Investigations into Unsupervised Category Learning:

The Role of Working Memory in

Learning Category Structures

by

Matthew W. Hayes, M. A.

A Thesis/Dissertation

In

EXPERIMENTAL PSYCHOLOGY

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

Roman M. Taraban

Ruth Hipple Maki

William S. Maki

Elizabeth M. Williamson

John Borrelli
Dean of the Graduate School

August, 2007

ACKNOWLEDGMENTS

No work of this scope ever results from the solitary efforts of a person alone. This work is the product of an ongoing dialog, a long collaboration with people interested as much in the process as in the final product. A dedication of a work to a wife or husband cannot hope to capture in a few words the sacrifice and support given in its creation. Her help contacting and running subjects was essential, but was secondary in importance to her willingness to listen to my attempts to put my thoughts into words; the formative moments without which this work would not exist. I also need to acknowledge the support of my advisor, Dr. Roman Taraban, whose patience never ran out and whose careful guidance helped me get here. To my two greatest supporters and collaborators, Amy and Roman, thank you.

I also need to acknowledge the support of Dr. Kate Bleckley and the Texas Tech University Graduate School. Kate's statistical expertise and extensive knowledge of working memory were absolutely critical. This dissertation was also supported by a Summer Dissertation Research Scholarship from the Texas Tech University Graduate School.

TABLE OF CONTENTS

ABSTRACT

The present research explored the role of working memory (WM) in unsupervised category learning, learning without an external tutor or even knowing that categories exist, by investigating its role using a pattern-sequence manipulation. A pattern-sequence manipulation compares learning when items from categories are presented together (blocked) versus when the items are presented in random order (mixed). Experiment 1 extended the pattern-sequence manipulation to assess category knowledge separate from paired-associate learning. Participants performed equally well on new and studied items, supporting the hypothesis that the pattern-sequence manipulation results in the acquisition of category information, not simply memory for item-feature associations. Experiment 2 introduced a WM factor, administering the method used in Experiment 1 to a group of high and low WM span participants. High WM span was predicted to interact with the pattern-sequence effect to produce greater learning when the items were blocked than mixed. There was reliable support for a role of WM span in the discovery and acquisition of category knowledge, but this role was different from the one hypothesized. The high WM span participants exhibited higher overall accuracies than the low WM span participants. This result supports a role for WM in unsupervised category learning, but did not benefit more from the pattern-sequence effect than did the low WM span participants as predicted. Implications for theories of category learning and WM are discussed.

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND OVERVIEW

Introduction

Will that animal scratch me? Does it like to play fetch? Will I break out in hives, or can I pet it? In a world defined by limitless variety, we see groups or categories of similar objects or events. These categories reduce the impossible burden of learning the properties of each and every object or event as a disconnected and unique individual, to learning the properties of far fewer categories and to which of those categories the object or event belongs. Our experience of the world is largely defined, not by a random and disordered collection of idiosyncratic individuals, but by a structured and organized set of categories and their members. Knowing what type of animal a friend purchased as a pet will allow you to infer many of its properties without directly experiencing them (Ashby & Maddox, 2005; Love, Medin, & Gureckis, 2004; Medin 1989; Medin & Schaffer, 1978; Nosofsky, 1986, 1988; Kruschke, 1992; Markman & Ross, 2003; Posner & Keele, 1968). For example, knowing that the pet is a cat allows you to draw conclusions about its behavior (it is unlikely to play fetch), its physical characteristics (cats have claws), and other relevant information about it (it may scratch and, if I'm allergic to cats, I had best not pet it). Categories allow people to stretch a little information a long way. We can draw reasonably accurate conclusions and make reasonably accurate predictions about specific individuals provided we have reasonably accurate knowledge of the category to which it belongs.

Research on category learning has addressed several issues and explored many areas including theoretical mechanisms of categorization, number of category learning systems, types of tasks used, and supervised versus unsupervised learning (for reviews, see Ashby & Maddox, 2005; Markman & Ross, 2003). The majority of category learning experiments have investigated *supervised category learning*. This approach operates explicitly and intentionally, informing people that there are categories in the stimuli they have been given, and providing feedback about the accuracy of their categorization decisions (Ashby & Maddox, 2005; Kruschke, 1992; Medin 1989; Medin & Ross, 1989; Medin & Schaffer, 1978; Nosofsky, 1986, 1988; Posner & Keele, 1968). Supervised category learning involves an external tutor who provides corrective feedback about category information of the items at hand. It is not difficult to recall instances of this happening in everyday experience, such as a parent explaining to a child that this animal is a cat, that dogs gnaw on bones, or that the candles and cake are for a birthday party, but supervised category learning does not describe how we acquire all category knowledge.

In contrast, there are many situations involving unsupervised learning, where categories are acquired without explicit feedback, or even without the knowledge that categories exist. *Unsupervised category learning* places the burden of learning the structure and members of categories entirely on the learner, who uncovers category structure without external directive feedback. Real world examples of unsupervised category information include learning linguistic classes (Taraban & Roark, 1996; Taraban & Kempe, 1999; Taraban, 2004, 2006), and general information (Clapper & Bower, 1994, 2002; Kaplan & Murphy, 1999; Kelly, Burton, Kato, & Akamatsu, 2001;

Love et al., 2004; Taraban & Hayes, 2000, 2001). For example, a parent might tell a child to give tuna to Mittens, the family cat, and a bone to Bowser, the family dog, without explicitly stating that cats like tuna and dogs like bones. In this case, the child acquires knowledge about the categories (cats and dogs) without the relationships between the categories and their features being explicitly stated; rather, the learner must infer the category structure.

There have been several theories that explain unsupervised category learning by appealing to the concept of memory limitations (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Clapper & Bower, 2002; Love et al., 2004). Some of these models locate the memory limitation within working memory (WM; Ashby et al., 1998; Love et al., 2004), yet very little research has been done to confirm that the WM system is the human memory system wherein these memory limitations arise. The little work that has been done has compared clinical and normal populations (Amos, 2000; Filoteo, Maddox, Ing, Zizak, & Song, 2005; Price, 2005, 2006) or developmental differences (Gureckis & Love, 2004). To date, no work has confirmed the existence of a connection between WM and unsupervised category learning in healthy adults. The next section reviews the unsupervised category learning literature and reviews the theoretical basis for WM as a key component in category learning when the learner does not receive explicit feedback about the category structure.

Unsupervised Category Learning

There have been two main areas of interest in unsupervised category learning, corresponding to two main experimental paradigms—i.e., classification learning and category structure learning. The first of these, classification learning, focuses on how people learn to categorize items, emphasizing the differences between categories. Typical unsupervised classification learning experiments require participants to sort instances into groups without feedback about classification accuracy (Ashby, Queller, & Berretty, 1999; Fried & Holyoak, 1984; Pothos & Chater, 2005; Spalding & Ross, 1994). This type of investigation examines knowledge of diagnostic feature dimensions for belonging to a given category, presenting participants with exemplars belonging to two or more categories based on feature values or combinations of feature values. Participants then sort the exemplars into groups based on their features. The organization of exemplars into categories is either defined by the experimenter (e.g., Ashby et al., 1999) and/or by the participant (Fried & Holyoak, 1984). Classification learning depends heavily on the use of learned exemplars to make categorization decisions, and results in knowledge about which features discriminate members of different categories (Yamauchi, Love, & Markman, 2002).

Category structure learning has many names in the research literature including schema abstraction (Elio & Anderson, 1981), comparison-based learning (Spalding & Ross, 1994), inference learning (Anderson, Ross, & Chin-Parker, 2002; Yamauchi et al., 2002; Yamauchi & Markman, 1998), similarity-based learning (Love, 2002), and category induction (Taraban, 2004, 2006; Taraban & Hayes, 2000, 2001). These studies

differ in how the experimental task was structured, what information was provided to participants, and what information was necessary to complete the task. What they all have in common is their focus on people's representation of the category structure: the relationship amongst features within a category. In unsupervised category structure learning, participants' knowledge of the category structure is examined by probing knowledge of the feature correlations within a category, often by querying the value of a missing feature. Participants are exposed to training exemplars from one or more categories, without those exemplars being labeled as belonging to one category or another, and their knowledge of the category structure is assessed based on their ability to predict the values of missing features (Billman & Heit, 1988; Billman & Knutson, 1996; Taraban, 2004, 2006; Taraban & Hayes, 2000, 2001), or to determine if a new item conforms to the structure of an existing category or categories (Love, 2002; Reber, 1967, 1989). Unsupervised category structure learning depends on the ability to form abstract representations of the relationships between category features and results in knowledge of the category prototype (Love, 2002), the category prototype being an abstract summary representation of the category defined by a family resemblance structure (Yamauchi et al., 2002; Yamauchi & Markman, 1998; Rosch & Mervis, 1975). Prototypes, in turn, are best learned if the categories are linearly separable, in other words, the categories could be differentiated based on the additive values of the features (Yamauchi et al., 2002; Yamauchi & Markman, 1998).

The main assumption regarding unsupervised category structure learning is that knowledge of within category structure results from experience of the items themselves.

Exactly how category knowledge arises from experience with category exemplars is debatable, as are the properties of the learning system responsible for category learning. Some researchers assert that category knowledge occurs automatically as a result of learning the exemplars themselves (Love, 2002). Other researchers have found that knowledge of the category structure results only when people focus on the category level (e.g., Fried & Holyoak, 1984) or when attention is directed to similarities amongst items through the distribution and frequency of the learning items (Taraban, 2004, 2006; Taraban & Hayes, 2000, 2001).

Several theories of unsupervised category structure learning have been proposed. The first of these explains this type of learning as the product of an implicit learning system. According to this view, representations are built up gradually, relying on a perceptual or procedural memory system (Ashby et al., 1998; Knowlton & Squire, 1993, 1996; Reber, 1967, 1989; Reber & Squire, 1994). The resulting representation of category structure is implicit because it resides in a non-verbalizable memory system (i.e., a perceptual and/or procedural memory system) and is independent of memory for the learning items (Knowlton & Squire, 1993).

The second theory proposes that unsupervised category learning operates via explicit iterative hypothesis testing (Billman & Knutson, 1996, Ashby et al., 1998). When people encounter an object or event, they form a hypothesis about the relationship between two or more of its features, then test that hypothesis when they encounter the next instance. If the hypothesis is supported, additional hypotheses are generated, related to the first. If the hypothesis is not supported, then a new hypothesis is formed and

tested. This iterative process is the same rule-based mechanism used in supervised category learning, with the additional requirement that participants must generate their own feedback regarding category assignment.

Clapper and Bower (1994, 2002) and Love (Gureckis & Love, 2003a, 2003b; Love, 2002; Love et al., 2004) provide a third theory for unsupervised category learning that falls somewhere between the other two. According to these researchers, category formation in unsupervised category learning occurs when people notice that there are significant differences in the features of the items they experience. The unexpected values cause the learner to invent a new category on the fly (Clapper & Bower, 1994, 2002) or add additional representational clusters (Gureckis & Love, 2003a, 2003b; Love, 2002; Love et al., 2004) to account for the new instance (category invention; Clapper & Bower, 1994, 2002).

Support for the category invention hypothesis in unsupervised category learning was demonstrated by Clapper and Bower (1994, 2002), who found that correlated features alone were not enough to result in category learning. Participants were able to acquire knowledge of the feature patterns of two categories, but only when the unlabeled stimuli were presented in a blocked sequence containing a series of items from one category before items from the other category were presented. When the stimuli were intermixed, participants did not learn the features associated with the categories and treated the items as though they all belonged to the same category. Incremental category learning models (e.g., Medin & Schaffer, 1978) are unable to explain this pattern-sequence effect. Clapper and Bower (1994, 2002) explained the pattern-sequence effect

as the result of early instances establishing the range of feature variability that characterizes a category, with greater initial variability (i.e., intermixed presentation) making it less likely that subsequent items will result in the creation of a new category.

Amongst the many models of category learning, there are very few that can account for both supervised and unsupervised learning as well as classification and category structure learning: Anderson's (1990, 1991) rational model and SUSTAIN (Supervised and Unsupervised Stratefied Adaptive Incremental Network, Gureckis & Love, 2003b; Love et al., 2004) are two that can. Both models employ incremental complexity mechanisms consistent with the category invention hypothesis. These models are parsimonious in the sense that they will assume no more categories than are necessary to represent experience; initially assuming that objects conform to a simple category structure, and adding additional representational units only when needed to capture more complex category structures in the environment.

