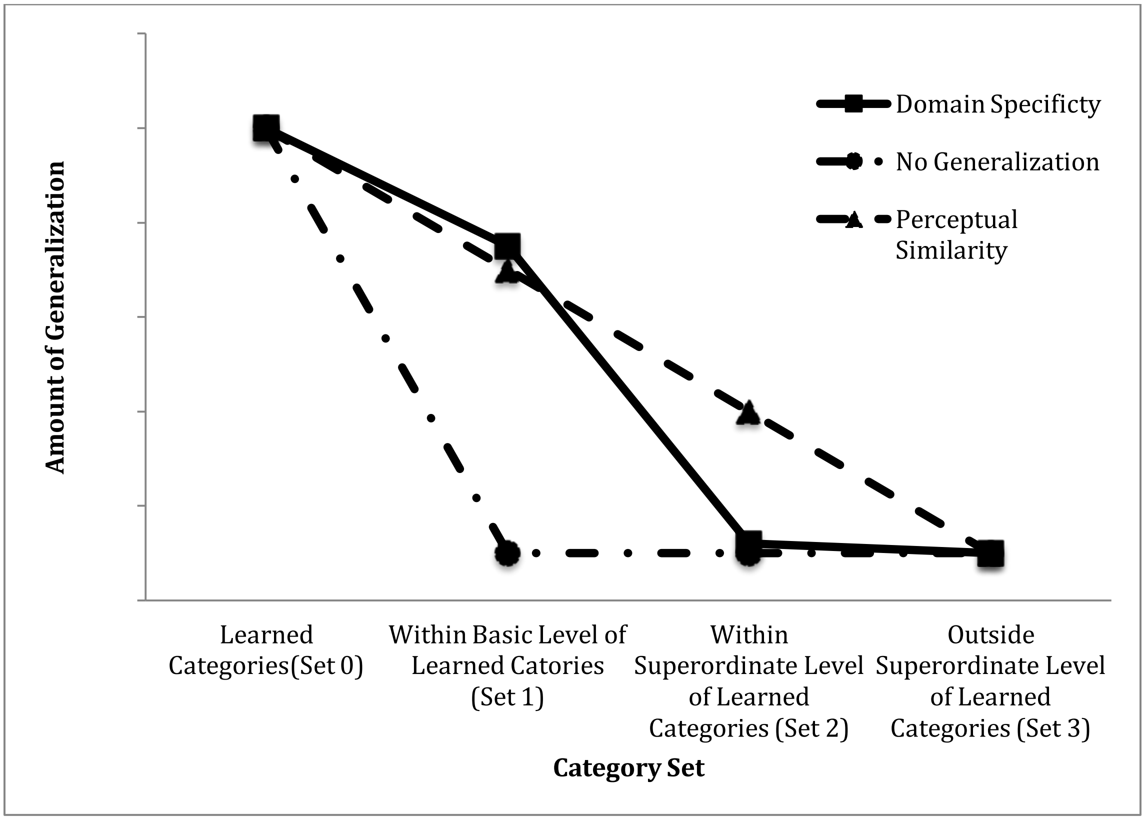


Figure 1. Theoretically plausible outcomes of converging measures of generalization. Set number is distance from learned categories as indicated by Figure 2.

The predicted outcome is a one of limited domain specificity. Here, there is a large amount of generalization to new members of the learned categories, but also a large amount of generalization to unlearned categories within the basic category where subordinate categories were learned, but little to no generalization farther away from the learned categories in the hierarchy. This is expected because the categories within the same basic level likely share some if not all relevant diagnostic dimensions; thus, attentional persistence theory predicts that having learned one set of categories in that basic level category will facilitate generalization to other categories within that higher order level, but not outside it where common relevant dimensions are rare. While domain specificity is predicted, there are other plausible outcomes. One is of no generalization where high performance on the transfer measures is only found in new members of the learned categories. A third plausible outcome is where there is generalization, but it conforms only to perceptual similarity and not any function of hierarchical distance.

Experiment 1

The purpose of experiment 1 is to test the domain specificity, or the limits of generalization, within a partially learned hierarchy. The key question is whether there is a distance from learned categories, where transfer will not occur; and if there is, what is the size of that specificity, or distance. This experiment differs in several key ways from past research (e.g. Tanaka et al., 2005; Scott et al., 2008). Tanaka et al. (2005) taught participants to differentiate at the order level (e.g. wading birds and owls) and at the species level of one type of order (e.g. Great Blue Herons). One difference is that the subordinate level categories are not at the species level, but are instead at the family level of the hierarchy (e.g. cranes instead of blue cranes). Making this change creates subordinate level categories that have more variability in them, which according to Homa and Cultice (1984) create stronger category representations. Other work (Lancaster & Homa, 2012) has shown that species level learning creates asymptotic performance within a few blocks, which indicates little variability within categories. A second key difference in the current study is how generalization will be measured. In addition to measuring discrimination, generalization will also be measured by speed of learning of new categories within the hierarchy. The final key difference was two sets of control categories, one at the same distance as the other learned order category but was not experienced before the transfer tests, and one where no generalization is expected to occur, to act as a baseline comparison condition. Figure 2 shows the kind of hierarchy explored in the present study. As can be seen in Figure 2, the baseline control categories (in this case, types of lizards) are only related to birds at two levels above the basic level.

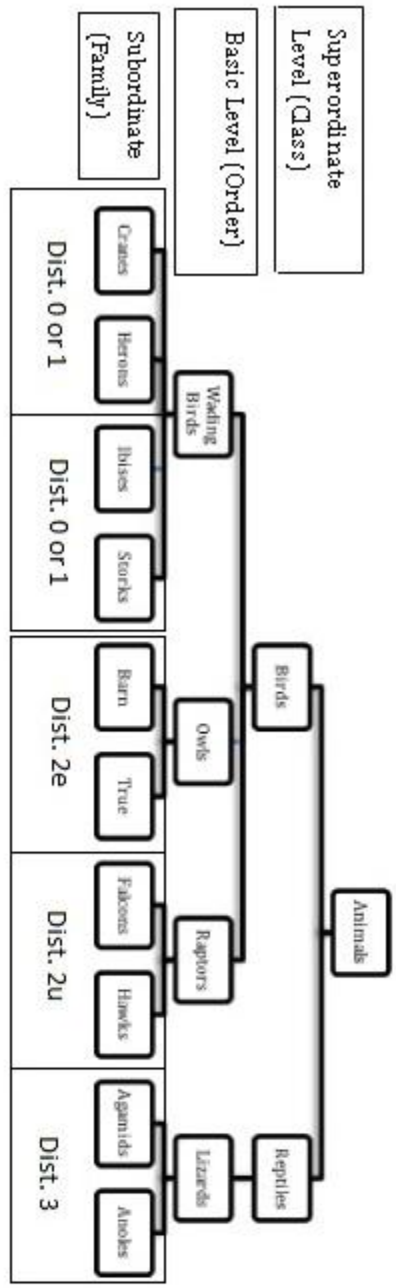


Figure 2. Hierarchical structure used for both experiments. Levels are listed as hierarchical levels and taxonomic levels. Sets are labeled based on hierarchical distance from learned group and previous experience with groups. Set 0 is the learned pair, Set 1 is within the same Order level group as the learned pair, Set 2e is within the same Class level group as the learned pair and had been experienced before, Set 2u is within the same Class level group as the learned pair and had not been experienced before, Set 3 is within the same Kingdom level group as the learned pair.

Method

Participants. One hundred eighty-five participants were recruited from the Introductory Psychology participant pool from Arizona State University. They received partial course credit for participating. Eight participants were removed from analyses because of experimenter or computer error, another 7 were removed from data analyses because they did not reach the learning criteria in phase 1 or phase 2, and 6 did not meet criteria in phase 4 and so were not included in analyses involving phase 4.

Design and materials. This experiment consisted of 15 different conditions split along 2 factorial dimensions. Table 1 illustrates those dimensions and the flow of experiment 1.

Table 1

Experimental conditions organized by experimental phase experience.

Learning Group	Phase 1 Order Level Learning	Phase 2 Family Level Learning	Phase 3 Discrimination Test (Completed all groups)	Phase 4 Learning Test (Completed 1 group)
Heron/Crane (H/C)	Owls/Wading Birds	Herons/Cranes	Heron-Crane Ibis-Stork True-Barn Owl Falcon-Hawk Agamid-Anole	Heron/Crane Ibis/Stork True/Barn Owl Falcon/Hawk Agamids/Anoles
Ibis/Stork (I/S)	Owls/Wading Birds	Ibises/Storks	Heron-Crane Ibis-Stork True-Barn Owl Falcon-Hawk Agamid-Anole	Heron/Crane Ibis/Stork True/Barn Owl Falcon/Hawk Agamids/Anoles
No Learning (NL)	No Learning	No Learning	Heron-Crane Ibis-Stork True-Barn Owl Falcon-Hawk Agamid-Anole	Heron/Crane Ibis/Stork True/Barn Owl Falcon/Hawk Agamids/Anoles

The first dimension, learning group, was split into 3 groups. One group, the No Learning (NL) group, acted as a control and did not complete Phase 1 or 2 of the experiment. The other two groups, the Heron-Crane (H-C) Learning group and the Ibis-Stork (I-S) Learning group, both completed identical Phase 1's. Then, in Phase 2 they learned the appropriate set, e.g. the H-C group learned to distinguish herons and cranes. Phase 3 was the same for all participants. In Phase 4, the learning groups were further split into 5 groups each. Each of these groups participated in a learning test by learning 1 of the 5 sets outlined in Figure 1. For example, participants in one condition learned the H-C group in Phase 2 and then learned the Falcon-Hawk distinction in Phase 4.

All stimuli were shown on windows computers using e-prime software. All the stimuli were colored photographs of bird (or lizard) species at various angles and in various poses (see Figure 4 for examples of stimuli). The stimuli were grouped into a hierarchical structure as depicted in Figure 3. Participants learned only two levels of the structure (Order and Family), and further only learned two categories (in the learning phase) at each level.

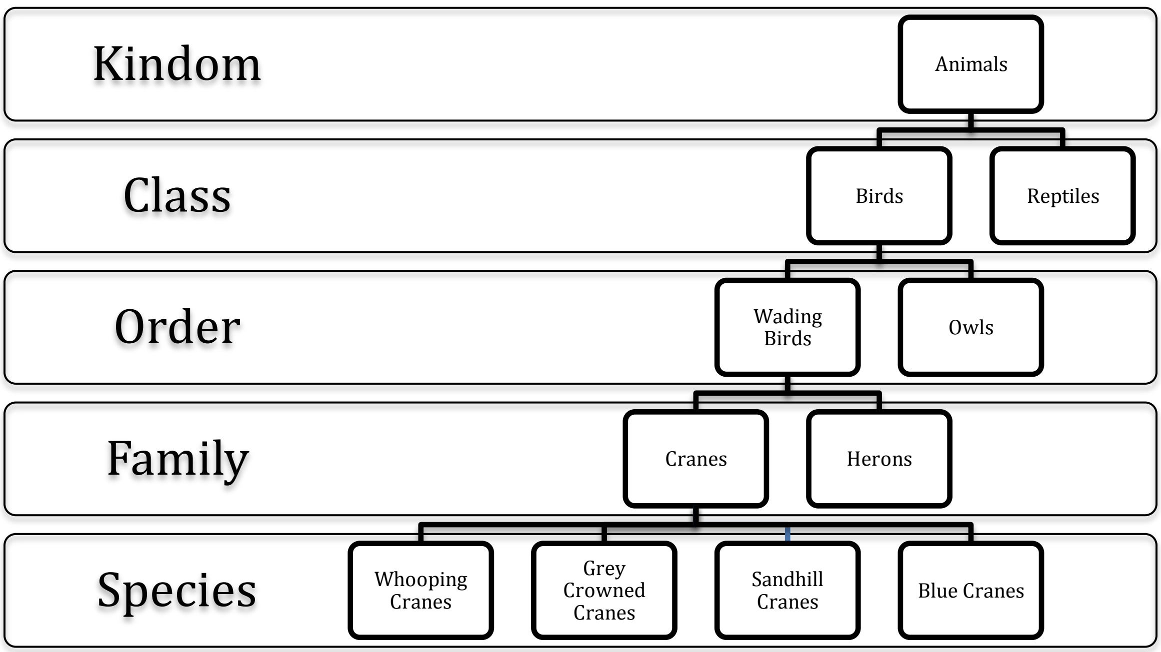


Figure 3. Partial hierarchy, showing 1 full branch. Participants will only explicitly learn at the order and family level. Species will never be distinguished, but will make up the family level stimuli.

The categories were constructed thusly: each wading bird, family level category (e.g. herons) consisted of 44 stimuli, split between 4 species in each group. For an example and as can be seen in Figure 2, the crane category consisted of 11 pictures each of the Grey Crowned Crane, Whooping Crane, Sandhill Crane, and Blue Crane species. Of these, 1 image from each species was used for learning in Phase 1, 4 from each appropriate species were used in Phase 2, 2 from each species were used in Phase 3, and 4 from each species were used during appropriate conditions in Phase 4.


Grey Crowned Crane


Great Blue Heron

Figure 4. Two examples of stimuli to be used in Experiment 1. Stimuli here are named at the species level, but will only be differentiated at the family level (e.g. cranes and herons).

The owl groups consisted of 8 images from each of 4 species from each family level group. Two images were used from each species in Phase 1 and Phase 3, and 4 from each species were used in the appropriate conditions for Phase 4. The images for the raptor and lizard groups followed a similar pattern; 2 from each species were used in Phase 3 and 4 from each species were used in the appropriate conditions for Phase 4. Table 2 illustrates the number of images used in each phase of experiment 1.

Table 2

Number of unique stimuli used in each phase for each family level category as a function of bird species.