The rational model (Anderson, 1990, 1991) assumes that people maximize their categorization behavior to reflect the structure of the environment. Therefore, the model is sensitive to the expectations and experiences of the individual. If the person expects X categories, then instances are assigned to those X categories. If the person has no reason to assume that there are multiple categories (e.g., in unsupervised category structure learning), then the system is less likely to produce new categories to accommodate the learning items and the production of new categories depends almost entirely on the number of items seen. Anderson's (1990, 1991) rational model assumes perfect memory; people have perfectly veridical access to all stored exemplars. This perfect memory over-

predicts human performance and cannot account for the pattern-sequence effect. Though the model is sensitive to order effects based on the frequencies of training exemplars, it is unable to explain the pattern-sequence effect because it tends to predict that people will form multiple categories given significant variation in feature values, regardless of whether the items are presented in a mixed or blocked order.

Clapper and Bower (2002) modified Anderson's (1990, 1991) rational model, incorporating a parameter reflecting the limits of the memory system, as an explanation for the pattern-sequence effect. In Clapper and Bower's (2002) implementation of the rational model, imperfect memory reflects the fact that human learners do not have error-free access to all stored exemplars and their features. Rather, there are severe memory limitations that permit the accurate recall of a relatively small proportion of an item's features. When stimuli are drawn from two categories presented in a mixed sequence, sufficient feature forgetting will result in not enough feature information being recalled to cause the model to invent a new category when an exemplar from a new category is encountered. For example, when a Category B item follows a Category A item, if forgetting causes the loss of too much of the critical feature information of the Category A item, the model cannot distinguish between the two, causing it to group the items into a single category. The resulting category representation now has a greater amount of feature variability (from those feature values that were recalled), making it less likely that future Category A and Category B items will cause the creation of a new category. When stimuli are blocked by category (e.g., several Category A instances before any Category B instances are seen), even relatively substantial forgetting will not prevent the model

from forming a representation of Category A over the course of several exemplars, a representation complete enough to differentiate a new exemplar drawn from Category B.

The SUSTAIN model (Gureckis & Love, 2003b; Love, 2002, 2005; Love, et al., 2004) is very similar to the rational model and also provides an explanation for the pattern-sequence effect. The SUSTAIN model operates by recruiting representational "clusters" as needed. More categories and/or more complex category structures recruit more clusters to represent the added complexity. Clusters can represent anything from individual exemplars (which means that SUSTAIN behaves like an exemplar model), to groups of features within categories, to representations of an entire category (which means that SUSTAIN behaves like a prototype model). In unsupervised category learning, cluster recruitment is based on unexpected feature values (i.e., surprise). Surprise, in turn, depends on the similarity of the present item to existing clusters and the magnitude of the difference necessary to recruit a new cluster. Similarity is calculated from the interaction of selective attention to critical features and prior variability on those features. When more attention is tuned to relevant feature dimensions and when there is less variability of previous exemplars on those dimensions, smaller differences on those feature dimensions will result in the item being judged as significantly different than existing categories. An exemplar that is sufficiently different than existing clusters on those features results in the recruitment of an additional cluster, where the magnitude of the difference required to recruit a new cluster is governed by a threshold parameter in the model.

SUSTAIN is sensitive to order effects, such as the pattern-sequence effect, because initial exemplar variability interacts with the selective attention and threshold parameters to produce more or less cluster recruitment. A homogeneous set of exemplars tunes the attentional parameter so that a little difference in feature values on later exemplars results in the recruitment of an additional cluster. If, on the other hand, the initial set of exemplars was highly variable, then the attentional parameter would be desensitized, requiring greater differences to result in the recruitment of additional clusters. Modeling simulations were able to account for differences in unsupervised classification and category structure learning by adjusting only the threshold parameter (Gureckis & Love, 2003a). Unsupervised category structure learning resulted in a lower setting for the threshold parameter, which meant that the items were treated as more similar. This was explained as a general task-related difference where the formation of abstract representations (category structure learning) depends on the tendency to perceive items as more similar, whereas the ability to sort items into categories (classification learning) depended on the tendency to see the differences between exemplars.

WM and Category Structure Learning

In the remainder of this introduction, I will review research that suggests that WM is implicated in many of the processes essential for unsupervised category structure learning and will develop an argument for the hypothesis that WM plays a role in unsupervised category learning in general and in explaining the pattern-sequence effect in

particular. While several models and studies posit WM as an explanation for category learning effects, only a handful actually explored the connection empirically.

Research has revealed a critical role of WM in many areas of human cognitive activity involving reasoning (Kyllonen & Christal, 1990), problem solving (Hambrick & Engle, 2003), learning (Baddeley, 1986), memory (Baddeley, 1986; Baddeley & Hitch, 1974; Cowan, 1995), comprehension (Engle, Carullo, & Collins, 1992; Just & Carpenter, 1980), attention (Bleckley, Durso, Crutchfield, Engle, & Khanna, 2003; Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001), and awareness (Baddeley & Hitch, 1974; for overviews & reviews, see Baddeley, 1986; Cowan, 1995; Miyake & Shah, 1999). In addition to these general WM functions, there are several more that theoretically establish a role for WM in unsupervised category structure learning.

The memory parameter in Clapper and Bower's (2002) extension of the rational model affects the probability of estimating a given feature value for a feature dimension on a known exemplar. Though they briefly describe the memory limitation as resulting from "severe constraints on its short-term memory" (p. 920), Clapper and Bower (2002) do not provide any justification for locating the memory limitation to this memory system, and the processes behind the memory limitation are not specifically defined. It is possible that the memory constraints may reflect the ability to retrieve instances accurately from long-term memory, to maintain multiple instances in an active state for comparison, and/or to carry out the comparison process. WM has been shown to be involved in each of these functions. Bunting and Cowan (2005) demonstrated that WM is involved in retrieving the relevant feature information from memory, especially in the

face of distraction or interference (see also Conway & Engle, 1994). WM is important for resolving and preventing interference (Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999; Jha, Fabian, & Aguirre, 2004; Li, 1999), a critical aspect of correctly labeling instances and recalling the feature values of related instances. McElree (1998) found a retrieval advantage for items that belonged to a recently studied category, indicating that category membership may aid retrieval. The capacity of WM defines the number of instances or chunks that can be mutually active and available for comparison (Baddeley, 1986; Hummel & Holyoak, 2003). The comparison process may be between several stored instances as suggested by exemplar models (Kruschke, 1992; Nosofsky, 1986, 1988) or between the new item and a stored abstraction or prototype as suggested by prototype models (Posner & Keele, 1968; Smith & Minda, 2002).

Similarly, Elio and Anderson (1981) argued that category generalizations are formed more easily when their instances are "simultaneously available in WM for patterns" (p. 402), yet knowledge of those generalizations does not have to be explicit. As category structure learning operates best when the categories are linearly separable, represented by a prototype (i.e., they are represented by a single abstraction; Yamauchi et al., 2002), it follows that WM plays an important role in category structure learning to the extent that it depends on the formation of prototypes. In addition to representing the tolerance for dissimilarity, SUSTAIN's threshold parameter also represents memory in the model (Gureckis & Love, 2004). As a memory parameter, it represents the ability to bind features together into a cluster and, as subsequent similar items are encountered and represented by the cluster, it becomes an abstract representation similar to a prototype.

The ability to bind features together into clusters depends on the functioning of the hippocampus (Love & Gureckis, 2004), a part of the WM system. This is similar to Hummel and Holyoak's (1997, 2003) description of the biological limit of WM capacity as the number of chunks (i.e., clusters in SUSTAIN) that can be simultaneously active and mutually distinct. The threshold parameter represents memory in SUSTAIN and has been shown to capture the developmental trajectory of infant's sensitivity to within category feature correlations (Gureckis & Love, 2004).

WM is implicated in the control of attention (Bleckley et al., 2003; Engle, 2002; Kane et al., 2001). Selective attention plays a major role in category learning, both for directing attention to relevant feature dimensions (Kruschke, 1992; Wagar & Dixon, 2005) and ensuring that only relevant information is active in WM (Hasher & Zacks, 1988; Hasher et al., 1999). Selective attention has been implicated in both category structure learning (Love et al., 2004) and classification learning (Nosofsky, 1986; Nosofsky et al., 1994; Spalding & Ross, 1994). Brooks, Kempe, and Sionov (2006) found that executive attention was associated with learning grammatical categories in second language learning.

Neuropsychological studies with clinical populations and using brain imaging technology indicate that explicit rule-based learning depends on frontal lobe functioning (Ashby et al., 1998; Filoteo, Maddox, Ing, Zizak, & Song, 2005; Price, 2005, 2006; Tracy et al., 2003). Explicit rule testing requires WM to generate a hypothesis relating feature values, maintain the hypothesis across instances, evaluate the hypothesis, and then store

or revise it as needed. If participants are engaging in explicit hypothesis testing as Billman & Knutson (1996) suggest, then WM is certainly involved.

The present study tests the hypothesis that unsupervised category learning depends on WM resources. The pattern-sequence effect is used as a means to test WM effects in unsupervised category structure learning. Both SUSTAIN and Clapper and Bower's (2002) implementation of the rational model are consistent with the hypothesis that WM is involved in unsupervised category structure learning. SUSTAIN's account of the difference in infant's unsupervised category learning hinged upon the threshold parameter in the model, with higher values required to successfully acquire knowledge of the category structure of the two categories (Gureckis & Love, 2004). This parameter represents tolerance for dissimilarity and also represents memory in the model. It is theoretically tied to hippocampal function (Gureckis & Love, 2004), meaning that it is actually a WM parameter. As Clapper and Bower (2002) stated, the pattern-sequence effect will be obtained when sufficiently low expectations for multiple categories is combined with sufficiently poor short-term memory. Expectations can be manipulated through experimental instructions (unsupervised learning sets a low expectation of multiple categories). I propose that poor short-term memory in this case refers to the limited capacity of WM. The following experiments empirically test the prediction that WM resources are involved in the pattern-sequence effect by extending Clapper and Bower's (2002) hypothesis to predict that greater WM capacity will produce a greater unsupervised category learning and a larger pattern-sequence effect than will low WM capacity.

Just as there are multiple theories of WM, there are multiple ways of measuring WM span including reading span (Just & Carpenter, 1992; Engle, Tuholski, Laughlin, & Conway, 1999), operation span (Turner & Engle, 1989; Engle et al., 1999), counting span (Engle et al., 1999), and spatial span (Shah & Miyake, 1996), among others. All of these measures are dual-task procedures, requiring participants to simultaneously store and process information. Engle et al. (1999) demonstrated that, despite the variety of processing tasks used in the operation span (evaluating mathematical equations), reading span (reading sentences), and counting span (counting the number of targets in a field of distracters) and material to be remembered (words and numbers), these tasks all load highly on the same factor whereas general fluid intelligence measures and simple memory span measures (single-task short-term memory measures) load on separate factors. An automated version of the operation span task (Aospan; Unsworth, Heitz, Schrock, & Engle, 2005) was chosen as the measure of WM capacity. Aospan correlates well with other measures of WM span; $r = .499$ with operation span and $r = .603$ with reading span. These values are similar in magnitude to previous work (e.g., Engle et al., 1999, found correlations between three measures of WM ranged from $r = .32$ to $r = .51$). Aospan also loaded on the same factor as the operation span and reading span tasks (Unsworth et al., 2005). Aospan has the advantage of being delivered and scored automatically by computer. This measure of WM requires participants to evaluate the correctness of a series of mathematical operations and remember the letters that follow each operation (e.g., *is (1*2)+1=3? P*). The length of the series varies from 3 to 7 operation-letter pairs, and participants must recall the letters in the correct serial order.

Overview of Experiments

Experiment 1

The purpose of Experiment 1 was to validate an experimental method capable of producing the pattern-sequence effect using a modification of Clapper and Bower's (2002) method. The experimental paradigm used in Clapper and Bower (2002) was altered for two reasons. First, I wanted to make the task sufficiently difficult to reveal the role of WM in unsupervised category structure learning using the pattern-sequence effect. WM involvement will not be apparent under learning conditions that are too easy; processing demands would be insufficient to tax WM and reveal the differences between high and low WM span participants. It is possible that the robust pattern-sequence effect reported by Clapper and Bower (1994, 2002) was due to the very low processing demands required by the task, enabling lower WM span participants to benefit from the blocked presentation sequence while demonstrating no learning in the mixed presentation condition. Simultaneous presentation of all of the features (Clapper & Bower, 1994) or instant access to any feature dimension (Clapper & Bower, 2002) places a low processing load on WM. Therefore, the method used must impose a greater strain on WM resources while still producing a reliable pattern-sequence effect to reveal WM differences in the blocked presentation condition. The interaction of WM and the pattern-sequence effect is described in more detail in Experiment 2. Second, a new generalization manipulation was added to assess inference learning of the category structure, similar to the task used in earlier experiments by Taraban (Taraban, 2004, 2006; Taraban & Hayes, 2000, 2001).