		Number of Unique Images			
		Phase 1	Phase 2	Phase 3	Phase 4
Family group (e.g. Heron)	Species 1	1 (2 for Owls)	4	2	4
	Species 2	1 (2 for Owls)	4	2	4
	Species 3	1 (2 for Owls)	4	2	4
	Species 4	1 (2 for Owls)	4	2	4

Note. Images only used where appropriate for experimental design (e.g. Falcon images were only used in phases 3 and 4

It is important to note that no images were reused in other phases of the experiment. For example, for the Heron-Crane learning group, the images seen in Phase 2 were of the same species seen in Phase 1, but were not the exact, same pictures. All the images were collected from past research (e.g. Tanaka, et al., 2005) or from online image searches. Images were present in color on a plain white background (as is shown in Figure 3, no extraneous background information was available).

During the experiment, category names were developed from the appropriate Latin name for the order or family (for example, the Crane family used the name: Gruid). This had the advantage of providing meaningful names (as opposed to group A, etc.) as well as using uncommon names that participants would not recognize. Responses to the stimuli were keyboard presses of a prominent letter of the appropriate Latin name (with the condition that none of the groups used the same response letter). Responses during the discrimination test were 'z' or 'm' for same or different group.

Procedure. At the start of the experiment all participants received instructions describing the outline of the experiment and describing how animals can be grouped into multiple hierarchical groups. Experiment 1 consisted of 4 phases.

In Phase 1, all learning participants learned the basic (order) level categories of owls and wading birds. Each category consisted of 18 images as outlined in the materials section above. They did not learn the group raptors so that it could act as an unexperienced control group to the owl group. Participants learned until they reach a 93% or higher correct criteria on one block. A block consisted of presenting all the stimuli from both categories once in a random order.

After the learning group participants reached criterion on the basic (order) level groups, they completed Phase 2, where they learned two subordinate (family) level groups of wading birds, which are the two categories indicated as set 0 from learned group in Figure 1. The Heron-Crane group learned the set of herons and cranes, while the Ibis-Stork group learned the set of ibises and storks; thus, those respective sets were designated as set 0 for the two learning groups. These were learned to the same criteria as the basic (order) level categories. During these 2 learning phases, each trial consisted of presenting one stimuli in the center of the screen with the two response options listed below the image. Once a participant responded, the response options disappeared and the correct answer appeared for one second. After an inter-trial interval of 500 milliseconds the next trial started. The No Learning group did not experience Phase 1 or 2.

Phase 3 consisted of a discrimination test at the subordinate (family) level. Two stimuli were presented simultaneously with a category label between them. Participants decided whether they belonged to the same or different subordinate (family) level groups. No feedback was provided. Instructions before the phase began emphasized that at least one of the creatures always matched the given label, and it was the participant's task to decide if both were or not. Comparisons were always within a set. For example, a participant saw a comparison between two True Owls or between one True Owl and one Barn Owl, but would have never seen a comparison between a True Owl and a Falcon. There were 20 same and 20 different pairs for each of the 5 sets depicted in Figure 1) for a total of 200 discrimination trials.

Phase 4 was a learning test. A subset of participants from each of the 3 learning groups learned each of the sets described in Figure 1. For example, one group of participants in the Heron-Crane Learning group relearned set 0, or the heron and crane groups. Another group of participants in the Heron-Crane Learning group learned set 2u, the falcon and hawk groups. Phase 4 learning occurred in the same way as Phase 1 and 2 learning, with feedback. Participants learned to a criterion of 85% correct. This criteria, while not as high as initial learning is high enough to show differences in speed of learning. The number of blocks to criterion was the measure of the speed of learning.

After participants finished the transfer learning test, they completed a short questionnaire. The survey asked about the participants general knowledge of birds, then asked participants to list, for each of the categories in each phase where they learned one, the features they used to classify that category. After completing the questionnaire, participants were debriefed, given credit, and thanked for their time.

Results

Analyses were conducted on the learning in phase 1 and 2 to test whether the Heron-Crane group and the Ibis-Stork group learned differently. For Phases 1, 2 and 4, learning stopped when criteria was met. For analyses involving multiple blocks, the accuracy the participant achieved on the block they reached criteria was inputted for all further blocks.

Phase 1. In a 5x2 (Learn Block x Phase 2 Learning Group) Mixed Model ANOVA, there was a significant effect of learning block, $F(4, 416) = 28.478$, $p<.001$, but no effect or interaction involving learning group, p's >.05. Because of the imputation of a participant's final block score on blocks after he or she reached criteria, there was a concern that the assumption of sphericity was violated; so a 2x2 (first 2 learning blocks x Phase 2 Learning Group) Mixed Model ANOVA was conducted. This analysis also showed no main effect or interaction involving learning group, p's >.05. Figure 5 shows the learning across the 5 blocks of Phase 1.

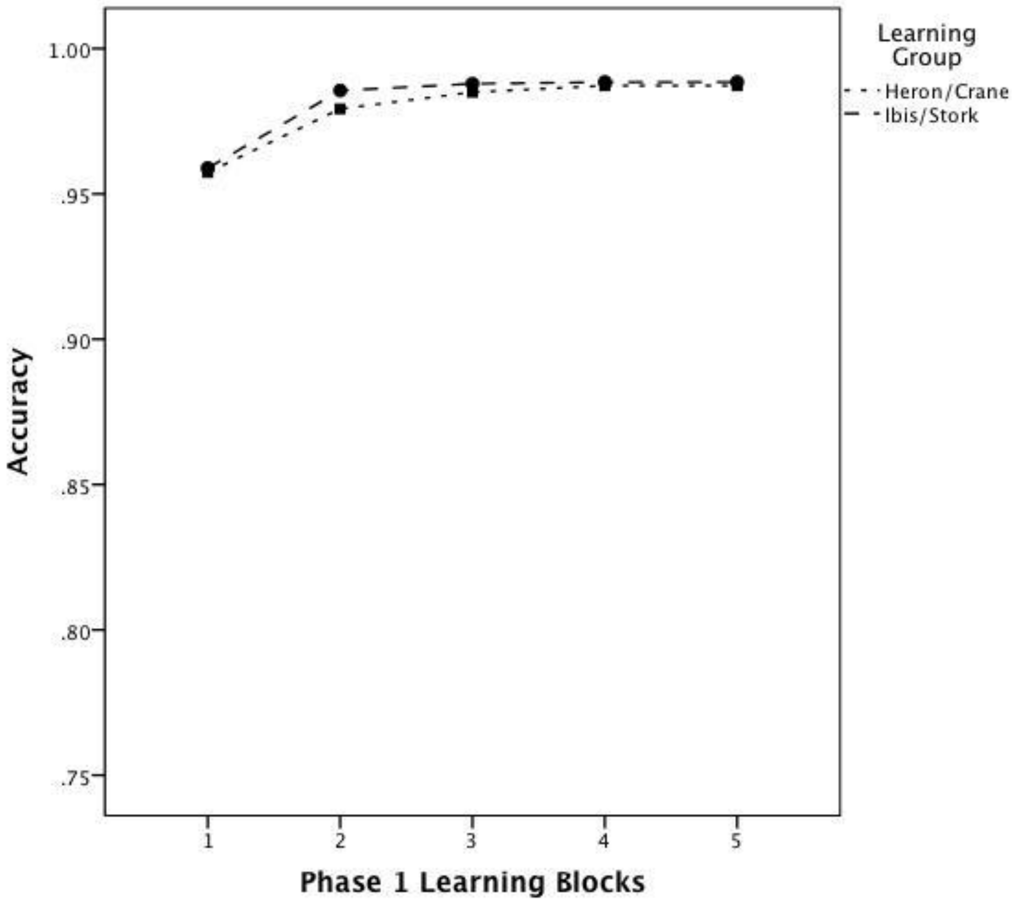


Figure 5. Proportion accuracy across the 5 learning blocks of Phase 1 for both learning groups.

Phase 2. In a 15x2 (Learn Block x Phase 2 Learning Group) Mixed Model ANOVA, there was a significant effect of learning block, $F(14, 1456) = 216.813$, $p<.001$, but no effect or interaction involving learning group, p's > .05. An independent samples t-test also showed that there was no difference between the Heron/Crane learning group ($M=6.963$, $s=2.946$) and the Ibis/Stork learning group ($M=6.192$, $s=3.320$) in the number of blocks required to reach the criteria of 93% correct, $t(104) = 1.265$, $p=.209$. Because of the imputation of a participant's final block score on blocks after he or she reached criteria, there was a concern that the assumption of sphericity was violated; so a 3x2 (first 3 learning blocks x Phase 2 Learning Group) Mixed Model ANOVA was conducted. This analysis also showed no main effect or interaction involving learning group, p's >.05. Figure 6 shows learning across the 15 blocks of Phase 2.

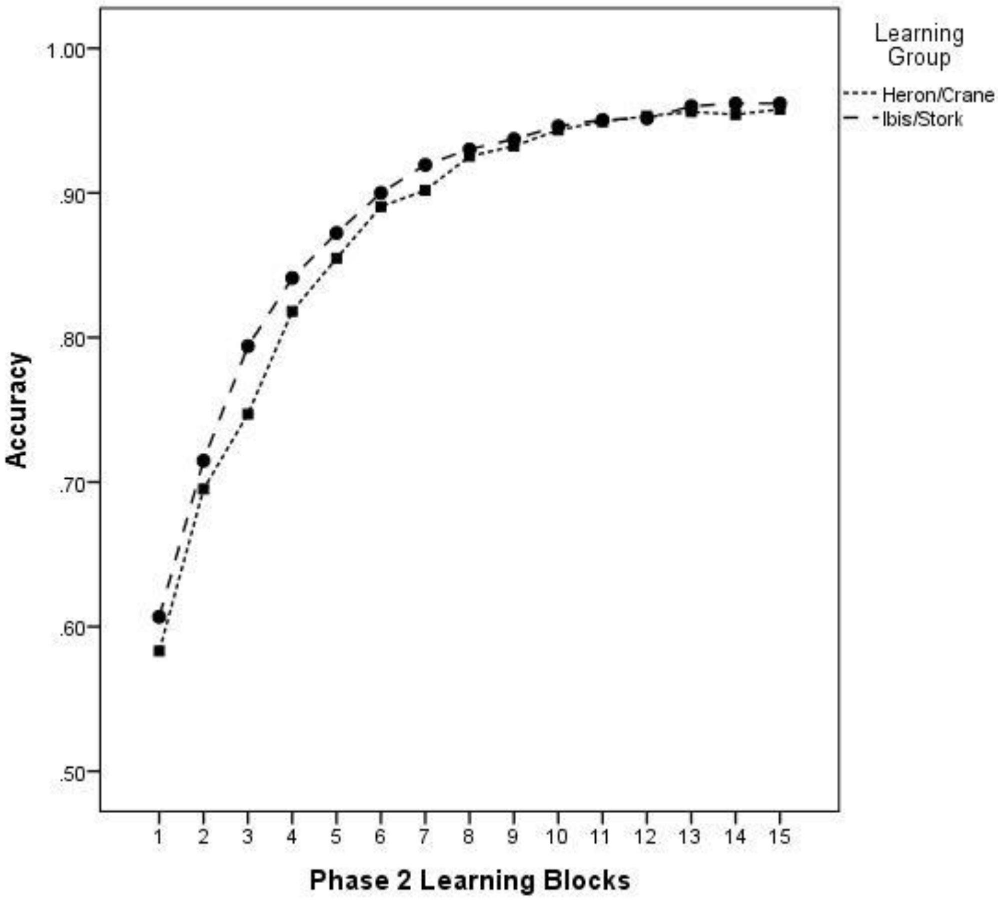


Figure 6. Proportion accuracy across the 15 learning blocks of Phase 1 for both learning groups.

Phase3. Analyses in phase 3 focused on accuracy of each creature set (hits plus correct rejections), sensitivity, and bias. Sensitivity was analyzed using d', which was calculated by subtracting the inverse of the cumulative normal distribution of false alarms from the cumulative normal distribution of hits. Hits were participants correctly responding yes, that images were members of the same family level group. False alarms were participants incorrectly responding yes to the same situation. Table 3 below displays the hit and false alarm rates for each category set.

Table 3

Mean hit and false alarm rates and standard deviations for each category set divided by learning conditions.

Category Sets		Learning Group Heron/Crane	Ibis/Stork	No Learning
Heron-Crane	Hits (SD)	.75(.14)	.73(.19)	.62(.17)
	False Alarms (SD)	.37(.17)	.48(.22)	.40(.21)
Ibis-Stork	Hits (SD)	.71(.22)	.83(.10)	.69(.18)
	False Alarms (SD)	.53(.21)	.35(.16)	.53(.20)
Owls	Hits (SD)	.76(.22)	.77(.18)	.76(.15)
	False Alarms (SD)	.55(.31)	.53(.29)	.76(.15)
Raptors	Hits (SD)	.74(.20)	.71(.19)	.68(.18)
	False Alarms (SD)	.61(.21)	.59(.21)	.54(.21)
Lizards	Hits (SD)	.59(.21)	.62(.18)	.59(.20)
	False Alarms (SD)	.40(.22)	.42(.19)	.39(.22)

Accuracy. In a 3x5 (Learning Condition x Creature Set) Mixed Model ANOVA, there are main effects of Learning Condition, $F(2, 166) = 4.274$, $p=.015$ and Creature Set, $F(4, 664) = 26.184$, $p<.001$. There is also an interaction between the two, $F(8, 664) = 17.930$, $p<.001$.

A priori comparisons using LSD showed that, for the Heron-Crane learning group, the Heron-Crane set was more accurately discriminated than all the other sets, p's < .001. Likewise, for the Ibis-Stork learning group, their accuracy on the Ibis-Stork set was higher than any other set, p's < .001. For the Heron-Crane set, Heron-Crane learners

(M=.70, SE=.013) were significantly more accurate than non-learners (M=.612, SE=.012) or Ibis-Stork learners (M=.629, SE=.013). For the Ibis-Stork set, Ibis-Stork learners (M=.751, SE=.013) were significantly more accurate than non-learners (M=.583, SE=.012) or Heron-Crane learners (M=.600, SE=.013). There were no other significantly differences of learning groups within the different sets.