Clapper and Bower (2002) presented participants with a series of 32 trees, each with 12 features that could assume one of four values. Participants had access to all 12 features of each item and were given 24 s to learn the feature values, but could only view one feature at a time (the other features were masked). The participants were not told that the trees belonged to two categories based on their feature values on nine of the features. The remaining three features were unrelated to group membership. The feature values of the two categories are represented as *111111111xxx* and *22222222xxx*. The primary measure of category knowledge was assessed by viewing times for the three idiosyncratic features. High accuracy on all 12 features coupled with longer viewing times on the uncorrelated features indicated that the participant learned the category structure and was able to direct his/her attention to learning those feature values that were not predicted by category membership.

Similar to Clapper and Bower (2002), participants in the present study learned information about a series of plant samples. Latin plant names (e.g., *Acer*) were paired sequentially with one of the features of the plant sample (leaf shape, root type, stem type, flower type). Each feature assumed one of two values, and the categories were formed by the feature values and were perfectly reliable. Category A plants always had pointed leaves, taproots, smooth stems, and headed flowers. Category B plants always had rounded leaves, fibrous roots, woody stems, and spiked flowers. Participants worked individually at a computer. A learning trial began with a query to the participant for the feature value of a plant sample (e.g., *Acer, stem is*:). The participant used the computer keyboard to enter his or her response (e.g., *woody*), after which the computer provided

feedback about the accuracy of the response (correct or incorrect) and displayed the correct answer.

The memory effect specified by Clapper and Bower (2002) does not directly control the acquisition of abstract category information. It controls the ability of the individual to recall the features of items. It is critical, therefore, to establish whether the pattern-sequence effect results in the acquisition of abstract category knowledge as Clapper and Bower (1994, 2002) propose, or merely facilitates the acquisition of item-feature knowledge, i.e., learning the features of each exemplar separately from the others. There are several mechanisms by which item memory may account for the pattern-sequence effect, though the particular mechanism is not important for the present study. The pattern-sequence effect may create less interference during encoding, placing lower demands on WM resources to correctly match the features to the items and making it more likely that the exemplars are correctly encoded and retrieved (Bunting & Cowan, 2005; Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999; Jha, Fabian, & Aguirre, 2004; Li, 1999). Alternatively, the blocked condition benefits from massed practice, whereas the mixed condition does not.

The first main difference between the procedure used here and that of Clapper and Bower (2002) is that the present method employed a more restricted sequential presentation format. Like Clapper and Bower (2002), participants saw the features of one exemplar before moving on to the next. Unlike Clapper and Bower (2002), only one feature was displayed for an exemplar at a time, and participants were not able to select which feature to view and were not able to revisit a feature. The second main difference

is that this study did not present all of the features for each exemplar during learning. One feature was withheld during learning for each plant and used to assess category structure learning during a subsequent test phase. Participants saw all of the feature values for the categories used, but not together on any one exemplar. An additional set of transfer items was included to reduce experimenter-demand effects for generalization, which may have been present in earlier studies (See Appendix E & F).

Experiment 2

Experiment 2 tested the hypothesis that WM is involved in unsupervised category structure learning by testing whether the pattern-sequence effect depends on WM capacity. Aospan (Unsworth et al., 2005) was used to identify high and low WM span participants which were then compared in an extreme groups design. The performance of these groups was compared using the experimental task used in Experiment 1. The rational model and SUSTAIN are both consistent with the hypothesis that the pattern-sequence effect relies on sufficient memory resources. Based on the formulations of these models and the hypothesized functions of WM within those models, WM should interact with the order of presentation manipulation that gives rise to the pattern-sequence effect. High WM high span participants were predicted to drive the pattern-sequence effect, exhibiting significantly greater learning in the blocked condition relative to the mixed and control conditions. As difficulty is a relative matter, low WM span participants were expected to exhibit a smaller pattern-sequence effect. If the task was sufficiently easy, even low WM span participants would exhibit an advantage of the

blocked presentation over the mixed presentation because low WM span participants lack the WM resources to utilize the additional structure provided by the presentation order manipulation. High span participants, on the other hand, possess sufficient WM to discover the structure, and will demonstrate a robust pattern-sequence effect. Thus, the presence of an interaction of WM and the ordering manipulation would support the theoretical assertion that WM is a significant factor in unsupervised category learning. No specific prediction was made for the high WM span participants in the mixed condition. If the task was sufficiently easy they might demonstrate some category learning.

The alternative hypothesis posits that, while WM is required for item memory, i.e., learning item-feature pairs (a form of paired-associate learning), category learning depends on an implicit learning system (Ashby et al., 1998; Knowlton & Squire, 1993). Knowlton and Squire (1993) demonstrated that amnesic participants were no different from normal controls in their ability to acquire knowledge of a family resemblance category structure, but were impaired in their ability to recognize the specific training items used. According to this hypothesis, greater WM capacity will result in faster learning and greater mastery of the training items but will not result in any greater learning of the category structure and will not enable high WM span participants to perform any better than low WM span participants on novel items, even if they are consistent with the learned category structure.

CHAPTER II

EXPERIMENT 1

The present study combined the category induction methodology of Taraban and Hayes (2000, 2001) and the stimulus order manipulation used by Clapper and Bower (1994, 2002) that gave rise to the pattern-sequence effect to investigate the effects of initial exemplar exposure on unsupervised category learning when participants view one feature of an exemplar at a time (Taraban & Hayes, 2000, 2001). Participants in Clapper and Bower (2002) had access to all 12 features of an exemplar, 9 of which were perfectly informative of category membership, and could freely select which feature to view. This method places relatively little strain on WM. In order to reveal the role of WM in the pattern-sequence effect, the task must be sufficiently difficult. If the task is too easy, even low WM span participants will demonstrate the pattern-sequence effect. Likewise, if the task is too difficult, even those with high WM resources will fail to show the pattern-sequence effect. The two goals of Experiment 1 were to adjust the difficulty of the task, making it more taxing on WM than the task employed by Clapper and Bower (2002) by lowering overall performance, and to incorporate three tests of unsupervised category learning that could not be passed by recall or recognition of studied items (exemplar-feature pairs).

Clapper and Bower (1994, 2002) proposed that the pattern-sequence effect produces surprise that then causes the creation of a new category to represent the new category (the category invention effect). According to this position, the category invention effect results in the formation of abstract category representations. In order to

test this theory, the present study utilized a more restrictive serial presentation format than Clapper and Bower (2002) and added a test of category knowledge that included previously unseen item features that tested category knowledge in the absence of item memory (Taraban & Hayes, 2000, 2001). If the pattern-sequence effect results in the acquisition of category knowledge, then it should facilitate performance on the familiar items seen during learning as well as the new items because they conform to the category structure. If, on the other hand, the pattern-sequence effect facilitates item-feature pair memory, perhaps by reducing interference during encoding or retrieval of the exemplars resulting in more robust memory traces (Bunting & Cowan, 2005; Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999; Jha, Fabian, & Aguirre, 2004; Li, 1999), then it should facilitate performance on the familiar items seen during learning but not on the new items because, even though they conform to the category structure, they were not practiced during learning.

Four main tests of category knowledge were used. The first test of category knowledge tested participants' knowledge of unseen features of familiar exemplars. The second test of category knowledge used combinations of familiar features and feature values with new exemplars. The third test of category knowledge examined differences in response times for seen versus unseen items. The final test was an indirect measure of category knowledge using ratings that gauge participants' perceptions of the difficulty of learning the items.

Method

Participants

Participants were 75 undergraduate students at Texas Tech University who participated for course credit. Participants achieving less than 34% correct on the base-learning items were excluded from the analysis. This cutoff was established to eliminate participants who were not engaged in the task and had not learned the two response options for each feature and represents one standard deviation below the mean in a pilot study (see Appendix F). Nine participants failed to meet this criterion, five in the mixed condition and four in the control condition, resulting in 66 participants overall; 22 in each condition.

Materials

The stimuli were 24 Latin plant words (exemplars) selected from a dictionary of plant terms (Coombes, 1985), with four feature dimensions (root type, stem type, leaf type, flower type), and two feature values for each of those dimensions. A pilot study suggested that learning is possible with these exemplars, features, and feature values (see Appendix F). The 12 plant word exemplars were matched between categories on average word length (5.67 letters overall, 5.75 for Category A, 5.58 for Category B), number of syllables (2.33 overall, 2.33 for Category A, 2.33 for Category B), first letter, last letter, and last syllable. Thus, there were no known cues associated with the plant names themselves that would reliably signal the underlying category membership of the plants.

The three experimental conditions used the same 24 Latin plant name exemplars, four feature dimensions, and eight feature values, but differed in how those feature values were combined and the presentation order of the feature combinations. In the blocked and mixed conditions, a tacit categorical organization was formed by which feature values are associated with a plant name. In these two conditions, feature values are perfectly correlated and completely described all members of the category where Category A is represented by the feature value pattern *1111*, and Category B by the pattern *2222*. Category A exemplars were always smooth stem, taproot root, pointed leaf, and headed flower. Category B exemplars were always woody stem, fibrous root, rounded leaf, and spiked flower. Participants were not shown the category labels. The feature values in the control condition were intentionally uncorrelated. There are 16 possible feature value combinations (2 stems X 2 roots X 2 leaves X 2 flowers). Eight of these were selected for the control condition such that each feature pair occurred exactly twice for any pair of features (e.g., taproot and pointed leaves occurred exactly twice). See Appendix A for the complete stimulus set.

Procedure

Participants were told that the study was a calibration study to establish the learning difficulty of the materials for use in future studies. This cover story was chosen in an effort to bias participants away from thinking that there was a trick to the experiment. Post experimental questionnaires from two pilot studies revealed that participants tried to find relationships between the feature values and various aspects of

the Latin names, including the first letter and last letter, and the answer sequence (all of which were either counterbalanced or randomized). Evaluative instructions are typically employed to ensure that participants process the information without searching for other patterns in the stimuli and are frequently used in studies of implicit or incidental memory and learning (Love, 2002; Goshen-Gottstein & Kempinsky, 2001). To support this cover story, participants provided difficulty ratings at four points in the procedure: halfway through the learning phase, during the test phase, during the first block of transfer trials, and during the transfer-test phase.

Throughout the experiment, participants were presented with only three of the four features for each exemplar, though they were exposed to both feature values of all four features over the course of the experiment. For any given plant name, these were the same three features for all blocks in both pretraining and learning phases (base-learning items); the remaining feature from the learning phase were reserved for one of the generalization tests (base-generalization items). Likewise, the same three features were presented during the first two transfer-phase blocks for each of the transfer-learning items, the remaining feature also reserved for a test of category knowledge in the transfer-test phase (transfer-generalization items). Just like the procedure in Clapper and Bower (2002), participants saw all of the available features for an exemplar before moving on to the next exemplar, but unlike the Clapper and Bower procedure, they were not able to revisit an item. Participants viewed each exemplar-feature pair, one at a time, and entered a response before the next pair was presented. For example, a participant might have seen the sequence of *Azium, stem is:*, then *Azium, root is:*, then *Azium, leaf*

is:, before seeing the next exemplar (*Azium, flower is:*, would not appear in the learning phase). This method was more stringent than that used by Clapper and Bower (2002), who presented participants with all of the features of an exemplar and allowed them to decide which feature to view. In the present study, participants were exposed to the complete category structure but not for any single exemplar, meaning that performance on the generalization items could not have been based on rote memorization of those items, but rather on knowledge of the category structure, because the generalization items were composed of novel combinations of familiar exemplars, features, and feature values.

The three experimental conditions were adapted from Clapper and Bower (1994, 2002) and included a control condition with fixed randomized feature values for each exemplar (the control condition), and two experimental conditions containing exemplars belonging to one of two categories based on their correlated feature values (the blocked and the mixed conditions). The blocked condition differed from the mixed condition only in the presentation order of the initial exemplars. The experimental sequence, as well as the stimuli presented and data collected, for each of the experimental conditions are presented in Appendix A. Participants worked individually at computers and participated in groups of one to three. The main experimental task was a fact-learning task followed by a post-experimental paper-and-pencil questionnaire. The instructions to participants are included in Appendix B.