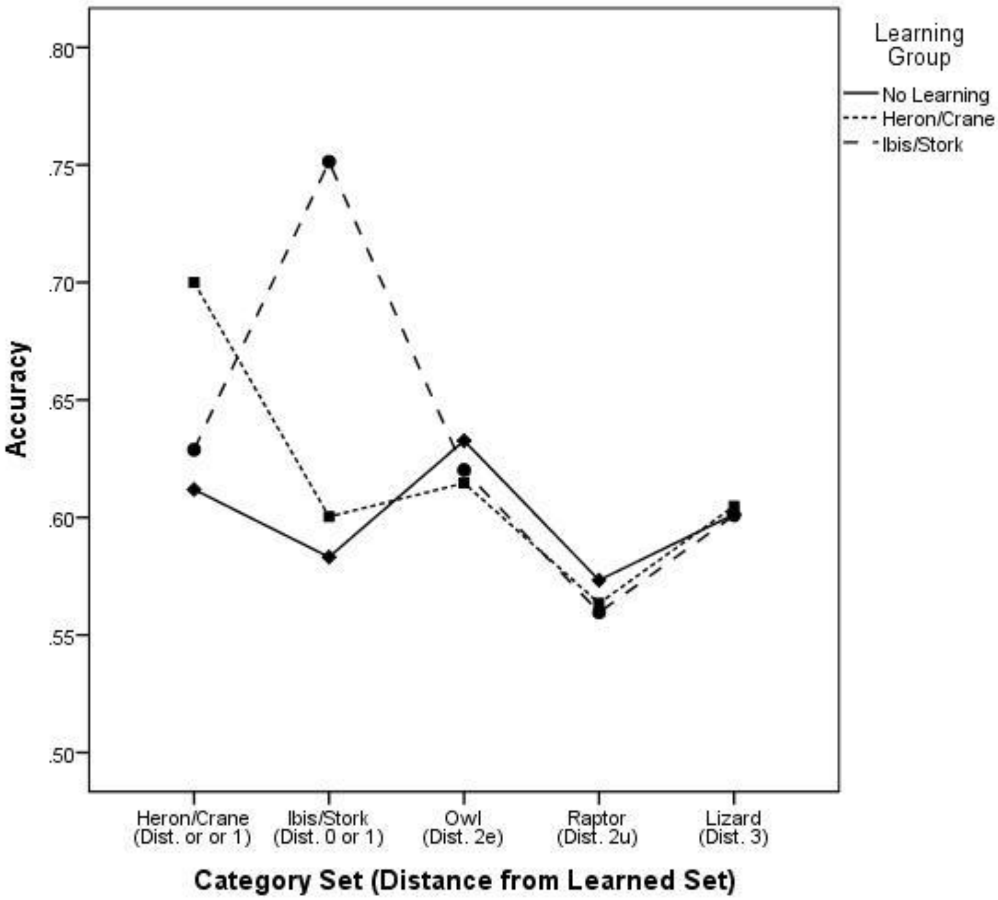


Figure 7. Accuracy of discrimination of creature sets by the learning groups in Phase 3. The only significant differences involved the Heron/Crane and Ibis/Stork sets. Distances from the learned set are described in Figure 2.

Sensitivity. In a 3x5 (Learning Condition x Discrimination Set) Mixed Model ANOVA, there were main effects of Learning Condition, $F(2, 166) = 4.759$, $p=.01$ and

Discrimination Set, $F(4, 664) = 23.392$, $p<.001$. There was also an interaction between the two, $F(8, 664) = 15.198$, $p<.001$.

A priori comparisons using LSD showed that the Heron-Crane learners (M=1.186, SE=.082) were better able to discriminate the groups of the Heron/Crane set than non-learners (M=.655, SE=.75) or Ibis-Stork learners (M=.794, SE=.083). For the Ibis-Stork set, Ibis-Stork learners (M=1.532, SE=.086) had a significantly higher d’ than non-learners (M=.516, SE=.079) or Heron-Crane learners (M=.623, SE=.085). There were no other significantly differences among the learning groups in the different sets. Figure 8 depicts these differences.

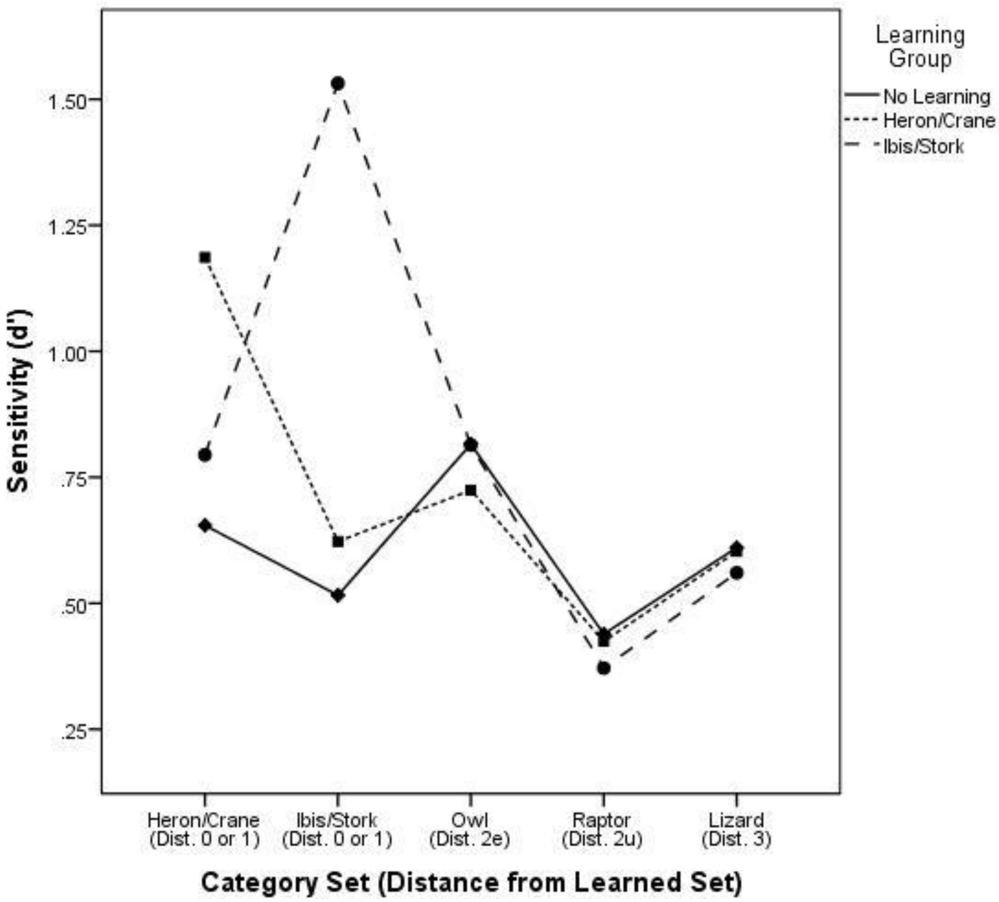


Figure 8. Sensitivity (d') of discrimination of creature sets by the learning groups in Phase 3. The only significant differences involved the Heron/Crane and Ibis/Stork sets. Distances from the learned set are described in Figure 2.

Bias. In a 3x5 (Learning Condition x Discrimination Set) Mixed Model ANOVA, there was a main effect of discrimination set, $F(4, 664) = 30.452$, $p<.001$, but no effect or interaction involving learning condition, p's > .05. Responses to the lizard set were more conservative than all other sets, p's < .001. Responses to the Heron/Crane set were the next most conservative, p's < .001. Responses to the Ibis/Stork set were significantly more conservative than the owl set, $p = .032$, and the raptor set, $p = .022$. The owl and raptor sets were not different from each other. Figure 9 illustrates these effects.

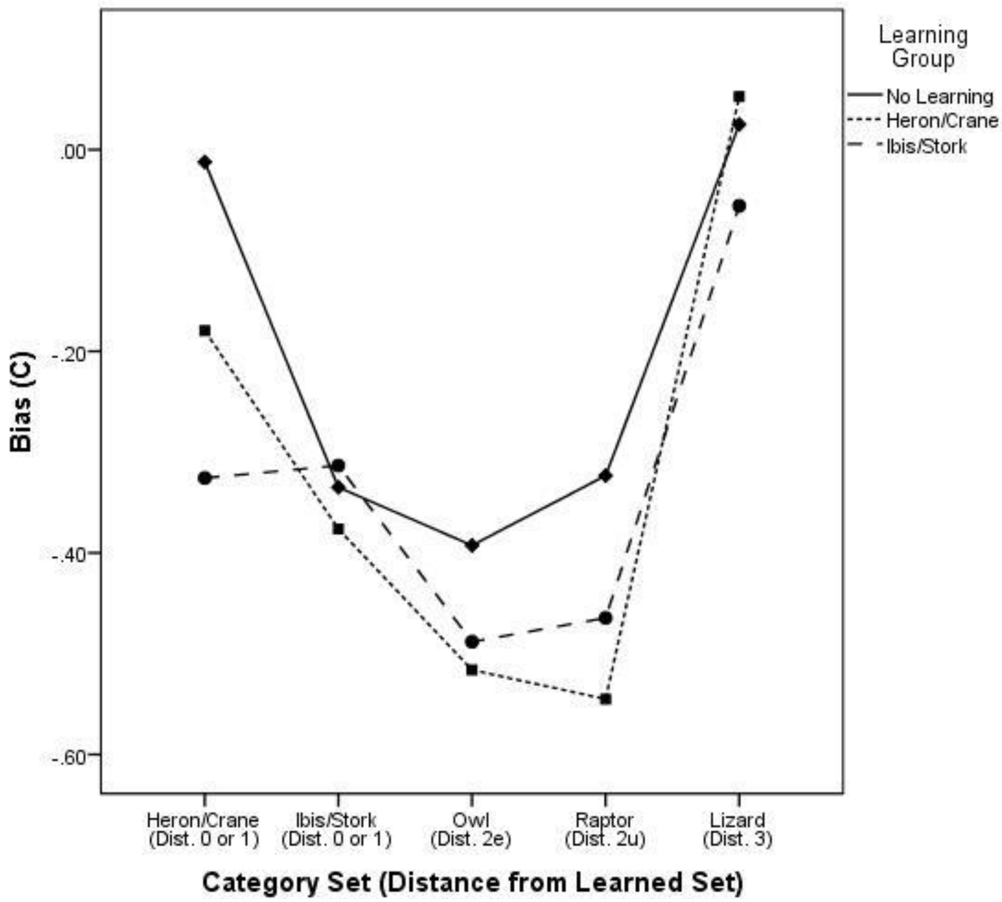
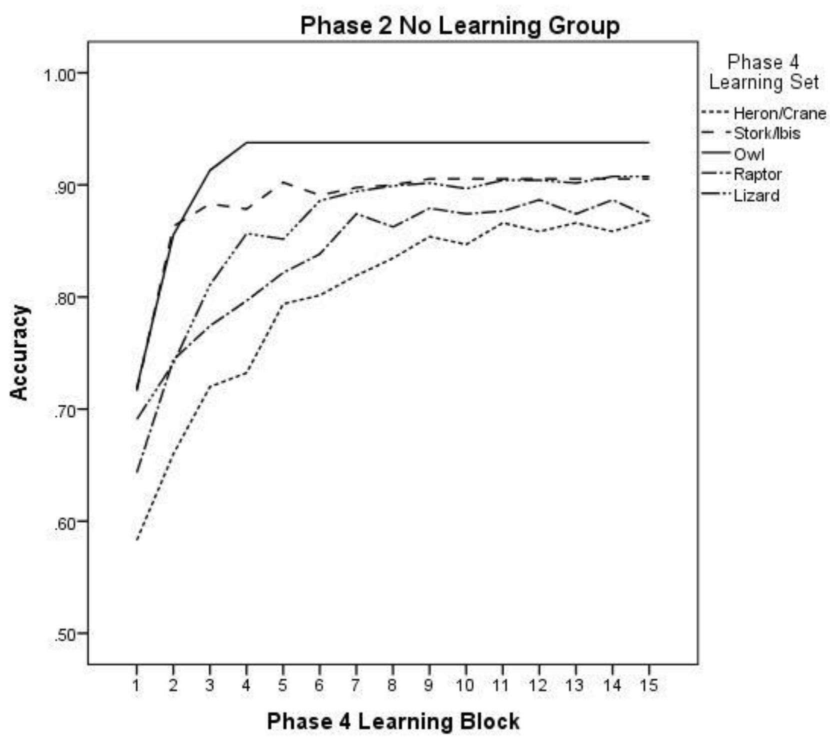


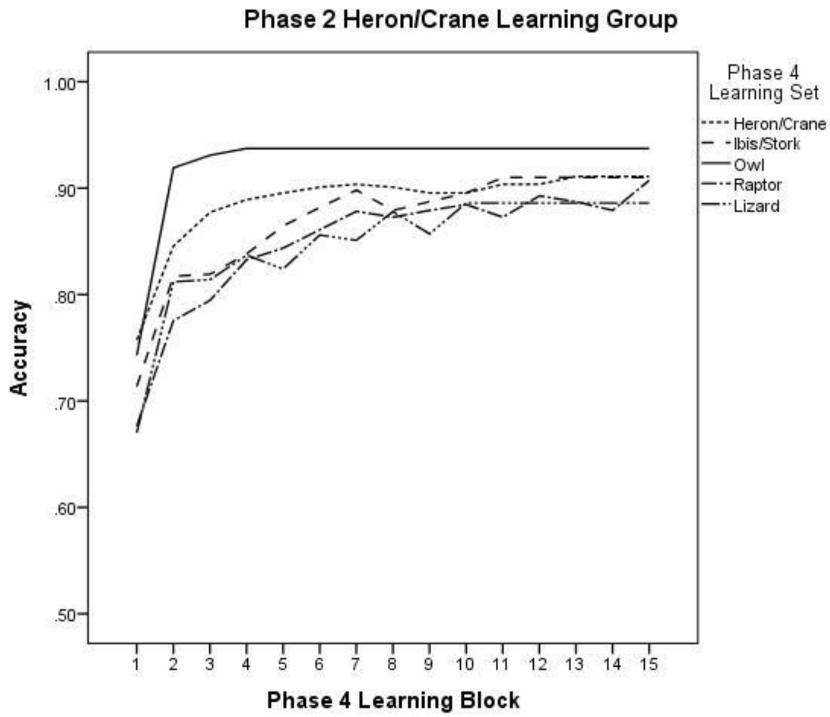
Figure 9. Bias (C) of discrimination of creature sets by the learning groups in Phase 3. There were no significant differences between learning groups. Distances from the learned set are described in Figure 2.