Based on the outcome of Pilot Study 2, the pattern-sequence effect was expected to produce a large effect size (Cohen, 1988). Power analysis using GPower (Faul &

Erdfelder, 1992) indicated that 66 participants (22 per group) were required for a power of .80.

Pretraining Phase

For all conditions, the pretraining phase consisted of three blocks of the learning items, for a total of 144 trials (3 blocks of 16 exemplars each with 3 features). The number of pretraining items and trials was selected based on the procedures employed by Clapper and Bower (1994, 2002) and by the results from Clapper and Bower's (1994) investigation of the impact of the amount of pretraining required for the pattern-sequence effect and was subsequently examined in two pilot studies (see Appendix E & F). Clapper and Bower (1994) used 16 pretraining exemplars and Clapper and Bower (2002) used 12 pretraining exemplars. The results of Clapper and Bower (1994; Experiment 3) indicated that there was a significant unsupervised learning advantage for a pretraining set of 12 over sets of 8 or 4. The stimulus set size of 16 exemplars used here contained fewer features to learn, but the results of Clapper and Bower (1994) indicated that more pretraining exemplars lead to greater category learning. A pilot study utilizing 16 pretraining exemplars also revealed more evidence for learning (Appendix E) than a pilot study utilizing 12 pretraining exemplars (Appendix F).

Participants in the blocked condition completed the items in a modified blocked design, completing three repetitions of the 24 Category A items (8 exemplars, each with 3 features) before completing the three repetitions of the 24 Category B items (8 exemplars, each with 3 features). Block 1 consisted of two replications of the 24 Category A items.

Block 2 consisted of one replication of the 24 Category A items followed by one replication of the 24 Category B items. Block 3 consisted of two replications of the 24 Category B items. This had the effect of presenting participants with 24 exemplars from Category A before they are exposed to the first Category B exemplar. The order of exemplar presentation was randomized within each block. Participants in the mixed condition saw three blocks consisting of the 24 Category A and 24 Category B items intermixed in a pseudo-randomized manner such that no more than three exemplars from the same category appeared in a row. Participants in the control condition completed the same number of trials but using stimuli from the control stimuli set. The order of exemplars was randomized as in the mixed condition.

In addition to the controls on exemplar order, all items in all conditions were presented in a randomized block design. Each exemplar was presented with three of its features sequentially in random order before the next item was displayed. The computer displayed a request for the feature value of an exemplar in the center of a computer monitor (e.g., *Azium, stem is:*). Participants used the computer keyboard to enter their response and the computer displayed feedback after each item, informing the participant whether their response for the feature value was correct or incorrect and indicating the correct response. This applied to all feedback throughout the experiment. Additional feedback was provided after the three trials for each exemplar indicating how many of the three items the participant got correct. At the beginning of the experiment, the experimenter informed participants that misspellings or extra spaces would result in the computer counting their response as incorrect.

Learning Phase

For all conditions, there were 96 learning trials organized into 2 blocks of 16 exemplars, each with 3 features. As in the pretraining phase, participants saw a request for a feature of an exemplar (e.g., *Acacia, stem is:*), and they entered their response using the computer keyboard. Feedback was provided for all participants for all items as described above. Exemplar order in all conditions was presented in pseudo-random order such that no more than three members of the same category were presented in a row. Item presentation in all three conditions was presented in the randomized block procedure described in the pretraining phase; each exemplar was presented sequentially with three of its four features, the fourth feature (base-generalization items) being reserved for the test phase.

In the second block of trials, participants completed a difficulty rating of each item after they receive response accuracy feedback. This was on a 7 point Likert scale, anchored at *not at all difficult* and *extremely difficult*. Participants used the computer keyboard to enter their response (a number from 1 to 7) before moving on to the next item.

Test Phase

The test phase was executed exactly as the second block from the learning phase except that each exemplar was presented with two features seen during the learning phase (base-learning items) and the remaining feature not seen during learning (base-

generalization item). The decision not to present all four features was made based on the logic that presenting three features consistently throughout the experiment would minimize the chance that participants notice or search for a pattern in the stimuli during the test because they are suddenly confronted with four features for an exemplar when previously they had only seen three features for any of them. Ideally, the test of category knowledge reflects knowledge gained prior to the test itself, and presenting four features instead of three might function as an experimenter demand effect, cuing participants to explicitly compare old instances in an effort to deliberately locate a pattern. The test phase appeared without special instruction after the final block of learning trials and included the same difficulty-rating task employed on block 2 of the learning phase.

Transfer-Learning Phase

The transfer-learning phase introduced eight new exemplars, four conforming to the category structure used for each of the two categories in the blocked and mixed condition learning stimuli, and was the same for all three conditions (see Table 17 and 15). There were two blocks wherein participants saw three of four features for each exemplar (transfer-learning items) presented pseudo-randomly in a randomized block design identical to that used in the learning phase. Feedback was provided following each item and after the three trials for each exemplar. Participants completed the difficulty-rating task on the first block only.

Transfer-Test Phase

The transfer-test phase followed the second block of transfer-learning trials and mirrored the procedure used in the test phase. As in the test phase, one of the transfer-learning items for each exemplar from the transfer-learning phase was replaced by the unseen transfer-generalization item (see Table 17 and 15). Feedback was provided following each trial and after the three trials for each exemplar. Participants again completed the difficulty-rating task following each item.

Post-Experimental Questionnaire

Following the fact-learning task, participants completed a funneled debriefing questionnaire (see Appendix C) designed to measure awareness of the category structure and probe for the type of learning strategies employed by participants. Participants completed the first two questions before moving onto the next question. Between each question (except 1 and 2), participants showed the questionnaire to the experimenter who ensured that the participant answered the question and asked for clarification if the participant's handwriting or statements were unclear. This measure was primarily included as a manipulation check on the pattern-sequence manipulation to assess how much awareness participants had for the category information and when and how they acquired it.

Debriefing

All participants were fully debriefed and instructed not to discuss the experiment with other students so as to not foil the cover story for future participants (see Appendix B).

Hypotheses

There were several hypotheses regarding learning and the effect of stimulus presentation order. Figure 1 presents the general predicted pattern of learning and generalization accuracy for both the base and the transfer exemplars. Figure 2 presents the general predicted pattern of learning and generalization response times for both the base and the transfer exemplars. Figure 3 presents the predicted pattern for the difficulty ratings for the base and transfer exemplars.

Hypothesis 1

There will be greater learning during the learning phase in the blocked than the mixed and control conditions. This is the pattern-sequence effect (Clapper & Bower, 1994, 2002). This hypothesis is presented in Figure 1 as the difference between the contrast, mixed, and control conditions for the learning items (filled triangles). Accuracy on the base-learning items will be above chance during the transfer-test phase for participants in the blocked condition. The blocked condition will outperform the mixed condition and control conditions. The mixed condition may or may not result in significantly greater learning than the control condition. Clapper and Bower (1994)

found evidence of learning in the mixed condition relative to the control condition, but Clapper and Bower (2002) did not.

Hypothesis 2

More learning will occur during the transfer-learning phase for the blocked condition than the mixed and control conditions. This is another extension of the pattern-sequence effect and follows the same pattern in Figure 1 as Hypothesis 2, though the line may be shifted upwards reflecting higher overall accuracy throughout the transfer phase. Participants who have acquired the category structure will have an advantage when learning the new exemplars. Accuracy on the transfer-learning items will be above chance during the transfer-test phase. The blocked condition will outperform the mixed condition and control conditions. The mixed condition may or may not perform above chance or outperform the control condition. Pilot studies suggest that accuracy for the control condition will not differ significantly from chance.

Hypothesis 3

Participants in the blocked condition will achieve above chance accuracy on the base-generalization items and the transfer-generalization items and will perform better than participants in the mixed and control conditions. This hypothesis is an extension of the pattern-sequence effect that assesses category knowledge for items that conform to the category structure but were not experienced during learning. This hypothesis is presented in Figure 1 as the difference between the contrast, mixed, and control

conditions on the generalization items (open triangles). Pilot work and previous research (Clapper & Bower, 1994) suggests that participants in the mixed condition are capable of learning the category structure, but at a much slower rate than participants in the blocked condition. They may perform above chance, but they should not outperform the blocked condition. Participants in the control condition are expected to perform at chance on the base-generalization items because even though their stimuli do not contain reliable feature correlations, they are expected to learn the two feature values for each feature.

Hypothesis 4

Participants in the blocked condition should exhibit faster response times for the new items (base-generalization and transfer-generalization items) than participants in the mixed and control conditions. This hypothesis is presented in Figure 2 as the difference between the three conditions on the generalization items (open triangles). Taraban (2006) found reliable category learning was associated with faster response times to the base-generalization items. The mixed condition may exhibit faster response times than the control condition depending on the amount of learning, but they should not outperform the blocked condition.

Hypothesis 5

Difficulty ratings for the base-learning, base-transfer, transfer-learning, and transfer-generalization items will be lower in the blocked condition than the mixed and control conditions. This is another test of the pattern-sequence effect. Participants who

have acquired the category structure will find new items that conform to a known structure less difficult to learn than participants who have not acquired the category structure. This is presented in Figure 3 as the difference between blocked, mixed, and control condition difficulty ratings. The greater the learning in the mixed condition, the greater the difference between the mixed and control condition and the more likely that the difficulty ratings in the mixed condition will resemble the blocked condition.

Data Screening

Latency scores were first normalized using natural log transformations of the raw response times (in milliseconds) for each of the response time dependent variables. Data were then examined for outliers and violations of normality using SPSS v15.0 Explore. Thirteen extreme scores were identified on eight of the dependent variables. Due to the small sample size, it was not feasible to exclude these cases. Instead, the extreme scores were Winsorized, the extreme scores were replaced with the next less-extreme score, separately for each dependent variable (Tabachnick & Fidell, 2001). A total of 20 replacements were made, accounting for no more than 6.06% on any single variable, and less than 4% of the data overall for these eight variables.

The normal and detrended PP plots of the transformed data set were visually evaluated for non-normality and exhibited no significant departures from normality. The normal PP plots were very linear for all dependent variables and the detrended PP plots contained no points outside of .7 deviations across all dependent variables.

Levene's test of error variances revealed several violations of equality of error variances on the following dependent variables: base-learning accuracy, transfer-learning accuracy, transfer-generalization accuracy, and base-generalization latency. These violations were not considered serious because MANOVA is robust to this type of violation especially when a large sample with equal cell sizes and two-tailed significance tests are used for the analysis (Tabachnick & Fidell, 2001).

Questions 4 and 5 on the post-experimental questionnaire were included to ensure that participants did not have substantial knowledge of the plant words used in this study that might conflict with the experimental materials. Technical problems resulted in the loss of data for Questions 4 and 5 for 12 participants: 5 from the blocked condition, 4 from the control condition, and 3 from the mixed condition. The number of Latin plant names listed as familiar are presented in Table 2. The Latin plant names listed as familiar, and the number of times they were mentioned by participants, are presented in Table 3. There were no significant differences between conditions in the number of participants identifying one or more of the Latin plant names as familiar, X^2 (2) = .337, p = .845, and the total number of plants rated as familiar was deemed acceptable.

Statistical Analyses

Replicating the pattern-sequence effect means that there should be greater learning in the blocked than the mixed or control conditions. The mixed and control conditions may or may not differ. Clapper and Bower (2002) found no difference

between the mixed and control conditions, but Clapper and Bower (1994) found that the mixed condition did perform slightly better than did the control condition.

Four assessments of category learning were used in the present study. Two were behavioral measures: accuracy, latency. The difficulty ratings were collected at select points throughout the experiment and measured participants' subjective evaluation of the task. The fourth was a post-experimental questionnaire included to assess explicit awareness of the category structure used in the task. Accuracy was the primary dependent measure. Taraban (2006) found evidence that faster response latencies were associated with greater learning of category information. The difficulty ratings were included as a way to assess participant's perceptions of the task and as another secondary indication of learning: lower difficulty ratings indicating greater learning.