Phase 4. Two types of analyses were conducted on the learning in Phase 4. It is important to realize that all pictures used in Phase 4 were novel and had not been seen in any of the previous phases. The first analyses explored the impact of prior Phase 2 learning, including the no learning control, on the learning across blocks of categories used in Phase 4. The second set of analyses compared the number of blocks needed to reach the Phase 4 criteria of 85% correct.

Overall learning. A 3 condition x5 category type x15 learning blocks mixed model ANOVA was conducted on categorization accuracy. There was a significant main effects of Phase 2 learning group, F(2,154)=5.245, p=.006, and Phase 4 learning set, F(4,154)=7.859, p<.001, as well as, a significant interaction involving the two, F(8,154)=2.710, p=.008. The main effect and interactions involving learning block were all significant, p's < .01. Figure 9 illustrates participants learning across the 5 different Phase 4 learning sets in each of the different Phase 2 learning groups. Overall, the Phase 2 Ibis/Stork learning group (M=.891, SE=.007) had a significantly higher level of learning than the Phase 2 no learning group (M=.859, SE=.007). When looking at the effects of Phase 2 learning group on individual Phase 4 learning sets, the Phase 2 no learning group (M=.798, SE=.015) had a significant lower accuracy than the Phase 2 Heron/Crane group (M=.887, SE=.016) and the Phase 2 Ibis/Stork group (M.878, SE=.016) when learning the Heron/Crane set in Phase 4, p's <.001. The Phase 2 no learning group (M=.834, SE=.015) also had lower accuracy than the Phase 2 Ibis/Stork learners (M=.886, SE=.016) when learning the Raptor set in Phase 4, p=.023.

Because of the imputation of a participant's final block score on blocks after he or she reached criteria, there was a concern that the assumption of sphericity was violated; so a 5x3x5 (first 5 learning blocks x Phase 2 Learning Group x Category Set) Mixed Model ANOVA was conducted. The analysis also showed a significant interaction between learning phase 2 learning group and category set, $F(8, 154) = 3.202$, p=.002. Pairwise comparisons showed the same pattern of significant results as the full 15 block analysis with the following addition; The Ibis/Stork learning group (M=.898, SE=.026) had significantly higher accuracy than the Heron/Crane learning group (M=.811, SE=.025) when learning the Ibis-Stork category set during Phase 4, p=.016. Figure 10 shows the learning for all of the Phase 4 learning conditions separately for the no learning, heron/crane, and ibis/stork training in phase 2. Figure 11 focuses on highlighting the differences between the Phase 2 learning groups in the Phase 4 learning of the Heron/Crane and Ibis/Stork sets.





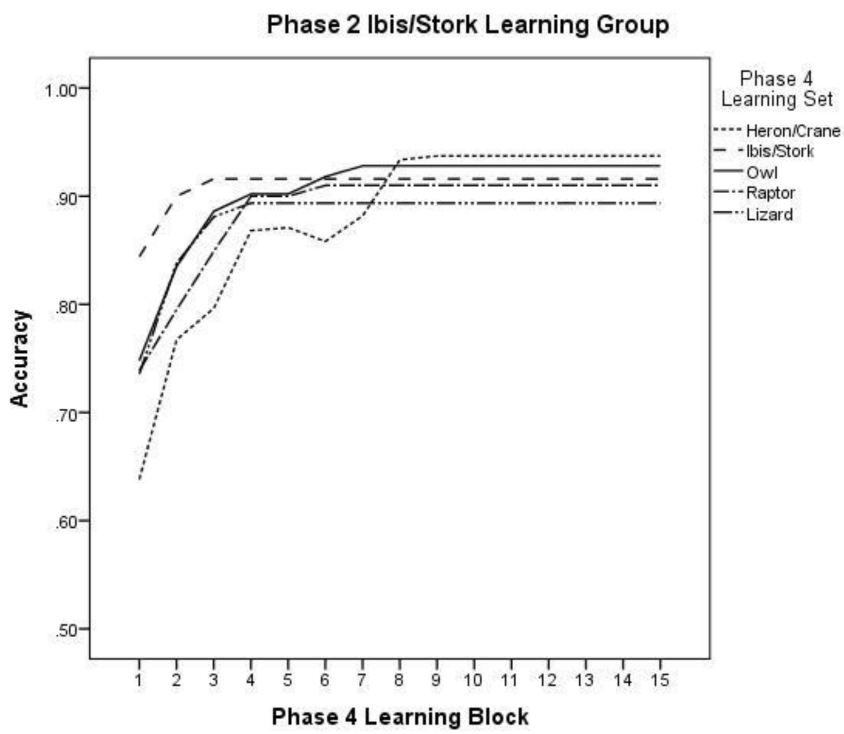
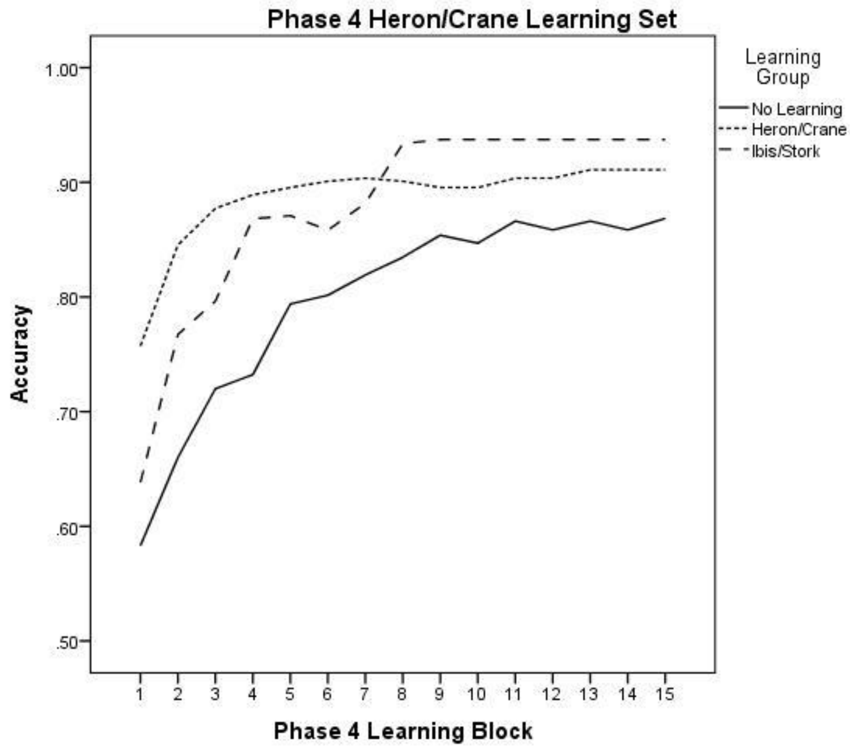


Figure 10. Accuracy across the 5 learning sets of Phase 4 for each of the 3 learning groups from Phase 2.



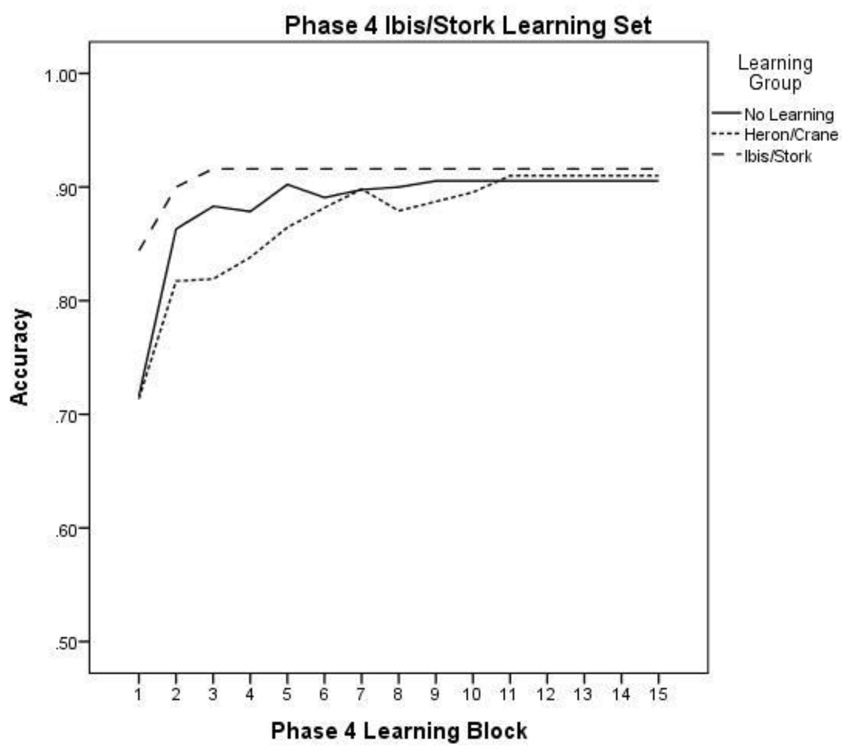


Figure 11. Phase 4 learning for the groups that learned the Heron/Crane and Ibis/Stork sets for each of the Phase 2 Learning Groups.

Blocks to criteria. A 3x5 (Phase 2 learning group x Phase 4 learning group) between subjects ANOVA was conducted on the number of blocks required to reach the criteria of 85% correct in Phase 4. Results showed a main effect of Phase 4 learning set, $F(4,150)=7.021$, $p<.001$, and a marginally significant interaction between the two factors, $F(8,150)=1.780$, $p=.085$.

However, the number of blocks to criterion is a relatively gross measure and has limited ability to reflect learning occurring within a block. Because of this, a second analysis was conducted explored the effects of the same factors on the number of half blocks required to reach a criteria of 85% correct. Each block of 32 trials was divided in half and the mean accuracy was calculated for each set of 16 items. Analyses showed that there was a main effect of Phase 4 learning set, $F(4,153)=5.766$, $p<.001$, and a significant interaction between Phase 4 learning set and Phase 2 learning group, $F(8,153)=2.489, p=.014$. Pairwise comparisons show that the Phase 2 no learning group ($M=4.815$, $SE=.377$) took a significantly higher number of half blocks to reach criterion than the Phase 2 Heron/Crane learning group ($M=3.704$, $SI=.404$) and the Phase 2 Ibis/Stork learning group ($M=3.658$, $SE=.412$), p's$<.05$. Analyses also showed that, when learning the Heron/Crane set in Phase 4, the Heron/Crane learning group ($M=2.727$, $SE=.894$) took significantly fewer half blocks to reach criterion than the no learning group ($M=7.500$, $SE=.856$) and the Ibis/Stork learning group ($M=6.455$, $SE=.894$), p's$<.01$. The Ibis/Stork and no learning groups did not differ in learning the Heron/Crane set in Phase 4. However, this pattern did not repeat for the Phase 4 Ibis/Stork set. While the Phase 2 Ibis/Stork learning group did reach criteria more quickly than the Heron/Crane or no learning groups, it did not do so significantly faster. Effects based on half blocks to criteria are depicted in Figure 12.

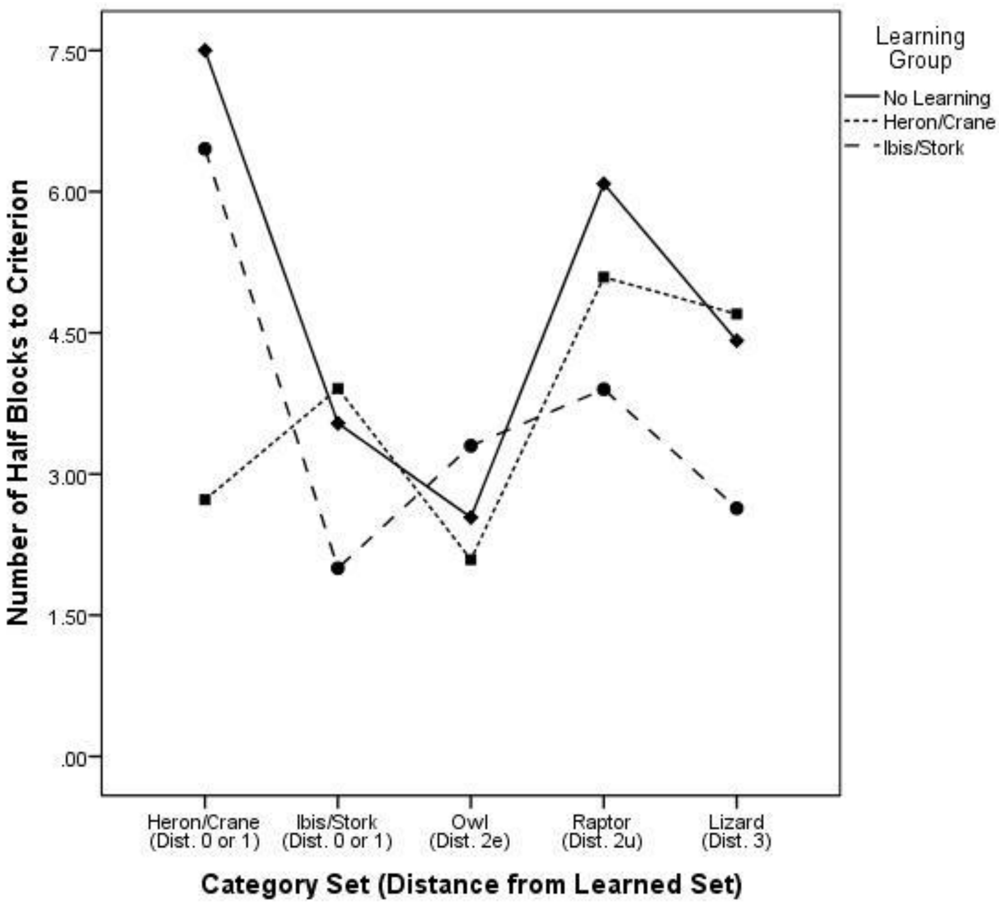


Figure 12. The number of half blocks required to reach the criteria of 85% accuracy in Phase 4. Distances from the learned set are described in Figure 2.