The two behavioral measures, accuracy and latency, were analyzed together in a doubly multivariate analyses of variance (MANOVA) using GLM in SPSS v15.0. This analysis allows statistical evaluation of between-subjects and within-subjects effects for two variables measured on different scales. This analysis was followed by two separate profile analyses, one with accuracies as the dependent measures, the other with latencies as the dependent measures. The difficulty measures were analyzed in a separate profile analysis. Each profile analysis was implemented through a 3 (condition: blocked, mixed, control) X 4 (measure: base-learning, base-generalization, transfer-learning, transfer-generalization) mixed design MANOVA. Condition was the between-subjects factor, and measure was the within subjects factor.

All MANOVAs were conducted using SPSS v15.0 GLM. The Wilk's Lambda criteria is the preferred statistic when there is more than one degree of freedom for an effect (Tabachnick & Fidell, 2001), unless there is unequal sample sizes, in which case Pillai's criteria is preferred. Follow-up tests for the within-subjects factor were conducted using SPSS repeated contrasts. Follow-up tests for the between-subjects factor were conducted using pairwise comparisons adjusted with a Tukey correction (Toothaker, 1993).

Profile analysis is the appropriate analysis for the present study for several reasons. It answers the critical experimental questions in a single analysis without inflating the Type I error rate or requiring severe corrections on multiple univariate tests that would obfuscate all but the largest results. Profile analysis is preferred over univariate repeated measures analysis in situations where participants cannot be randomly assigned to groups and when within-subjects effects are not counterbalanced and may have carry-over effects (Tabachnick & Fidell, 2001). The repeated measures used in this study occur in a fixed order and carry-over effects (i.e., learning) are expected. Additionally, Experiment 2 adds a WM subject factor, which cannot be randomly assigned, to the design. There are three results from a profile analysis: levels, flatness, and parallelism. The specific predictions for each test are outlined below.

Significant results on a test of levels indicate that at least one of the three conditions differ from one another across the four dependent measures. Hypothesis 1 and Hypothesis 2 together predict a significant levels effect for the accuracy measure due to differences between the blocked and mixed condition. Hypothesis 4 would be supported

by a significant levels effect for latency if participants in the blocked condition answered more quickly than participants in the mixed and control conditions. Hypothesis 5 predicts a significant levels effect for difficulty, with participants in the blocked condition exhibiting significantly lower difficulty ratings than the mixed and control conditions.

Significant results on a test of flatness indicate that the four dependent measures differ from one another. A significant effect for the test of flatness would provide support for the item-learning hypothesis if the learning items are better than the generalization items for any of the three dependent measures where generalization accuracies do not differ from chance. If the flatness effect is not significant, then the learning items do not differ from the generalization items. A null result or a finding that accuracy was higher on the generalization items than on the learning items would support Hypothesis 3, that learning was due to the acquisition of category knowledge (Clapper & Bower, 2002).

Significant results on a test of parallelism indicate that the differences between the categories differ for the four dependent measures. A significant parallelism effect could support either the category invention effect (Clapper & Bower, 2002) or facilitated item memory. A significant parallelism effect due to differences between groups on the learning items but not the generalization items would indicate that the pattern-sequence effect resulted from facilitated item memory for those items seen during learning. The test of parallelism, as well as levels and flatness, will be considered separately for each analysis below.

The initial analyses were followed by additional planned analyses that compared the first items with the mean of the second and third items (subsequent items). This is important because participants received feedback following every item, and the ideal learner would use the feedback from the first item to answer the following two items correctly. This comparison was achieved by adding a within-subjects sequence factor to the profile analysis. A doubly multivariate design was used to evaluate these results. Differences on the first items would be more likely to reflect differences in item memory. Differences in subsequent-item performance without similar differences in first-item performance would indicate that the results are not due to item memory. Due to the small number of generalization items used in Experiment 1 and the randomization procedure, there were no first-item generalization measures (accuracy, latency, or difficulty) for four participants, one in the blocked, and three in the mixed condition.

All multivariate tests for the overall measures were evaluated using Wilk's criteria. Pillai's *T* was used to evaluate the multivariate tests involving comparisons of the first trial involving a plant name to subsequent trials involving the same name, due to missing data for four participants (Tabachnick & Fidell, 2001).

Results

The means and standard deviations for the accuracy, latency, and difficulty rating measures are presented in Table 4, Table 5, and Table 6, respectively. The post-experimental questionnaire was analyzed separately and is reported below. Significant

results for follow-up tests and additional analyses are at the $p < .05$ level unless otherwise reported.

Manipulation Check

Base-Item Pretraining and Learning Blocks

In order to verify that participants learned the answer options at the very least, and to evaluate the learning curves associated with the base-learning items, the accuracies for the six blocks of learning items was examined. The results are presented in Figure 4. A 6 (base-item block) X 3 (condition) mixed design ANOVA, with block as a within subjects variable and accuracy as the dependent variable was conducted to determine if there was improvement across the six repetitions of the base-learning items and if there were any differences between groups. The blocked condition had a different temporal presentation sequence for the Category A and Category B items than did the mixed and control conditions. For the purpose of this analysis, block 1 was composed of the First A and First B values, block 2 was composed of the Second A and Second B values, and block 3 was composed of the Third A and Third B values. See Table 1 for an outline of the experimental sequence and what data were obtained during each block. Blocks 4, 5, and 6 followed the same temporal presentation sequence in all three conditions. Block 4 and 5 were composed of the Learning Block 1 and Learning Block 2 data, respectively. Block 6 was composed of the base-learning items from the test block. The Greenhouse-Geisser correction is reported for all within-subjects univariate ANOVAs to correct for any violations of sphericity.

There was a main effect of condition, $F(3.525, 222.075) = 70.474$, $p < .001$, $\eta^2 = .528$ and a main effect of block, $F(2, 63) = 170.116$, $p < .001$, $\eta^2 = .844$ indicating that learning differed across conditions and across blocks. These main effects were qualified by a significant block X condition interaction, $F(7.050, 222.075) = 14.451$, $p < .001$, $\eta^2 = .314$ (see Figure 4). Univariate repeated measures ANOVAs were conducted for each condition separately to locate the source of the interaction. The effect of block was significant in each of the three conditions; blocked: $F(3.532, 74.173) = 77.993$, $p < .001$, $\eta^2 = .788$, mixed: $F(3.402, 71.434) = 29.630$, $p < .001$, $\eta^2 = .585$, and control: $F(2.756, 57.884) = 15.649$, $p < .001$, $\eta^2 = .427$. Repeated contrasts conducted using SPSS Contrasts revealed that all three conditions improved from the first to the second block, but only the blocked condition exhibited any change in any of the remaining blocks (see Figure 4). Blocked performance improved significantly from the second to the third block, then declined significantly from the third to the fourth block, coinciding with the transition from the consistent pretraining trials to the intermixed learning trials. This pattern is consistent with Clapper and Bower (2002), who reported significant learning across the pretraining trials for participants in the blocked condition only. A oneway ANOVA confirmed that the difference between the conditions was significant at the end of the pretraining phase (blocks 1-3), $F(2, 63) = 187.346$, $p < .001$, (see Figure 4). Tukey post-hoc tests revealed that participants in the blocked condition exhibited significantly higher accuracy ($M = .9763$, $SD = .0290$) than participants in the mixed ($M = .4991$, $SD = .1155$) and control conditions ($M = .4441$, $SD = .1265$), which did not differ.

First Versus Subsequent Item Analysis

Taking a closer look at the learning trends by separating them into performance on the first items and performance on the subsequent items reveals an identical pattern of results for the first items, but a different pattern of results for the subsequent items. The results of this analysis are presented in Figure 5. The effect of block was significant for the first items in each of the three conditions; blocked: $F(3.253, 68.320) = 80.084$, $p < .001$, $\eta^2 = .792$, mixed: $F(3.402, 71.434) = 29.630$, $p < .001$, $\eta^2 = .585$, and control: $F(2.756, 57.884) = 15.649$, $p < .001$, $\eta^2 = .427$. Repeated contrasts revealed an identical pattern of results to the overall accuracy analysis. All three conditions improved from the first to the second block. The blocked condition improved significantly again from the second to third block before declining significantly from the third to fourth block (see Figure 5). This pattern of pretraining accuracies is identical to the overall results. A oneway ANOVA confirmed that the difference between the conditions was significant at the end of the pretraining phase, $F(2, 63) = 128.239$, $p < .001$. Tukey post-hoc tests revealed that participants in the blocked condition exhibited significantly higher accuracy ($M = .9773$, $SD = .0454$) than participants in the mixed ($M = .4517$, $SD = .1172$) and control conditions ($M = .4801$, $SD = .1708$), which did not differ.

When subsequent accuracy was the dependent variable, a different pattern of results emerged. The effect of block was again significant for each of the three conditions; blocked: $F(3.656, 76.767) = 57.279$, $p < .001$, $\eta^2 = .732$, mixed: $F(3.038, 63.808) = 21.323$, $p < .001$, $\eta^2 = .504$, and control: $F(3.017, 63.358) = 13.244$, $p < .001$, $\eta^2 = .387$. Repeated contrasts revealed that all three conditions improved significantly

from first to second block, but that the participants in the blocked condition maintained their high level of performance in block 3, declined significantly in block 4, improved significantly in block 5, and declined significantly in block 6. This pattern of pretraining accuracies is different from the overall and first-item results because the participants in the blocked condition plateaued on the second block. A oneway ANOVA confirmed that the difference between the conditions was significant at the end of the pretraining phase, $F(2, 63) = 140.958$, $p < .001$. Tukey post-hoc tests revealed that participants in the blocked condition exhibited significantly higher accuracy ($M = .9759$, $SD = .0319$) than participants in the mixed ($M = .5227$, $SD = .1519$) who exhibited significantly higher accuracy than participants in the control condition ($M = .4261$, $SD = .1274$). In this case, only the participants in the blocked condition performed significantly above chance, $t(21) = 69.997$, $p < .001$. The difference between the mixed and control conditions was due to participants in the control condition performing significantly below chance, $t(21) = 2.719$, $p = .013$, whereas the participants in the mixed condition did not differ significantly from chance performance, $t(21) < 1$.

The Surprise Effect

Clapper and Bower (1994, 2002) proposed that the ordering manipulation produces surprise as an intermediate mechanism that then causes participants to create two separate categories in order to represent the items seen during learning. This is visible in their results as a sharp drop in accuracy between then end of the Category A trials and the first Category B trials, followed by a rapid increase in accuracy in the

Category B items to the high levels obtained for the Category A items for participants in the blocked condition. This effect is present in the current results presented in Figure 4 as the significant drop between the Third A and the First B items in the blocked condition, $t(65) = 13.098$, $p < .001$. The effect is also true for both the first and subsequent items in Figure 5, $t(65) = 9.453$, $p < .001$ and $t(65) = 10.753$, $p < .001$, respectively.

Omnibus Test: Accuracy and Latency

The behavioral measures of category learning, accuracy and latency, were first evaluated in a 3 (Condition: blocked, mixed, control) X 4 (Measure: base-learning, base-generalization, transfer-learning, transfer-generalization) doubly multivariate MANOVA conducted on the accuracy and latency dependent measures. This analysis is an omnibus test for Hypotheses 1-4 and the means are presented in Table 4 and 5 for accuracy and latency, respectively. The Condition X Measure interaction was not significant, Wilk's $\Lambda = .743$, $F(12, 116) = 1.550$, $p = .116$, permitting a clean interpretation of the main effects. The main effect of Condition was significant, Wilk's $\Lambda = .323$, $F(4, 124) = 23.518$, $p < .001$, $\eta^2 = .431$. Tukey post hoc tests revealed that participants in the blocked condition had higher accuracies than participants in the mixed and control conditions, which did not differ. There were no differences between conditions on the latency measures. The main effect of measure was also significant, Wilk's $\Lambda = .733$, $F(6, 58) = 3.515$, $p = .005$, $\eta^2 = .267$, because the transfer-learning and transfer-generalization items were answered more

quickly than were the base-learning and base-generalization items, $F(1, 63) = 12.690$, $p = .001$, $\eta^2 = .168$.

First Versus Subsequent Item Analysis

The analysis was repeated with the addition of a time factor (first vs. subsequent). The means are presented in Table 4 and 5 for accuracy and latency, respectively. The analysis revealed the same significant main effects of measure, Pillai's $T = .202$, $F(6, 54) = 2.282$, $p = .049$, $\eta^2 = .202$, and condition, Pillai's $T = .599$, $F(4, 118) = 12.626$, $p < .001$, $\eta^2 = .300$. There was a main effect of time, Pillai's $T = .653$, $F(2, 58) = 54.629$, $p < .001$, $\eta^2 = .653$, which was qualified by a significant condition X time interaction, Pillai's $T = .480$, $F(4, 118) = 9.327$, $p < .001$, $\eta^2 = .240$. Follow-up profile analyses were conducted on the accuracy and latency measures separately and are reported below.