Discussion

The primary goal of experiment 1 was to test the ability of people to generalize their previous learning to new unlearned categories. This was tested in two ways, through a discrimination test and a learning test. Analyses from experiment 1 revealed lawful patterns of behavior in learning and converging measures of generalization in the discrimination and learning tests. Below the results of the tests are concisely described.

The implications of these findings are then discussed with the findings of experiment 2 in the general discussion.

Analyses conducted on learning at the order and family levels, revealed no differences between the learning groups in how much or how quickly they learned either the Order level groups, Wading Birds/Owls, or the Family level groups, Herons/Cranes or Ibises/Storks. They also revealed that most participants were able to learn both the order level categories, wading birds and owls, and the family level categories during both the family learning phase and during the learning test in a relatively short number of blocks and to a high level of accuracy. This indicates that participants were learning the structure and essence of the categories and not just memorizing exemplars. This is important because, generalization would not be expected if participants were just memorizing exemplars.

The Heron/Crane and Ibis/Stork learning groups also generally followed the same pattern of transfer to other categories. Both groups revealed similar learning. In each group, the key comparison was between a learning group and the control no learning group.

Analyses showed that participants, who learned the Phase 2 family level categories at a high criterion of accuracy, also were more accurate at discriminating between the two learned categories than the no learning group during the discrimination test. They also relearned or reached criteria much faster in the learning test than the no learning group did. Both the learning test and discrimination used different images than were used during the initial learning of those categories. This, again, indicates that the categories were truly learned and that increased performance was not merely due to memorization of exemplars.

However, analyses showed that the family level learning did not generalize to other types of wading birds. Specifically, Heron/Crane learners were not better at discriminating or learning the Ibis/Stork set than the no learning group. Likewise, the Ibis/Stork learners were not better at discriminating or learning the Heron/Crane set than the no learning group. There was also no generalization from learning to the owl or raptor categories. Neither the Heron/Crane nor the Ibis/Stork learning group differed from the no learning group on the owl or raptor category sets. Generalization also did not occur for the lizard group, the learning groups did not differ from the no learning group. These results indicate that generalization did not occur beyond the new exemplars from the learned categories. Closely related categories (the other set of wading birds) that were within the same order (basic) level as the learned family level categories did not have better performance for learners than for non-learners in either the discrimination test or the learning test. Likewise, learners were not better than non-learners, in the discrimination and learning tests, on categories that were within the same class (superordinate) level but not the same order (basic) level as the family level learned categories (the owls and raptors category sets). Learners were also not able to generalize to categories that were outside the class (superordinate) level category as the family level learned categories. Experiment 2 was designed to provide more converging evidence for the above pattern of generalization.

Experiment 2

The purpose of experiment 2 was to use multi-dimensional scaling to provide converging evidence for the pattern of hierarchical generalization found in experiment 1. It also provides an objective measure of the changes in participants' categorical representations and organizational structure as learning levels of the hierarchy occurs.

Method

Participants. Sixty-four participants were recruited from the Introductory Psychology participant pool from Arizona State University. They received partial course credit for participating. Two participants were removed from analyses because they did not reach the learning criteria in the learning phases, another 2 were removed from data analyses because their responses during scaling did not vary with respect to item differences, e.g. they issued the same response for almost all item comparisons.

Design and materials. The same materials used in experiment 1 were used in this experiment. Specifically, the same learning stimuli for the Phase 1 learning and Phase 2 learning were used here for Phase 1 and Phase 2. For the scaling stimuli, 24 stimuli (3 from each of the 8 bird family level categories) from the discrimination task of Phase 3 in the first experiment were chosen at random. No stimuli from the lizard groups were used in this experiment, since their inclusion would likely minimize any differences in the birds (and so grouping analysis of the birds would not be possible).

Procedure. Participants were split into three groups. The first group did not complete any learning and simply scaled the scaling stimuli. To scale the stimuli, each of the 24 stimuli was paired with every other stimulus and presented one pair at a time in random order. As each pair was presented, participants rated on a 9 point scale the similarity of the two stimuli. Participants were instructed to use the whole scale to evaluate the bird pictures.

The other two groups completed Phase 1 and Phase 2 in the same manner and to the same criterion as the participants in experiment 1. One group learned the Heron-Crane distinction in Phase 2 and the other group learned the Ibis-Stork distinction in Phase 2. After which, they will scale the scaling stimuli in the same way as the other group.

Results

Learning. Learning analyses similar to analyses from Phase 1 and 2 from Experiment 1 were conducted. A 2x5 (Learning Group x Block) mixed model ANOVA showed no significant effects or interactions involving learning group, p's > .15, in Phase 1. The same was true for the 2x15 (Learning Group x Block) mixed model ANOVA conducted on the learning in Phase 2, p's > .24. There was also not a statistically significant difference between the Heron/Crane (M=7.89, s=3.234) and Ibis/Stork (M=.6.23, s=2.759) learning groups in the number of Phase 2 blocks required to reach the learning criteria, $t(38)=1.754$, p=.088.

Scaling. Each participant rated the similarity of each item pair once. From these ratings, dissimilarity matrices were constructed for each participant. Then, using those matrices, the dissimilarities for each condition were multidimensionally scaled in SPSS using the PROXSCAL function. Each condition was scaled separately in 2 to 8 dimensions. Figure 13 shows the Stress 1 values for each of the learning conditions. The values were very similar among conditions.

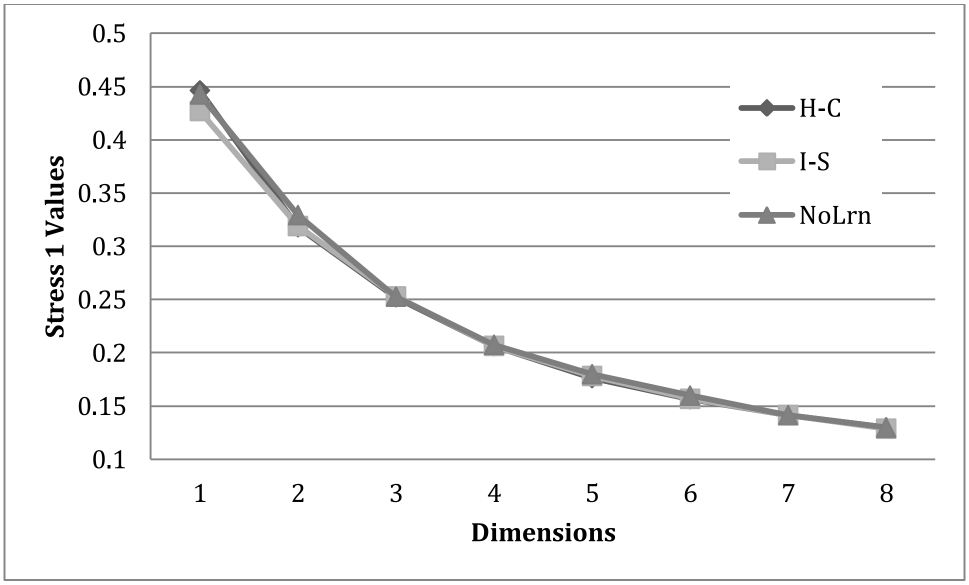


Figure 13. The Stress 1 values for each of the scaling solutions of the different conditions in 1-8 dimensions.

Based on the pattern of stress scores for each of the solutions, it was decided to use the 4 dimensional solutions for further analyses. The 4 dimensional stress 1 values were .2063, .2065, and .2076 for the Heron/Crane, Ibis/Stork, and No learning groups respectively.

Structure Analysis. I performed a number of analyses on the multidimensional solution in 4 dimensions. To anticipate, each MDS space contains 24 pictures, three examples from each of eight categories. Furthermore, the eight categories were composed of three basic levels (the orders of wading birds, owls, raptors), which could be subdivided into their subordinate categories (e.g., the families of Cranes, Herons, Ibis', and Storks). The analyses reported here address whether the space of 24 birds increasingly conformed to their natural groups, e.g., following training at the family level of Ibis and Storks, were these two families better separated in MDS space relative to the no learning control.

The distances between items from each of the 4 dimensional solutions were used to analyze the structure of each of the learning groups. Figure 14 shows a conceptual diagram that illustrates how structural values were calculated. Basically, for each item, the mean of the distances to within category items were computed and then divided by the mean of the distances to other category items. In these analyses, each item has its own structural ratio and acts as an individual entry for statistical tests. Smaller numerators in the ratio indicate a closer grouping of within category items, and larger denominators in the ratio indicate larger distances to other category items; thus, a smaller ratio indicates a more highly structured space.

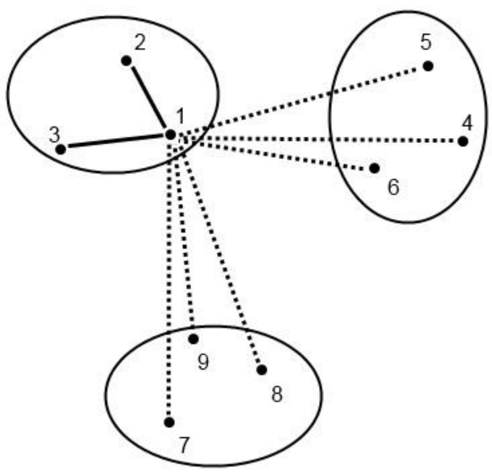


Figure 14. Conceptual diagram illustrating the general pattern for creating structural ratios from distances in multidimensional scaling solutions. For item 1, solid lines indicate within category distances and dotted lines indicate between category distances. For each item the mean within category distance divided by the mean between category distance constitutes the structural ratio for that item.

Because of the complexity of the hierarchical structure used in this experiment, several different, meaningful structural ratios can be calculated. The first ratio calculated and analyzed was at the order category level (e.g. wading birds). It was possible that learning a subset of wading birds, or learning to distinguish wading birds and owls could have differentially affected the participants' psychological space in comparison to those who received no learning. For this structural ratio (Order Ratio), the mean distance to other order level category items (e.g. for item Heron1 the mean distance to all other Herons, and all Cranes, Ibises, and Storks) was divided by the distance to all other items (e.g. for item Heron1 the mean distance to all True Owls, Barn Owls, Falcons, and Hawks). A 3x3 (Learning Group x Order level category) ANOVA was conducted using the Order Ratio as the dependent variable. The only significant effect was the main effect of order level category, $F(2,63) = 18.82$, $p<.001$. Planned comparisons showed that the Owl group was the most structured (M=.558, se=.023), followed by the Raptor group (M=.635, se=.023), with the Wading bird group (M=.724.se=.016) the least structured, all p's <.02. Figure 15 illustrates this structural difference.

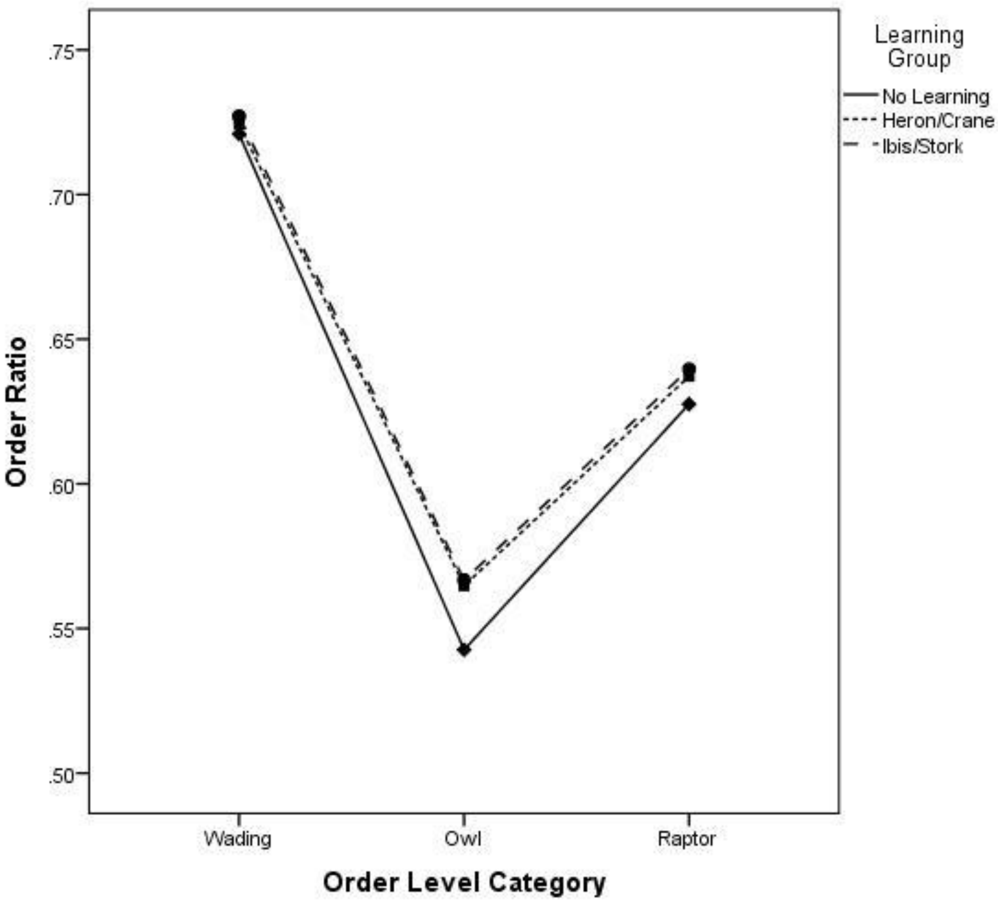


Figure 15. Order Ratio structure of order level categories split by learning group. Lower values indicate more structure.