Hypotheses 1, 2, and 3

Hypotheses 1, 2, and 3 were based on the accuracy measures and were evaluated together in the same profile analysis. The results are presented in Figure 6 and the mean accuracies are presented in Table 4. The test of parallelism was not significant, Wilk's $\Lambda = .906$, $F(6,122) = 1.031$, $p = .408$, $\eta^2 = .125$, permitting a clean interpretation of the test of levels and flatness, which are discussed below as they apply to Hypotheses 1-3.

Hypothesis 1

There will be greater learning during the learning phase in the blocked than the mixed and control conditions. This hypothesis was supported. The results are presented in Figure 7 as the difference between the blocked condition and the mixed and control conditions on the base-learning items. This effect was significant based on the significant levels effect, $F(2, 63) = 64.987.031$, $p < .001$, $\eta^2 = .674$, indicating that the conditions differed in accuracy. Tukey post-hoc tests revealed that the blocked condition had higher accuracies than the mixed and control conditions, which did not differ. Single sample t-tests revealed that both the blocked, $t(21) = 16.541$, $p < .001$, and mixed, $t(21) = 2.257$, $p = .035$, conditions performed above chance on the base-learning items. This replicates the findings of Clapper and Bower (2002).

Hypothesis 2

More learning will occur during the transfer-learning phase for the blocked condition than the mixed and control conditions. This hypothesis was also supported. The results are presented in Figure 7 as the difference between the blocked condition and the mixed and control conditions, which did not differ, on the transfer-learning items. This effect was significant based on the same significant levels effect, $F(2, 63) = 64.987.031$, $p < .001$, $\eta^2 = .674$. Tukey post-hoc tests established that the blocked condition had higher accuracies than the mixed and control conditions, which did not differ. Only participants in the blocked condition performed above chance, $t(21) = 17.158$, $p < .001$.

Hypothesis 3

Participants in the blocked condition will achieve above chance accuracy on the base-generalization items and the transfer-generalization items and will perform better than participants in the mixed and control conditions. This hypothesis was supported in two ways visible in Figure 7. First, the blocked condition achieved higher accuracies on these two measures than the mixed and control conditions, which did not differ. This effect was significant based on the significant levels effect, $F(2, 63) = 64.987.031$, $p < .001$, $\eta^2 = .674$, followed by Tukey post-hoc tests. Only participants in blocked condition performed above chance on the base-generalization, $t(21) = 13.853$, $p < .001$, and transfer-generalization items, $t(21) = 12.076$, $p < .001$ (see Table 4 for means and standard deviations). This is the critical result that supports the hypothesis that the pattern-sequence effect results in the acquisition of category knowledge, not paired-associate memory for the plant-feature pairs. Further evidence for the acquisition of category knowledge comes from the non-significant flatness effect, meaning that the learning accuracies did not differ significantly from the generalization accuracies, Wilk's $\Lambda = .896$, $F(3, 61) = 2.373$, $p = .079$, $\eta^2 = .104$. This is important because it means that the generalization items, which participants had not previously seen, were answered just as successfully as the learning items for which participants received feedback several times.

An additional oneway analysis of variance (ANOVA) was conducted on the accuracies from the first block of transfer items as an additional test of Hypothesis 3. Participants who have acquired category knowledge should perform above chance on

these new items because, even though they are unfamiliar, they conform to the known category structure. If participants rely only on memory for item-feature pairs, then they will perform at chance on this first block of unfamiliar items. The conditions were significantly different, $F(2, 63) = 37.038$, $p < .001$. Tukey post-hoc tests revealed that the blocked condition achieved higher accuracy ($M = 76.70\%$, $SD = 9.33\%$) than the mixed condition ($M = 55.3\%$, $SD = 15.71\%$), which achieved higher accuracy than the control condition ($M = 45.64\%$, $SD = 10.79\%$). Single sample t-tests against chance performance revealed that only the blocked condition performed above chance, $t(21) = 13.423$, $p < .001$. The control condition was marginally below chance, $t(21) = 13.423$, $p < .001$.

First Versus Subsequent Item Analysis

The addition of the time factor revealed that the differences observed between groups in the overall results were due to differences in the subsequent items, but not the first items, as evidenced by a significant Time X Condition interaction, Pillai's $T = .461$, $F(2, 59) = 25.252$, $p < .001$, $\eta^2 = .461$. Separate analyses were conducted on the first and subsequent items to locate the source of the interaction. The main effect of condition was not significant for the first items, $F(2, 59) = 1.613$, $p = .208$, but was significant for the subsequent items, $F(2, 59) = 62.224$, $p < .001$, $\eta^2 = .678$. Tukey post-hoc tests revealed that the blocked condition had significantly higher accuracies than the mixed and control conditions, which did not differ. There were no other significant interactions. The results are presented in Figure 7.

Hypothesis 1. These results support Hypothesis 1 because the differences between the blocked condition and the mixed and control conditions were on the subsequent items, which benefit the most from the category structure. Participants in all three groups performed equally well on the first items. Only the participants in the blocked condition were able to use the information in the feedback from the first item to improve on the subsequent items. Single sample t-tests revealed that participants in the blocked condition performed significantly better than chance on both the base-learning first-item and subsequent-item accuracy measures, $t(21) = 2.619$, $p = .016$, and $t(21) = 23.539$, $p < .001$, respectively. Even though participants in the mixed condition performed above chance on the subsequent base-learning items, $t(21) = 2.224$, $p = .037$, they did not differ significantly from the control condition.

Hypothesis 2. These results also support Hypothesis 2 for the same reason. Participants in the blocked condition performed no differently than participants in the mixed and control conditions on the first items, but were the only ones able to use the feedback to do significantly better on the subsequent items. Single sample t-tests revealed that all participants performed at chance on the first-items, all $ps \geq .100$, but only participants in the blocked condition performed significantly better than chance on the transfer-learning subsequent-item accuracy measure, $t(21) = 16.541$, $p < .001$.

Hypothesis 3. These results support Hypothesis 3 in two ways. Participants in the blocked condition performed no differently than participants on the base-generalization and transfer-generalization first-item accuracies, but were the only ones able to use the feedback to do significantly better on the subsequent items. Single sample

t-tests revealed that participants in the blocked condition performed significantly better than chance on the base-generalization and transfer-generalization subsequent-item accuracy measures, $t(21) = 26.138$, $p < .001$, and , $t(21) = 15.868$, $p < .001$, respectively. The non-significant flatness effect for the first and subsequent accuracies provides additional support for Hypothesis 3 because the learning items were no better than the generalization items, despite having the benefit of multiple repetitions with feedback during the pretraining and learning phases.

The same pattern of results was found for the first block of transfer items. There were no group differences on the first items, $F < 1$, but significant group differences on the subsequent items, $F(2, 63) = 48.692$, $p < .001$. Tukey post-hoc tests revealed that the blocked condition achieved higher accuracy ($M = 89.49\%$, $SD = 9.33\%$) than the mixed condition ($M = 55.97\%$, $SD = 21.69\%$) and the control condition ($M = 44.89\%$, $SD = 12.45\%$), which did not differ. This result supports Hypothesis 3 because the only way participants in the blocked condition could perform better than the other conditions on the subsequent items for the new trees would be to use the feedback on the first item to access the appropriate category knowledge to answer the subsequent items better.

Hypothesis 4

Participants in the blocked condition should exhibit faster response times for the new items (base-generalization and transfer-generalization items) than participants in the mixed and control conditions. This hypothesis was not supported. The results are presented in Figure 8. The blocked condition did not answer any more quickly than the

mixed or control conditions on any of the four measures; neither the test of parallelism nor the test of levels was significant, Wilk's $\Lambda = .835$, $F(6,122) = 1.921$, $p = .083$, $\eta^2 = .086$, and $F < 1$, respectively. The only significant finding for the latency measures was the faster response times for the transfer items than the base items. This was supported by a significant test of flatness, Wilk's $\Lambda = .818$, $F(3, 61) = 4.531$, $p < .001$, $\eta^2 = .182$. Repeated contrasts were performed using SPSS contrasts to locate the source of the significant flatness effect. The base-learning and base-generalization latencies did not differ from one another, but were slower than the transfer-learning and transfer-generalization latencies, $F(3, 61) = 12.690$, $p = .001$, $\eta^2 = .168$, which did not differ from one another. This effect of flatness is consistent with the category invention hypothesis because the effect was due to differences between the base and transfer items, not between the learning and generalization items.

First Versus Subsequent Item Analysis

The first and subsequent latencies did not change the conclusions from the overall data. The results are presented in Figure 10. The subsequent items were answered significantly more quickly than the first items, Pillai's $T = .608$, $F(1, 59) = 91.497$, $p < .001$, $\eta^2 = .608$. This does not support the abstract category learning hypothesis because there were no differences between groups on the subsequent items; everyone was equally faster on the subsequent items, as shown by the non-significant Condition X Time interaction ($F < 1$). No other effects were significant, though the Time X Measure

interaction approached significance, Pillai's $T = .108$, $F(3, 57) = 4.531$, $p = .087$, $\eta^2 = .108$.

Hypothesis 5

Difficulty ratings for the base-learning, base-transfer, transfer-learning, and transfer-generalization items will be lower in the blocked condition than the mixed and control conditions. This hypothesis was supported in two ways. The results are presented in Figure 11. First, the difference between the blocked condition and the mixed and control conditions is reliable, as shown by a significant test of levels, $F(2, 63) = 21.061$, $p < .001$, $\eta^2 = .325$. Tukey post-hoc tests revealed that the blocked condition had lower overall difficulty ratings than the mixed and control conditions, which did not differ. Second, participants rated the generalization items (both base-generalization and transfer-generalization) no more or less difficult than the learning items. The test of flatness was not significant, Wilk's $\Lambda = .908$, $F(3, 61) = 2.066$, $p = .114$, $\eta^2 = .092$. This result supports the hypothesis that participants acquired category knowledge rather than simply remembering the item-feature pairs because there was not advantage for the items seen during learning; the test of parallelism for the difficulty items was not significant, $F < 1$.

First Versus Subsequent Item Analysis

The difficulty data were reexamined with the addition of the time factor. The results are presented in Figure 11. The first items were rated as significantly more

difficult than the subsequent items as indicated by a significant main effect of Time, Pillai's $T = .374$, $F(1, 59) = 35.273$, $p < .001$, $\eta^2 = .374$. More importantly, the differences between conditions were not the same for the first and subsequent items as indicated by a significant Condition X Time test of parallelism, Pillai's $T = .369$, $F(2, 59) = 17.285$, $p < .001$, $\eta^2 = .369$. Participants in the blocked condition did not rate the first-items as any more difficult than did the participants in the mixed condition, but did rate these items as easier than did participants in the control condition; the levels effect was not significant for the first items, $F(2, 59) = 4.677$, $p = .037$. There were no differences between the four measures as indicated by a significant flatness effect, Pillai's $T = .089$, $F(3, 57) = 1.861$, $p = .146$. There were no significant interactions. Participants in the blocked condition rated the subsequent items as significantly easier to answer than did participants in the control conditions. This result was based on a significant levels effect, $F(2, 59) = 18.151$, $p < .001$, $\eta^2 = .381$, followed by Tukey post-hoc tests, which revealed that participants in the blocked condition had significantly lower difficulty ratings than the participants in the mixed and control conditions, which did not differ.

Post-Experimental Questionnaire

The post-experimental questionnaires were coded as described in Appendix C. The number of participants in each condition who reported actively searching for patterns and who exhibited awareness of the category structure are presented in Table 7. A chi-squared test of independence revealed a significant difference between the groups, X^2 (2) = 9.149, $p = .010$. Follow up 2 x 2 chi-squared tests revealed that the participants in the

blocked condition engaged in significantly less active hypothesis testing than did participants in the mixed condition, X^2 (1) = 5.939, p = .015, and the control condition, X^2 (1) = 8.282, p = .004 but did report significantly more awareness of the category structure than the mixed condition, X^2 (1) = 5.500, p = .019, but not the control condition, X^2 (1) = 3.492, p = .062.