A second structural ratio was calculated (family ratio). This ratio was calculated by dividing the mean distance of an item to other family level category items (e.g. for Heron1, the mean distance to Heron2 and Heron3) by the mean distance to all other non-family category items (e.g. for Heron1, the mean distance to all non Herons). This ratio can be used to explore differences in the structure of family level categories or sets of family level categories. These analyses were also conducted using a set ratio, which was calculated by dividing the mean of within family level category distances (e.g. the mean of the distance from Heron1 to Heron2 and Heron3) by the between group distances of its set pair from experiment 1 (e.g. the mean of the distances from Heron1 to Crane1 –

Crane3). The overall patterns of these analyses are similar to the analyses using the family ratio, and so, will not be reported in detail here.

Using family ratio as the dependent measure, a 3x8 (Learning Group x Family level category) ANOVA was conducted. Both main effects were significant, p's<.001, as was the interaction between learning group and family level category, $F(14,48)=4.837$, $p<.001$. Figure 16 illustrates the structure of each of the learning groups for each family level category. Planned comparisons showed that there were no significant differences between learning groups for the Heron, True Owl, Barn Owl, Hawk, or Falcon family level categories. For the Crane family level category, the no learning group (M=.527, SE=.048) was more structured than the Heron/Crane learning group (M=.708, SE=.048, $p=.01$) and the Ibis/Stork learning group (M=.696, SE=.048, $p=.016$), which were not different from each other. For the Ibis family level category, the Heron/Crane learning group (M=.983, SE=.048) had a less structured category than the Ibis/Stork learning group (M=.597, SE=.048, $p<.001$) and the no learning group (M=.594, SE=.048, $p<.001$). The Ibis/Stork and no learning groups did not differ. The Stork family level category follows the same pattern. The Heron/Crane learning group (M=.866, SE=.048) had a less structured category than the Ibis/Stork learning group (M=.573, SE=.048, $p<.001$) and the no learning group (M=.525, SE=.048, $p<.001$). The Ibis/Stork and no learning groups did not differ.

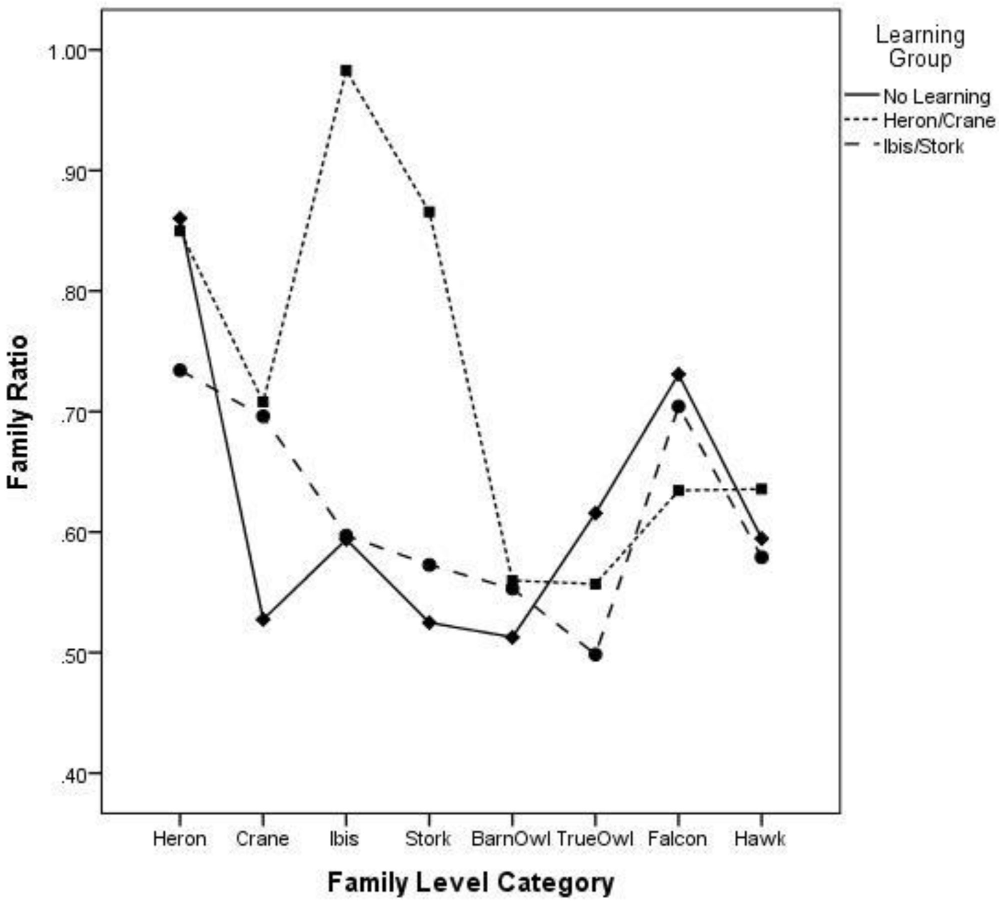


Figure 16. Family ratio values for each of the family level categories split by learning group.

A second analysis using the family ratio was conducted exploring the differences of learning group and the learning sets (e.g. Heron-Crane) described in experiment 1. This analysis is more directly comparable to the analyses in experiment 1, where learning set was one of the key variables. The 3x4 ANOVA showed main effects of learning group and learning set, p's<.001, and a significant interaction between the two factors, $F(6,60)=5.601$, $p<001$. Figure 17 shows the structure scores for each set of categories split by the learning groups. Pairwise comparisons showed no differences in learning group for the Owl, Raptor, or Heron-Crane sets. For the Ibis-Stork set, the Heron/Crane learning group (M=.924, SE=.042) was significant less structured than the Ibis/Stork learning group (M=.585, SE=.042, $p<.001$) or the no learning group (M=.559, SE=.042, $p<.001$), which were not different from each other.

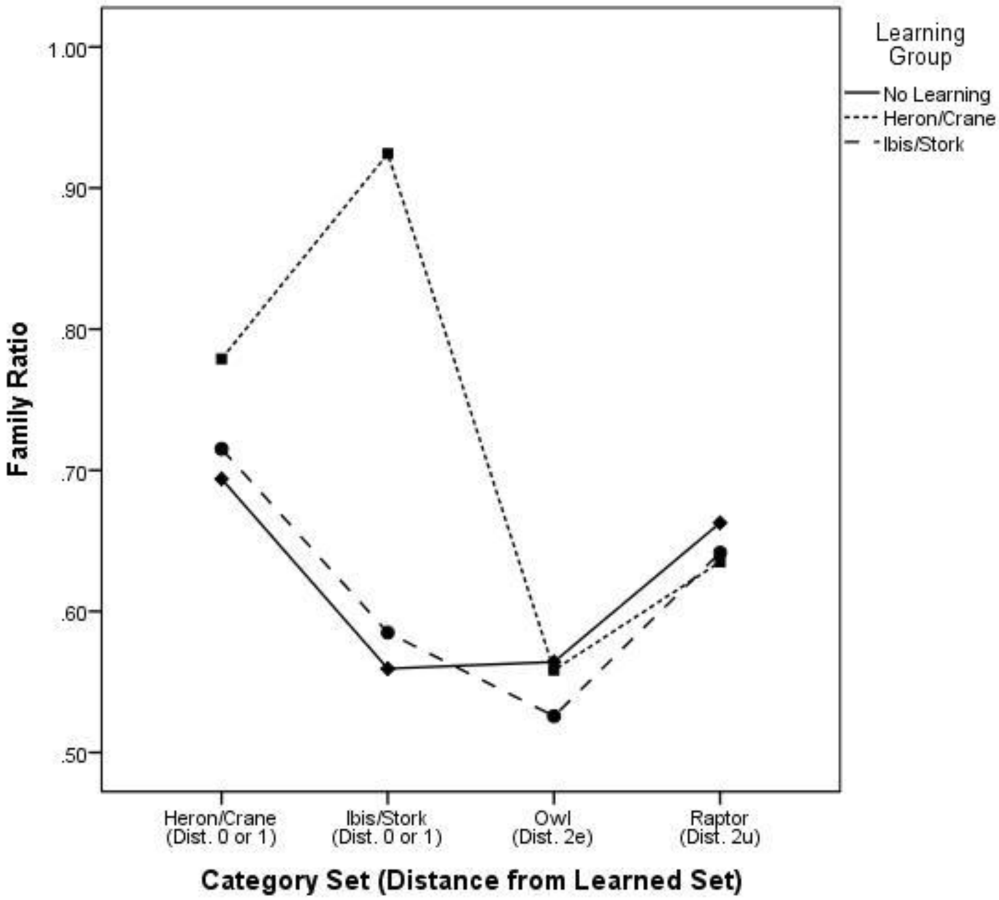
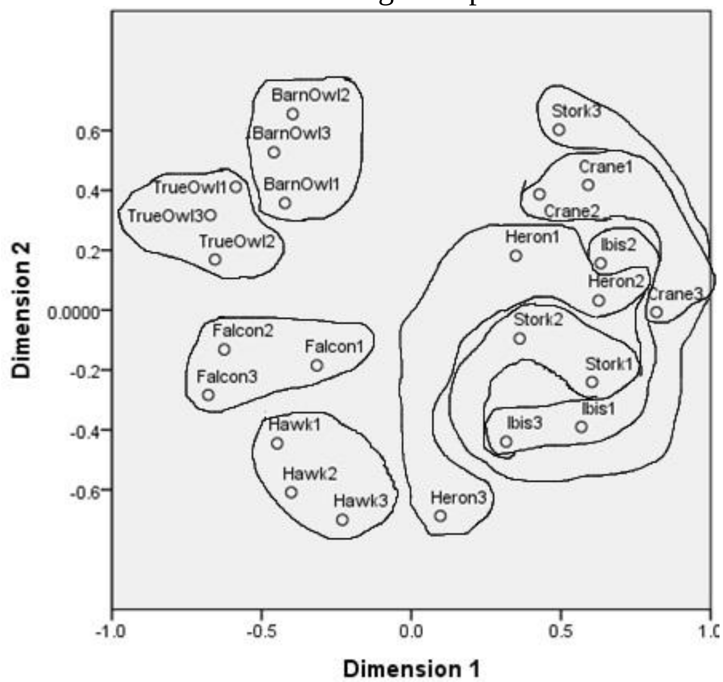


Figure 17. Family ratio values for each of the learning sets from experiment 1 split by learning group. Distances from the learned set are described in Figure 2.

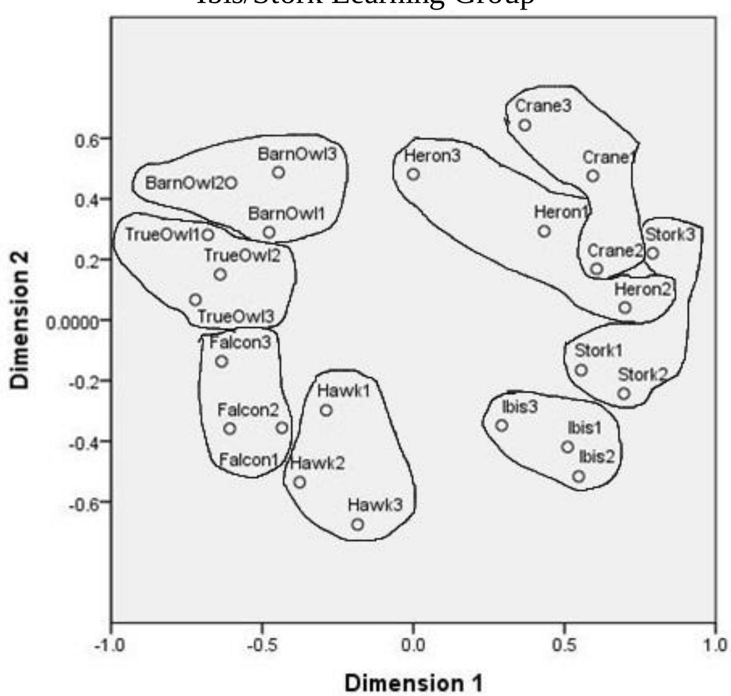
Because these patterns of effects differed to some extent from the findings in experiment 1, analyses were conducted on the 2 and 3 dimensional solutions. The 3 dimensional solutions followed the same pattern as the 4 dimensional solutions; however, the 2 dimensional solutions differed in several significant ways. As can be seen in Figure

18, the categories in the Ibis/Stork learning group were much more structured than in the no learning group.

No Learning Group



Ibis/Stork Learning Group



Heron/Crane Learning Group

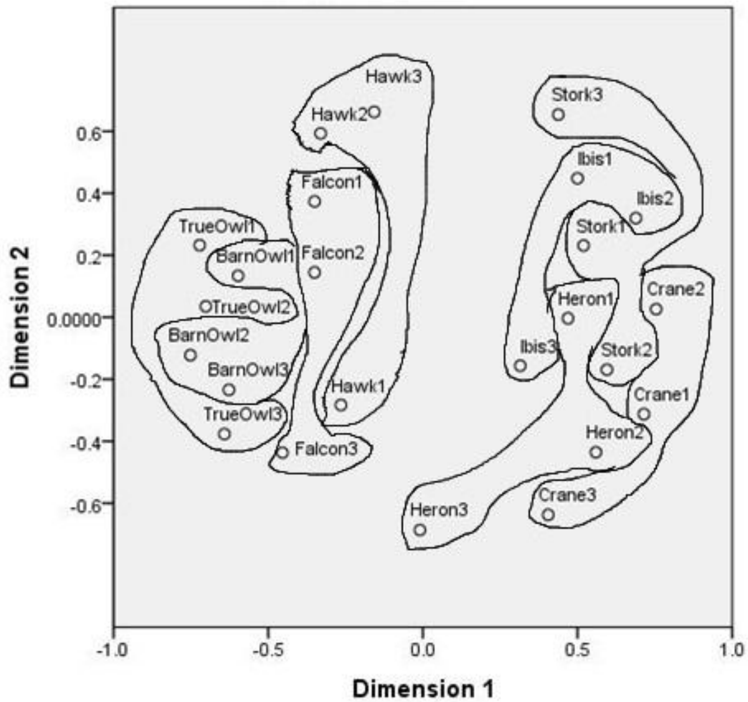


Figure 18. 2 dimensional solutions for the Ibis/Stork, Heron/Crane, and no learning groups. Each family level category is circled.

In a 3x4 (Learning Group x Set Name) ANOVA using the Family ratio calculated from the 2 dimensional solution distances, both main effects, p's<.001, and the interaction was significant, $F(6,60)=3.073$, $p=.011$. As can be seen in Figure 19, planned comparisons showed that for the Ibis-Stork set, the Ibis/Stork learning group (M=.309, SE=.061) was more structured than the Heron/Crane learning group (M=.610, SE=.061, $p=.001$) and the no learning group (M=.625, SE=.061, $p<.001$). There was no difference between learning groups in the Heron-Crane set. The key difference between the 2 dimensional and 4 dimensional solutions is that in 4 dimensions, the Heron/Crane learning group is worse than the Ibis/Stork and no learning groups, while in 2 dimensions, the Ibis/Stork learning group was more structured than the Heron/Crane and no learning groups. It is unclear as to why this change occurred.

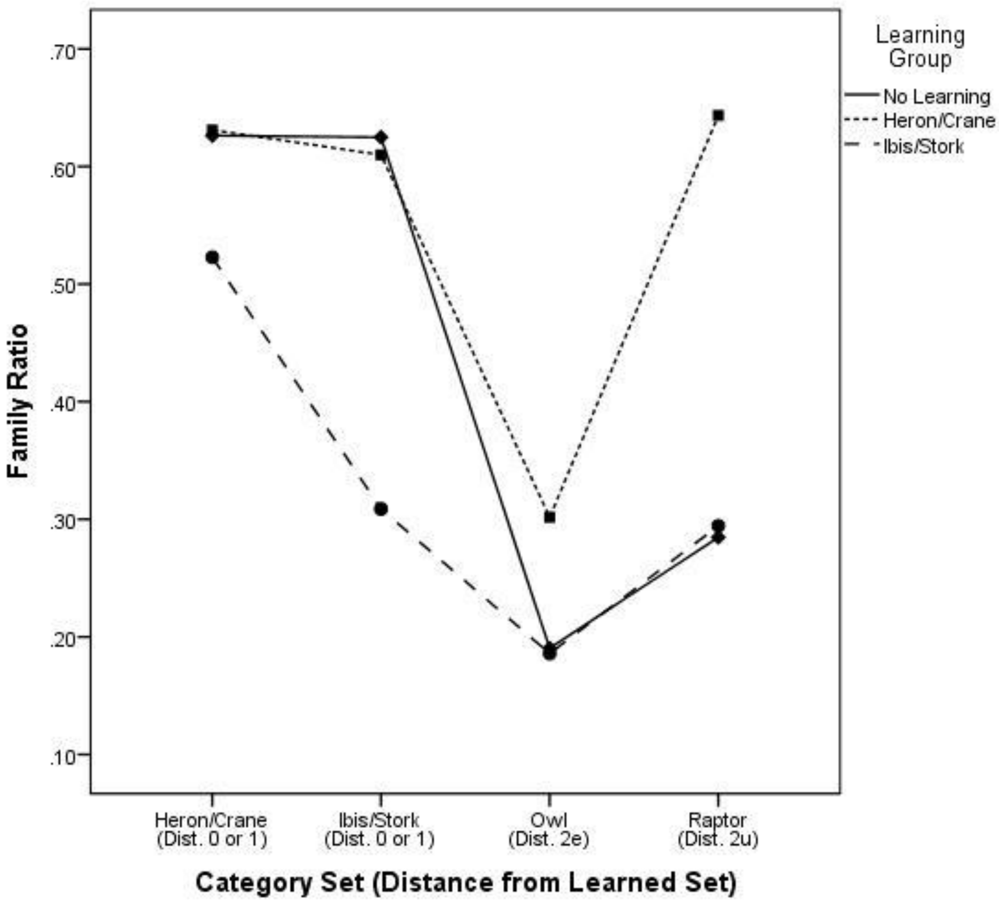


Figure 19. Family ratio values for the 2 dimensional solution split by set name and learning group. Distances from the learned set are described in Figure 2.

Discussion

The primary goal of experiment 2 was to determine if prior learning impacted, either selectively or globally, the psychological space containing these of learners. Because experiment 1 and 2 were designed to provide converging evidence for an overall pattern of generalization, the implications of the pattern of generalization found in experiment 2 will be discussed along side experiment 1 in the general discussion.

As in experiment 1, there were no differences between the learning groups in learning the order level, Wading birds/Owls, or in learning the family level, Herons/Cranes or Ibises/Storks. Multidimensional scaling the rating data indicated that Stress 1 values for each of the three groups, Heron/Crane learning, Ibis/Stork learning, and no learning, was similar to each other. This indicates that when comparing the structure of the categories for each of the three learning groups, differences were not due to a particular group having a different level of stress.

The two dimensional structure analyses revealed that the Ibis/Stork learning group had a stronger structure of the Ibis/Stork category set than the no learning group. They did not have a more structured space for any of the other family level category sets. Unlike in experiment 1, the pattern of results from the Heron/Crane learning group did not mimic the Ibis/Stork learning group. Here the Heron/Crane learning group did not differ from the no learning group on any of the family level category sets. These results indicate that, at least for the Ibis/Stork learning group, the learning selectively modified the learners' psychological space. Participants were able to better structure the categories they learned; this is impressive, especially because the exemplars used to measure the categorical structure were never before seen by the participants.

Because of relatively high stress values in the two dimensional solutions, analyses were also completed on the four dimensional solution. The pattern of results from the two dimensional solution did not reappear in the four dimensional solution. In the four dimensional solution, the Ibis/Stork learning group did not differ in category structure from the no learning group for any of the family level category sets; however, the Heron/Crane learning group did differ from the no learning group on the Ibis/Stork family level category set. The Heron/Crane learning group Ibis/Stork category set was less structured than the no learning group Ibis/Stork category set. There were no other differences between the Heron/Crane learning group and the no learning group. While the effect is different, and not as easily interpretable, it still indicates that learning selectively altered the category structure of participants. The effects also indicate that the category structure from the learners' psychological space did not change for any categories that were not learned; thus, the learning did not generalize to any other category sets.

The implications, with regard to the ability to transfer knowledge from learned categories to unlearned categories, of these results and the results of experiment 1 will be discussed in the general discussion.

General Discussion

The primary goal of these experiments was to mimic some aspects of perceptual expertise and then test the extent of generalization and transfer from one set of subordinate set of categories to other sets of subordinate level categories varying in relatedness from the learned categories. Participants learned multiple levels of a hierarchy, including typically subordinate level categories. They also learned them to a high degree of accuracy. Participants learned to differentiate two order categories from each other and then learned one set of family level categories. They were then tested on their ability to generalize to other family level categories. Those categories included the learned category set (either Heron/Crane or Ibis/Stork). The second set was a family level category set that was within the same order as the learned set, in this case the other wading bird set. The third and fourth family level sets were the owls and raptors. These sets were within the same class level as the learned family level categories (i.e. Birds), but were not with the same order level category (i.e. wading birds). The reason there were two sets at this hierarchical distance was for one set, the raptors, to act as a control for the owls, which had been experienced during the order level learning. The final family level category set, the lizards, was outside the class of the learned family level learned categories, and acted as a control or baseline for performance. These category sets and the hierarchical organization can be seen in Figure 2.

There were several theoretically plausible patterns of behavior that could have occurred. One possibility is that there would be no generalization to categories other than to new exemplars of the learned ones. A second plausible result was the domain specificity hypothesis, the idea that there would be a limited influence, where learning would generalize within a domain, or boundary, and there would be very little generalization outside the domain. The size of the domain could vary based on a number of factors, would be determined by the locus of learning within the hierarchy where learning occurred. For example, one possible pattern of results was a large amount of generalization to the two wading bird sets, but no generalization to any of the other bird category sets or the lizard set. This would suggest that learning can transfer to other categories within the same one level higher category, but not to categories outside of that higher-level category. In this instance, the one level higher category is the order level category; so learning would transfer with the family level wading bird categories, because that is where the original learning occurred, but could not transfer to other family level categories in other orders, e.g. owls or raptors. A final plausible pattern of results was for there to be generalization, but no specificity. Here the degree of generalization that occurred would be based on the degree of perceptual similarity between the learned categories and the other categories. So a large amount of transfer would occur for the non-learned wading birds, less would occur for the non-wading bird birds, and little to no transfer would occur for the lizards. Each of these patterns is illustrated in Figure 1.

Experiment 1 and 2 were designed to provide converging evidence of one of the patterns of behavior described above. There were several measures of transfer providing converging evidence: the accuracy and sensitivity of the discrimination test in experiment 1, the learning and number of blocks to criterion in the learning test of experiment 1, and the structure measurement from the scaling solutions of experiment 2. Mostly, these measures did follow the same pattern of results. In the accuracy and sensitivity (d`) analyses of the discrimination test, participants who learned a category set did much better than those who did not learn, in discriminating between that set. The learning group did not do any better than the no learning group on any of the other category sets. Likewise the learning groups had higher overall learning than the no learning group on the corresponding category sets. Similarly, the Heron/Crane learning group required fewer half blocks of learning to reach criteria than the no learning group, when the learning test consisted of the Heron-Crane set. The Ibis/Stork learning group also took fewer half blocks to reach criteria on the corresponding category set, though not significantly so. Also similar to other results, the Ibis-Stork category structure in the multidimensional scaling solutions for the Ibis/Stork group was greater than the no learning group. The same was not true for the Heron/Crane learning group for the Heron-Crane category set. All of this evidence points to little or no generalization beyond new exemplars of the learned category, which is depicted in Figure 20.

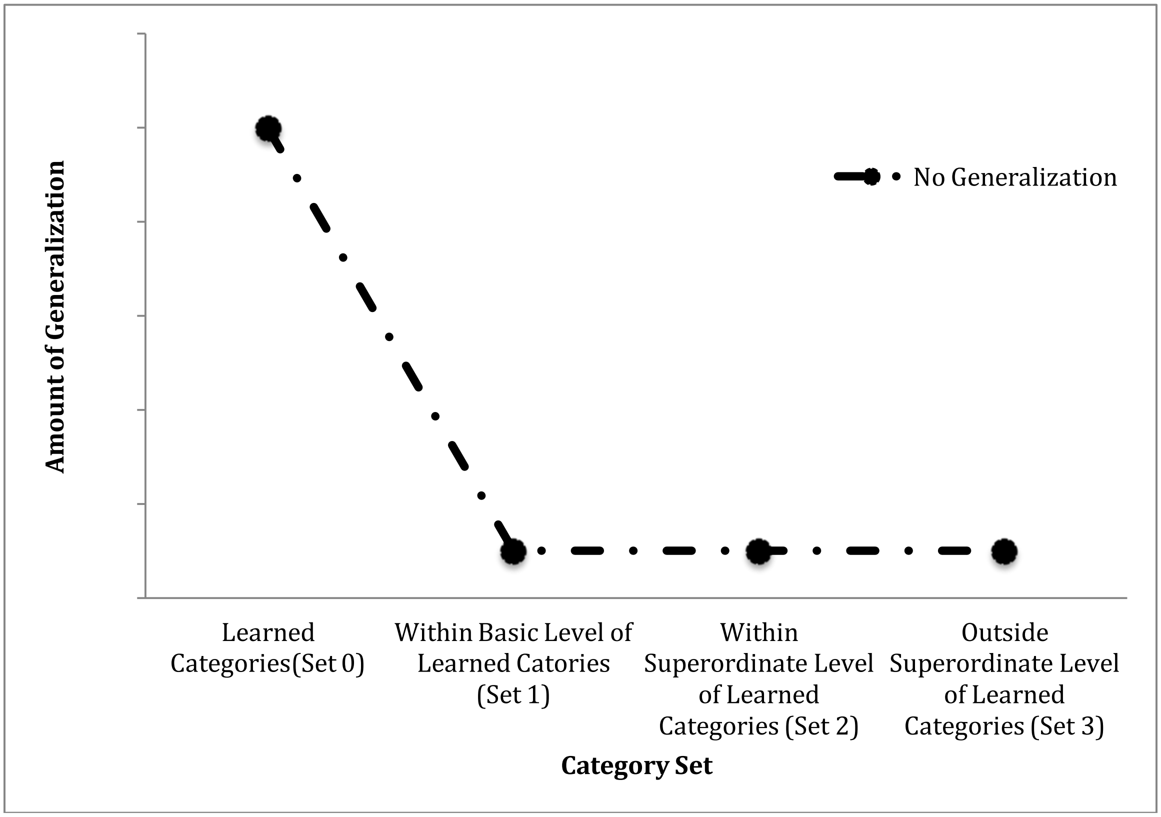


Figure 20. General pattern of results across the discrimination test, learning test, and structure of categories during multidimensional scaling. The set number is a function of hierarchical distance as outlined in Figure 2.

This result, no generalization beyond the learned categories, fits well with previous research. Scott, et al. (2008) found that learning models of cars did not improve discrimination of other car models, nor did it improve performance in discriminating older car models, which had not been discriminated before. The current results match Scott, et al.'s (2008) very well. Participants did not generalize to new wading bird categories, nor did they generalize to owls or raptors, which are analogous to the older car models. Tanaka, et al. (2005) found that when participants learned wading bird species, they were better at discriminating new wading bird species than at discriminating owl species. The current results match their results fairly well. Participants in both studies were able to discriminate between new exemplars of learned categories. Tanaka, et al. (2005) found a small, but significant improvement for unlearned wading bird species, while here there tended to be a small increase in ability to discriminate unlearned wading bird categories; however, the increase was not significant.

While the current results match past work in a number of ways, the current studies go beyond past work. First, previous work has only used discrimination tests to assess generalization. Here generalization was tested using a variety of methods. These included a discrimination test, a prepared for future learning test, and a measure of the learners' psychological space and organization of categories. All of these measures showing a common trend of behavior provide strong support for the conclusion that transfer of perceptual knowledge between categories is extremely limited.