Discussion

Experiment 1 replicated the pattern-sequence effect of Clapper and Bower (1994, 2002) and supported a process of category invention for the acquisition of abstract category information. The pattern of data showed a surprise effect that can be taken as support for category invention (Clapper & Bower, 1994, 2002). Participants in the blocked condition performed significantly better than chance and significantly better than participants in the mixed and control condition on the base-learning items. Even though participants in the mixed condition performed significantly above chance on the base-learning items, it was not significantly different from the performance of the participants in the control condition.

The present experiment extended Clapper and Bower (1994, 2002) in five ways that support the hypothesis that the pattern-sequence effect results in the acquisition of abstract category knowledge and not solely improved item-feature memory. First, the flatness effect was not significant for the accuracy measures, which indicates that the generalization items were no different than the learning items, even though participants did not receive previous practice and feedback on these items as they did the learning

items. The levels effect coupled with the non-significant flatness effect indicates that participants in the blocked condition performed significantly better than participants in the mixed and control condition on the generalization items as well as the learning items. This extension of Clapper and Bower (1994, 2002) indicates that the pattern-sequence effect is not due simply to enhanced memory of the presented items but rather leads to abstract knowledge of the category structure independent of memory of specific items because participants had not previously seen the plant-feature combinations in the generalization trials.

Second, this pattern of participants in the blocked condition performing significantly better than chance and significantly better than participants in the mixed and control condition on the new features was also true for the transfer items (transfer-generalization items). This is a replication of the base-generalization results and provides convergent evidence that participants acquired abstract category knowledge because they were able to apply correlated feature knowledge to new items.

Third, there was a significant difference between the groups on the first block of trials for the new plant names (transfer-learning items). The blocked condition performed significantly above chance on the first block of transfer items whereas the mixed and control conditions did not. This supports the abstract category knowledge hypothesis because this was the first time the participants encountered these items. The only way they could have exceeded chance was to have applied knowledge of the category structure to the new items.

Fourth, further support for abstract category knowledge comes from the analysis of first trials associated with a particular plant name versus subsequent trials associated with that name. Participants in the blocked condition performed significantly better than chance and significantly better than participants in the control condition for the first base-learning items, but not for the first transfer-learning items, indicating that they were better able to learn the features associated with specific plant names that they saw six times, but not three times. When the first feature for a plant was unfamiliar (base-generalization and transfer-generalization items), participants in the blocked condition performed no better than chance. However, after receiving feedback, participants in the blocked condition were able to do significantly better than chance on the subsequent items, both the familiar learning items and the unfamiliar generalization items – i.e., for both the items that they had more practice on (base items) and the items that they had less practice on (transfer items). This would suggest that participants in the blocked condition, but not the mixed or control condition, were able to determine the correct category based on the feedback from the first item and go on to get the subsequent items correct. The fact that first-item base-learning accuracy was above chance in the blocked condition does not eliminate the possibility that the pattern-sequence effect improves item memory, but if item memory were the only mechanism, there would be no differences on the generalization items. In the present learning paradigm, it is more likely that participants acquire the abstract category representations first, then learn which exemplars belong to which category (exemplar-category memory).

Fifth, Hypothesis 5 was supported, providing additional evidence that participants acquired category knowledge and used that knowledge to answer the items. The non-significant flatness effect for difficulty ratings means that participants found the base-generalization and transfer-generalization items just as easy to answer than the base-learning and transfer-learning items. Further support for the acquisition of category knowledge came from the significantly lower difficulty ratings on the subsequent items than the first items, indicating that participants used the feedback from the first items to answer the second and third items for each plant.

Category knowledge was evident in differences in accuracy and difficulty ratings, but not in latency. Participant's difficulty ratings mirrored their accuracy. Contrary to Taraban (2006), latency was not a sensitive measure of category knowledge. Participants spent approximately the same time whether they got the item correct or incorrect. This would indicate that participants did not engage in a speed-accuracy trade-off. The lack of difference in reaction time also supports the conclusion that learning the category information did not require significantly more time than not learning it.

Another alternative explanation for the pattern-sequence effect is that participants in the blocked condition engaged in more active hypothesis testing (Ashby et al., 1998; Billman & Knutson, 1996). The analysis of the post-experimental questionnaire does not support this alternative. Participants in the blocked condition engaged in the same amount or less active hypothesis testing than participants in the control and mixed condition, yet exhibited greater awareness of the category structure than participants in the control and mixed conditions. Even though participants in the mixed condition

looked for patterns just as much as participants in the control condition, they were no more aware of the structure. This supports Clapper and Bower's (1994, 2002) assertion that the mere presence of structure is not enough to guarantee that it will be learned.

The mixed condition did not differ substantially from the control condition. There were no significant differences between these two groups on any of the primary accuracy, latency, or difficulty measures (i.e., base-learning, transfer-learning, base-generalization, transfer-generalization). This result replicated Clapper and Bower's (2002) finding that the mere presence of category structure in the stimuli was not enough to permit learning of that structure. There is no need for two control conditions since the mixed condition was not significantly different from the control condition. The mixed condition is a better control for the blocked condition as both use the same structured stimuli. The control condition was not included in Experiment 2 for this reason.

There are three main conclusions from Experiment 1. First, Experiment 1 replicated and extended the pattern-sequence effect (Clapper & Bower, 1994, 2002), supporting the conclusion that the pattern-sequence effect is due to the acquisition of abstract category knowledge and not to enhanced memory for category consistent exemplars. Second, presentation of the learning materials was just as important as the materials themselves as evidenced by the similarity between the mixed and control conditions, which is consistent with Clapper and Bower (1994, 2002) and Taraban (2004, 2006). A third conclusion, not directly associated with the initial hypotheses, is that participants find learning the feature-feature relationships easier than learning the item-feature relationships. This results in a strategy of ignoring the Latin plant names and

guessing on the first feature in order to obtain the feedback. Participants then used the feedback to answer the second and third features perfectly. This behavior has characteristics of localist problem-solving heuristics, although it may be the way categories, even implicitly acquired ones, are learned (e.g., Taraban & Kempe, 1999).

Table 1. Data, Stimuli, and Presentation order by Experimental Phase.

Experimental Block	Phase	Stimuli	Presentation Order: Blocked	Presentation Order: Mixed & Control	Data
Block 1	Pretraining	Base	8 A then 8 A	8 A + 8 B intermixed	First A, Second A
Block 2	Pretraining	Base	8 A then 8 B	8 A + 8 B intermixed	Third A, First B
Block 3	Pretraining	Base	8 B then 8 B	8 A + 8 B intermixed	Second B, Third B
Block 4	Learning	Base	8 A + 8 B intermixed	8 A + 8 B intermixed	Learning Block 1
Block 5	Learning	Base	8 A + 8 B intermixed	8 A + 8 B intermixed	Learning Block 2
Block 6	Test	Base	8 A + 8 B intermixed	8 A + 8 B intermixed	Base-Learning & Base Generalization
Block 7	Transfer	Transfer	8 A + 8 B intermixed	8 A + 8 B intermixed	Transfer Block 1
Block 8	Transfer	Transfer	8 A + 8 B intermixed	8 A + 8 B intermixed	Transfer Block 2
Block 9	Transfer-Test	Transfer	8 A + 8 B intermixed	8 A + 8 B intermixed	Transfer-Learning & Transfer Generalization

Note. The stimuli for each category for the three conditions are presented in Appendix A.

Table 2. Number of Familiar Plant Names Reported by Condition for Experiment 1.

Number Reported	Condition			Total
	Blocked	Control	Mixed	
0	11	12	11	34
1	2	6	6	14
2	2	0	2	4
3	2	0	0	2
Missing	5	4	3	12
Total	22	22	22	66

Table 3. Familiar Plant Names by Condition for Experiment 1.

	Condition			
Latin Plant Name	Blocked	Control	Mixed	Total
Carex		3	1	4
Juglans	1			1
Juncus	1			1
Mentha			2	2
Nepeta	2			2
Pelia		1		1
Prunus	2	1	3	6
Rumex	1			1
Salvia	1		2	3
Vicia	3	1	1	4
Vitex	1		1	2
Total	12	7	10	29

Table 4. Accuracy Measures by Condition for Experiment 1.

Measure		Blocked		Mixed		Control		Overall	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
				Base-Learning					
	Overall	0.7997*	0.0850	0.5696*	0.1446	0.4929	0.0806	0.6207	0.1687
	First	0.5812*	0.1454	0.5072	0.1555	0.4643	0.1345	0.5175	0.1512
	Second	0.9033*	0.0804	0.5955*	0.2015	0.5105	0.0981	0.6698	0.2173
				Base-Generalization					
	Overall	0.7926*	0.0991	0.5256	0.1721	0.4574	0.1355	0.5919	0.1998
	First	0.5627	0.2464	0.4971	0.2331	0.5125	0.2272	0.5241	0.2338
	Second	0.9184*	0.0751	0.5441	0.2385	0.4344	0.1871	0.6323	0.2740
				Transfer-Learning					
	Overall	0.8068*	0.0839	0.5256	0.1527	0.5227	0.1501	0.6184	0.1874
	First	0.5734	0.2002	0.4638	0.2143	0.5029	0.2382	0.5133	0.2196
	Second	0.9212*	0.0866	0.5714	0.2184	0.5383	0.1629	0.6770	0.2385

Table 4 Continued.

Measure	Blocked		Mixed		Control		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Transfer-Generalization								
Overall	0.7841*	0.1103	0.5398	0.2295	0.4432	0.1321	0.5890	0.2179
First	0.4849	0.3301	0.5614	0.3641	0.4439	0.2802	0.4938	0.3228
Second	0.9061*	0.1200	0.5310	0.2860	0.4292	0.2151	0.6221	0.2978
Total								
Overall	0.7958	0.0939	0.5401	0.1757	0.4790	0.1290	0.6065	0.1960
First	0.5513	0.0238	0.5055	0.2464	0.4809	0.2242	0.5123	0.2369
Second	0.9122	0.0909	0.5605	0.2354	0.4781	0.1750	0.6516	0.2621

Note. Condition accuracies marked with * were significantly above from chance (.500).

Table 5. Latency Measures by Condition for Experiment 1.

Measure	Blocked		Mixed		Control		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Base-Learning								
Overall	8.371	0.231	8.404	0.320	8.389	0.276	8.388	0.274
First	8.521	0.309	8.563	0.427	8.510	0.329	8.531	0.354
Second	8.304	0.217	8.309	0.259	8.327	0.256	8.313	0.241
Base-Generalization								
Overall	8.411	0.192	8.412	0.366	8.424	0.350	8.416	0.308
First	8.612	0.241	8.556	0.427	8.570	0.391	8.580	0.357
Second	8.288	0.240	8.330	0.377	8.356	0.344	8.325	0.322
Transfer-Learning								
Overall	8.349	0.235	8.346	0.308	8.255	0.322	8.317	0.289
First	8.490	0.313	8.483	0.416	8.371	0.387	8.448	0.373
Second	8.289	0.251	8.261	0.287	8.205	0.320	8.252	0.285

Table 5 Continued.

Measure	Blocked		Mixed		Control		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Transfer-Generalization								
Overall	8.412	0.337	8.431	0.367	8.249	0.331	8.364	0.350
First	8.601	0.403	8.782	0.629	8.431	0.481	8.596	0.520
Second	8.301	0.338	8.308	0.346	8.190	0.362	8.266	0.348
Total								
Overall	8.386	0.251	8.398	0.337	8.329	0.325	8.371	0.312
First	8.556	0.319	8.589	0.481	8.471	0.401	8.532	0.399
Second	8.296	0.261	8.302	0.316	8.269	0.326	8.290	0.307

Note. The latency measures reported here are natural log transformations of the raw latency data.

Table 6. Difficulty Rating Measures by Condition for Experiment 1.

Measure	Blocked		Mixed		Control		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Base-Learning								
Overall	2.572	1.210	3.818	1.308	4.257	1.073	3.549	1.384
First	3.520	1.278	4.078	1.039	4.271	1.145	3.956	1.184
Second	2.101	1.325	3.687	1.476	4.245	1.082	3.344	1.577
Base-Generalization								
Overall	2.534	1.221	3.932	1.280	4.375	1.022	3.614	1.405
First	3.439	1.463	4.128	1.300	4.448	1.253	4.005	1.387
Second	2.080	1.308	3.790	1.453	4.381	0.960	3.417	1.581
Transfer-Learning								
Overall	2.585	1.228	3.892	1.433	4.676	1.338	3.718	1.576
First	3.512	1.416	4.190	1.340	4.778	1.480	4.160	1.485
Second	2.126	1.366	3.718	1.608	4.605	1.346	3.483	1.758

Table 6 Continued.