Second, the current work used a category structure with more variability within a category than past work. Scott, et al. (2008) used pictures of models of cars and Tanaka, et al. (2005) used bird species. Both categories use images that are highly similar, meaning that discrimination between them was relatively easy, e.g. participants in Scott et al.'s (2008) work were over 70% accurate in discriminating between them even before learning. The current work used bird families as categories, which consisted of several related bird species. The families were more variable as indicated by the no learning group only reaching approximately 60% accuracy in discriminating between them. Previous work, e.g. Homa and Cultice (1984), has shown that having a more variable category structure improves generalization to new exemplars from the learned categories over just seeing highly similar exemplars during learning. It was hypothesized that this more variable category structure would also improve generalization to unlearned categories, in addition to learned categories. This appears to not be the case, because there was no significant generalization outside of the learned categories. One potential possibility for the unexpected lack of an effect of the increased variability is in the nature of the variability. In Homa and Cultice (1984), the categories were artificial and the variability was spread out evenly around a centroid. In the current work, the increased variability could not be evenly distributed. Each family level category consisted of 4 exemplars from each of 4 species. While this is more variable than the subordinate categories that Tanaka et al. (2005) used, which were just exemplars of a single species, it is not evenly distributed. The distribution of a subordinate (family) level category here consisted of several clusters of highly similar exemplars. Thus, it may be functionally impossible to create evenly distributed categories, in the manner of Homa and Cultice (1984), using real world hierarchical categories.

The third way that the current work goes beyond previous work, is that here a larger more complex hierarchy was used. The hierarchy used here had two distinct differences from previous ones used. First, the hierarchy used here included a set of control categories that were outside the class level category of the learned categories, i.e. the lizards were outside the bird category. This group was specifically added to act as a baseline where no generalization was expected to occur. The second primary difference between the current and previous hierarchies was the inclusion of 2 sets of categories that were within the class level category of the learned sets, but outside the order level category of the learned sets. Previous work has used a single set of categories of this nature, e.g. the owl species from Tanaka et al. (2005). The addition of the second category set of this type allowed for a test to explore if experiencing the category set, but not differentiating within it improves generalization, e.g. the owls were seen during the order level learning, but participants did not learn to differentiate between True and Barn owls, while the raptor category set was never seen before the transfer tests. These additions did not reveal key patterns of behavior here, because generalization did not extend beyond the learned categories; however, future work that intends to measure generalization within a partially learned hierarchy will need to for these factors, a true baseline category and experience, but not learning of the categories.

There are several potential reasons why no generalization beyond the learned categories was found. The domain specificity hypothesis, where there was a great deal of generalization within a certain hierarchical distance of the learned categories and none or very little beyond that distance, was predicated upon the idea that transfer would be facilitated by attention to particular features of the exemplars within that category. Previous work by Goldstone and colleagues (Goldstone & Steyvers, 2001; Kersten et al., 1998; Schyns et al., 1998) has shown that learning categories draws attention to diagnostic features or sets of features. The domain specificity hypothesis predicted that closely related categories would have the same or highly similar set of diagnostic features as the learned categories. Thus, learners would already have their attention focused on the relevant dimensions for closely related categories, but not for less closely related categories. Being focused on the appropriate dimensions would facilitate the discrimination of those categories. However, as was discussed above, learning did not facilitate discrimination between unlearned categories, regardless of how closely related the categories were to the learned categories. There is still a question about why the directed attention did not facilitate further learning. One possibility is that such attention never has an effect, that when people are cued to new categories, they start from scratch their search for relevant dimensions. However, that possibility does not fit well with Kersten, et al.'s (1998) idea of attentional persistence, where more attention is allocated to dimensions or features that are predictive. Another possibility is that, because real life categories were used, it cannot be guaranteed that the same dimensions useful in distinguishing between herons and cranes were also useful in distinguishing between ibises and storks. Real world categories were used for several reasons, including the ability to construct large categories and a large hierarchy, both of which were highly externally valid; however, using real world categories did have some drawbacks. One potential drawback was the inability to guarantee that some categories shared relevant dimensions. Another being that ideally each of the category sets would have an equal baseline from the no learning group; however, that was not the case, some categories were inherently easier to discriminate than other. This impeded direct comparisons across category sets within a condition. While there is no evidence to confirm that the closely related category sets of the wading birds had different diagnostic features, it is a possibility.

Another potential issue was the lack of complete expertise. While perceptual expertise was mimicked in several ways in this experiment, participants learned multiple levels of a hierarchy and learned to a high degree of accuracy, there are several other aspects of perceptual expertise that were not captured here. One aspect is the extensive experience with the learned categories. The learning in the current experiments did result in high accuracy, but was not prolonged and only lasted until participants were proficient in categorizing. While some work, e.g. Tanaka et al. (2005), had more prolonged training and still did not show much if any generalization beyond the learned categories, that work did not include the characteristics described above to enhance the ability to detect generalization. It is possible that more aspects of expertise along with more sensitive measures are needed to detect generalization.

The current work here indicates that people who learn one set of wading bird families are not able to generalize their knowledge to other wading bird, owl, raptor, or lizard categories. This result does fit with some past work, e.g. Scott et al. (2008), but does not appear to fit with Kersten et al.'s (1998) concept of attentional persistence, where previous learning should improve future learning if the categories share diagnostic dimensions. Thus, we can conclude that generalization beyond learned categories is not widespread, as in across a hierarchy, and does not provide large easily detectable differences, at least not in many cases, but it may be premature to conclude that no perceptual generalization ever occurs. To provide more control over the learning situation, future work could focus on creating artificial hierarchies. Doing this would create categories that are equal in structure and the degree of overlapping diagnostic dimensions can be closely controlled. Future work should also include multiple measures of generalization and more direct measures of people's attention to test whether previous learning does cue attention to focus on previously relevant dimensions. Regardless, the results of the present study confirm that categorical training on detailed, pictorial stimuli can generate substantial generalization to that category but this generalization appears not to extend beyond the categories learned.

References

Ashby, F. G. & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33-53.

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition, 84,* 1-53.

Biederman, I. & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(4), 640-645.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120*(3), 278-287.

Bukach, C. M., Phillips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Attention, Perception, & Psychophysics, 72*(7), 1865-1874.

Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.). *Creative Thought: An investigation of conceptual structures and processes* (pp. 209-234). Washington, DC: American Psychological Association.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 21-30). New York, NY: Cambridge University Press.

Chi, M. T. H., Glaser, R., & Rees, E. (1982) Expertise in problem solving. In J. Sternberg (Ed). *Advances in the Psychology of Human Intelligence (Vol. 1),* (pp. 7-75). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Chi, M. T. H. & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19*(1), 29-39.

Coley, J. D., Hayes, B., Lawson, C., & Moloney, M. (2004). Knowledge, expectations, and inductive reasoning within conceptual hierarchies. *Cognition, 90,* 217-253.

Corter, J. E. & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin, 111*(2), 291-303.

Deneault, J. & Ricard, M. (2005). The effect of hierarchical levels of categories on children's deductive inferences about inclusion. *International Journal of Psychology, 40*(2), 65-79.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117.

Ericsson, K. A. (1996). The Acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1-50). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A. (2006) The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 683-703). New York, NY: Cambridge University Press.

Gobbo, C. & Chi, M. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development, 1*, 221-237.

Gobet, F. & Charness, N. (2006). Expertise in chess. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 523-538). New York, NY: Cambridge University Press.

Goldstone, R. L. & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130*(1), 116-139.

Griffee, K. & Dougher, M. J. (2002). Contextual control of stimulus generalization and stimulus equivalence in hierarchical categorization. *Journal of the Experimental Analysis of Behavior, 78*, 433-447.

Hodges, N. J., Starkes, J. L., & MacMahon, C. (2006). Expert performance in sport: A cognitive perspective. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 471-488). New York, NY: Cambridge University Press.

Homa, D. & Chambliss, D. (1975). The relative contributions of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 351-359.

Homa, D. & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 83-94.

Homa, D., Rhoads, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory, 5*(1), 11-23.

Homa, D. & Rogers, D. (2011). The cumulative modification of categorical knowledge: Testing a method for further investigation. *Unpublished senior honors thesis.*

Homa, D. & Silver, R. (1976). Triadic decision making in lexical memory. *Memory & Cognition, 4*(5), 532-540.

Johnson, K. E. (2001). Impact of varying levels of expertise on decisions of category typicality. *Memory & Cognition, 29*(7), 1036-1050.

Johnson, K. E. & Mervis, C. B. (1994). Microgenetic analysis of first steps in children's acquisition of expertise on shorebirds. *Developmental Psychology, 30*(3), 418-435.

Johnson, K. E. & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126*(3), 248-277.

Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1437-1458.

Lancaster, M. E. & Homa, D. (2012, May). *Generalization when learning different levels of a hierarchy.* Poster presented at the 2012 Association for Psychological Science Annual Convention, Chicago, IL.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome. *Cognitive Psychology, 32*, 49-96.

Minda, J. P. & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 775-799.

Murphy, G. L. & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior, 21*(1), 1-20.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25-53.

Nosofsky, R. M. & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 924-940.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life. *Science, 296*(5571), 1321-1323.

Patel, V. L., Kaufman, D. R., & Magder, S. A. (1996). The acquisition of medical expertise in complex dynamic environments. In K. A. Ericsson (Ed.). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 127-165). Mahwah, NJ: Lawrence Erlbaum Associates.

Posner, M. I. & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*(2), 304-308.

Quinn, P. C. & Tanaka, J. W. (2007). Early development of perceptual expertise: Within-basic-level categorization experience facilitates the formation of subordinate-level category representations in 6- to 7-month-old infants. *Memory & Cognition, 35*(6), 1422-1431.

Richler, J. J., Wong, Y. K., & Gauthier, I. (2011). Perceptual expertise as a shift from strategic interference to automatic holistic processing. *Current Directions in Psychological Science, 20*(2), 129-134.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Bream, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Saiki, J. (1998). The role of structural consistency between categories and attributes in hierarchical category learning. *Japanese Psychological Research, 40*(3), 144-155.

Schraagen, J. M. (2006). Task analysis. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 185-202). New York, NY: Cambridge University Press.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.). *Transfer of Learning from a Modern Multidisciplinary Perspective*, (pp. 1-51). Greenwich, CT: Information Age Publishing.

Schyns, P., Goldstone, R. L. & Thilbaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*, 1-54.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research, 1210*, 204-215.

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science, 16*(2), 145-151.

Tanaka, J. & Gauthier, I. (1997). Expertise in object and face recognition. *The Psychology of Learning and Motivation, 36*, 83-125.

Tanaka, J. W. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457-482.

Tversky, B. & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*(2), 169-193.

APPENDIX A

FULL SPECIES LIST USED FOR STIMULI DEVELOPMENT LISTED BY ORDER AND FAMILY LEVEL CATEGORIES

Order Category	Family Category	Species 1	2	3	4
Wading Birds	Herons	Cattle Egret	Great Blue Heron	Reddish Egret	Yellow-Crowned Night Heron
	Cranes	Blue Crane	Grey-Crowned Crane	Sandhill Crane	Whooping Crane
	Ibises	African Sacred Ibis	Glossy Ibis	Madagascar Ibis	White Ibis
	Storks	Wood Stork	Marabou Stork	Black Necked Stork	White Stork
Owls	True	Elf Owl	Northern Hawk Owl	Northern Saw-whet Owl	Eastern Screech Owl
	Barn	Barn Owl	Oriental Bay Owl	Australian Masked Owl	Ashy-faced Owl
Raptors	Falcons	Collared Forest Falcon	Common Kestral Falcon	Pygmy Falcon	American Kestral Falcon
	Hawks	Barred Hawk	African Harrier Hawk	Steller’s Sea Hawk	Short-toed Snake Hawk
Lizards	Agamids	Australian Water Dragon	Fan-throated Lizard	Eastern Bearded Dragon	Blanford’s Rock Agama
	Anoles	Allison’s Anole	Brown Anole	Neotropical Green Anole	Blue-lipped Forest Anole

APPENDIX B

MEANS AND STANDARD DEVIATIONS OF REACTION TIMES OF DURING THE DISCRIMINATION TEST IN EXPERIMENT 1

Learning Group	Trial Type	Category Set Reaction Time Mean (Standard Deviation)				
		Heron/Crane	Ibis/Stork	Owl	Raptor	Lizard
Heron/Crane	Correct	2121.62 (830.51)	1788.08 (895.18)	1381.74 (750.97)	1837.57 (784.21)	1771.62 (852.39)
	Incorrect	2120.11 (1262.82)	1937.19 (1044.76)	1580.52 (1212.52)	1994.70 (1045.22)	1911.39 (1352.93)
Ibis/Stork	Correct	2194.88 (993.25)	2170.24 (730.56)	1777.87 (1006.35)	2275.89 (1227.83)	1958.69 (1002.52)
	Incorrect	2.352.72 (1088.42)	2420.38 (1221.72)	1988.39 (1317.00)	2256.32 (1041.98)	2080.40 (1047.60)
No Learning	Correct	2125.91 (1172.23)	2136.26 (1194.26)	1707.05 (889.00)	2131.23 (976.92)	1912.28 (983.84)
	Incorrect	2154.99 (1333.88)	2046.19 (1159.22)	2168.67 (1410.87)	2225.51 (1128.69)	2168.67 (1410.87)

APPENDIX C

APPROVAL FOR HUMAN SUBJECTS FROM INTERNAL REVIEW BOARD AT

ARIZONA STATE UNIVERSITY

ARIZONA STATE UNIVERSITY
Research Compliance Office
Office for Research & Sponsored Projects Administration
P.O. Box 871503
Tempe, AZ 85287-1503

Phone
(480) 965-6788
Facsimile
(480) 965-7772

To:	Donald Homa PSY
From:	Mark Roosa, Chair Soc Beh IRB
Date:	08/25/2008
Committee Action:	**Exemption Granted**
IRB Action Date:	08/25/2008
IRB Protocol #:	0808003154
Study Title:	Making Inferences

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(2) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.