Measure	Blocked		Mixed		Control		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Transfer-Generalization								
Overall	2.693	1.236	3.960	1.411	4.739	1.361	3.797	1.567
First	3.803	1.397	4.118	1.344	4.603	1.378	4.184	1.393
Second	2.132	1.357	3.846	1.573	4.834	1.357	3.604	1.804
Total								
Overall	2.596	1.204	3.901	1.337	4.512	1.204	3.657	1.507
First	3.566	1.373	4.129	1.237	4.525	1.311	4.072	1.378
Second	2.110	1.316	3.761	1.503	4.516	1.199	3.446	1.709

Note. The difficulty ratings ranged from 1, “not at all,” to 7 “extremely.”

Table 7. Searching and Awareness by Condition for Experiment 1.

	No Search		Search		
Condition	Not aware	Aware	Not aware	Aware	Total
Blocked	1	19	0	2	22
Mixed	3	10	4	5	22
Control	6	9	3	4	22
Total	10	38	7	11	66

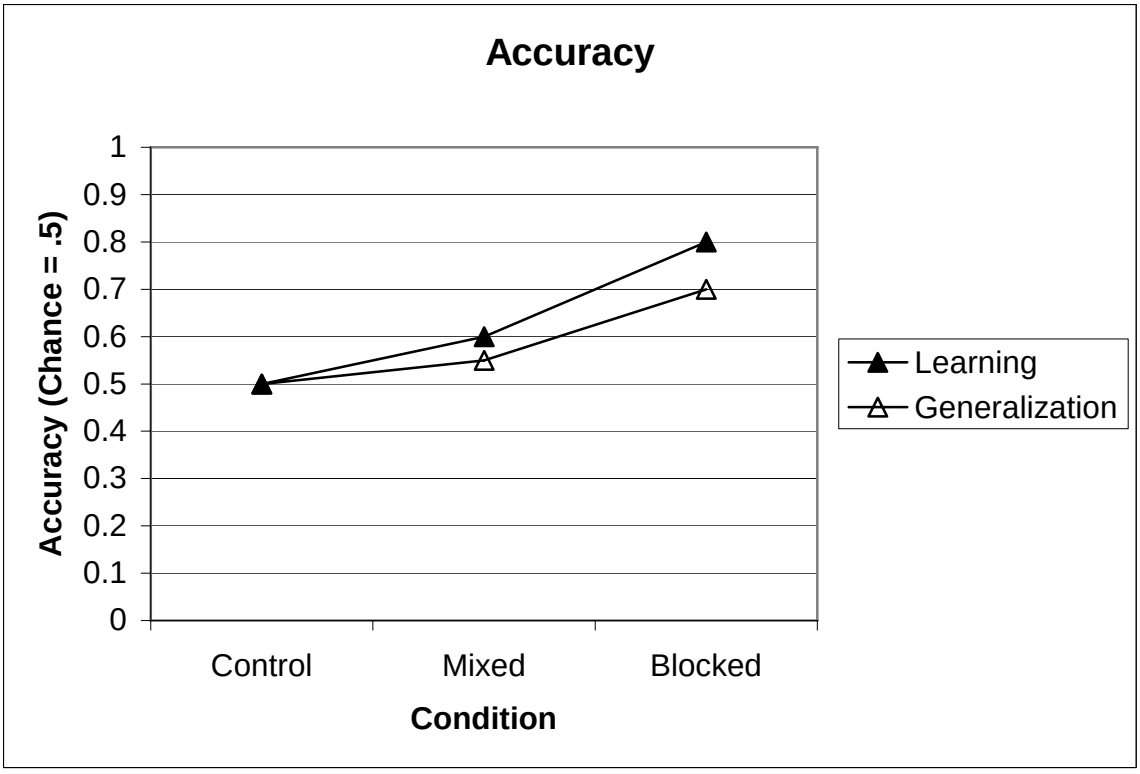

Figure 1 Predicted Pattern-Sequence Effect for Accuracy Dependent Measures.

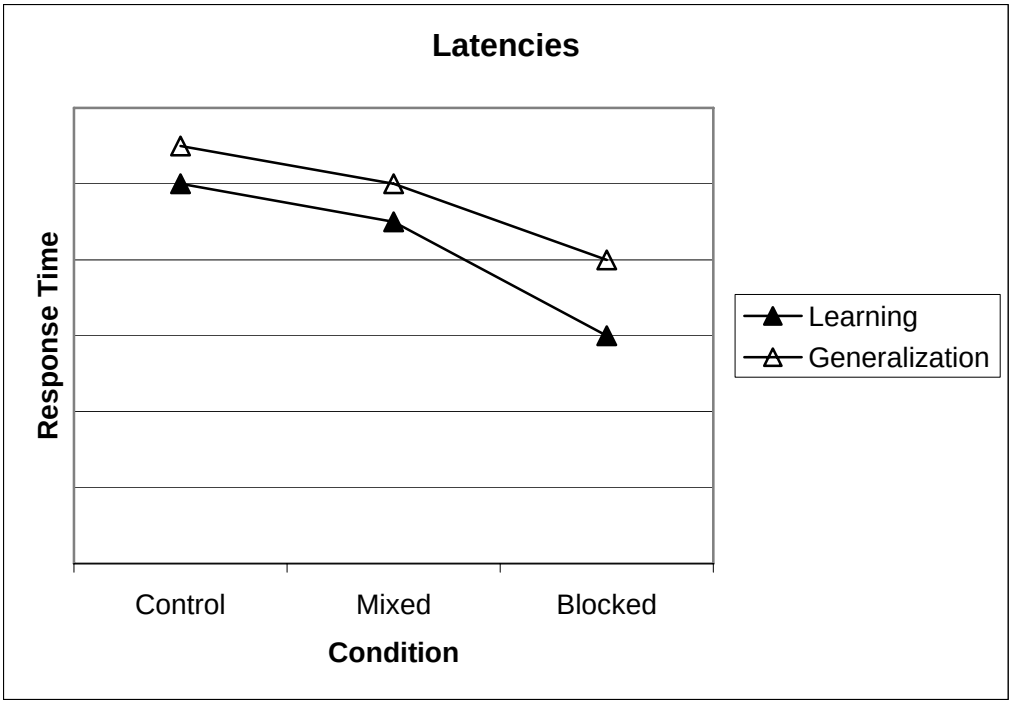

Figure 2 Predicted Pattern-Sequence Effect for Latency Dependent Measures.

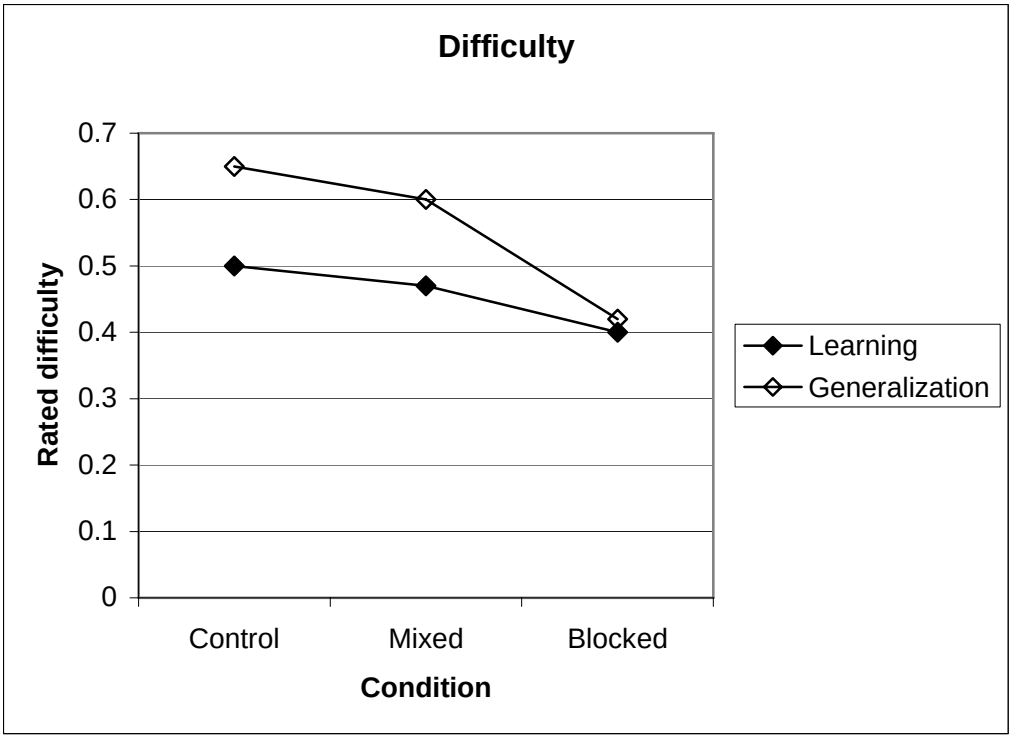

Figure 3 Predicted Difficulty Ratings for Base and Transfer Items.

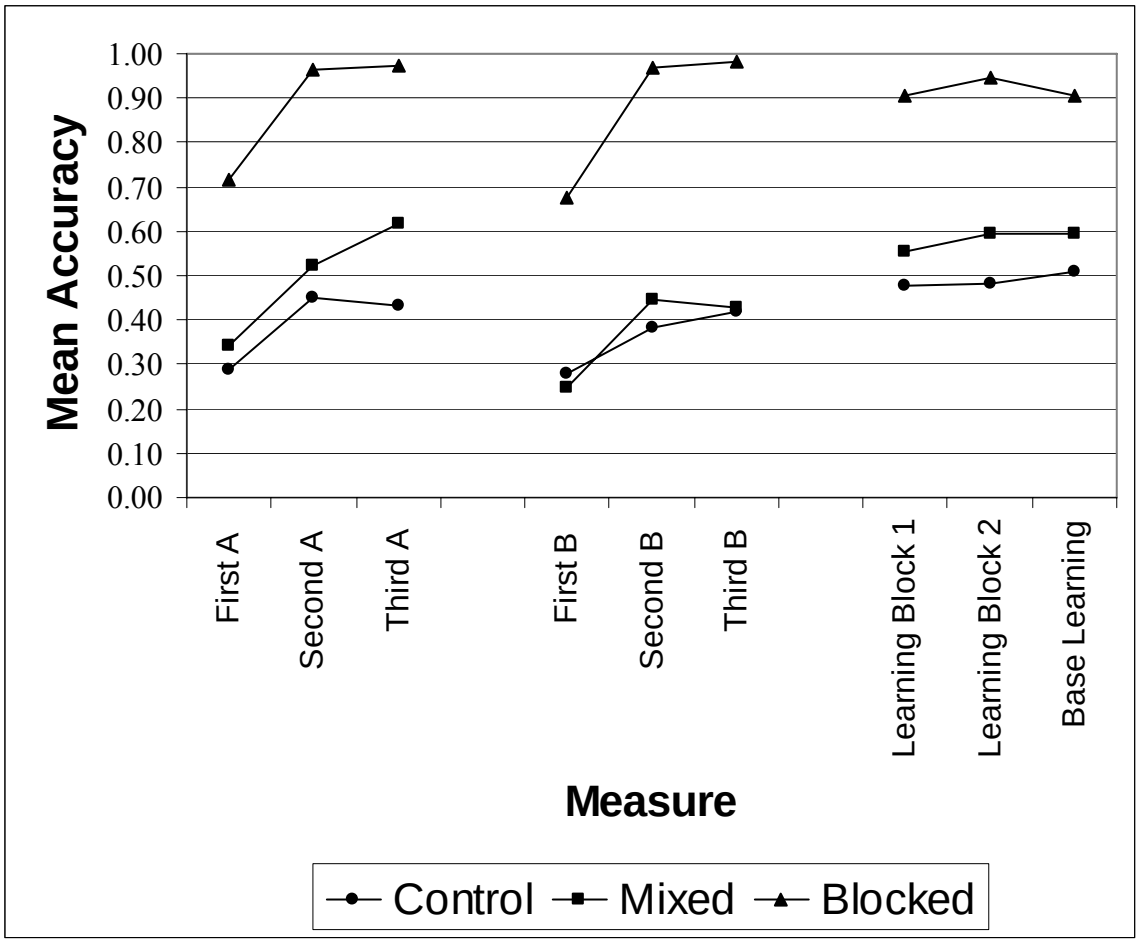

Figure 4. Overall Learning Accuracies for Base Items by Condition.

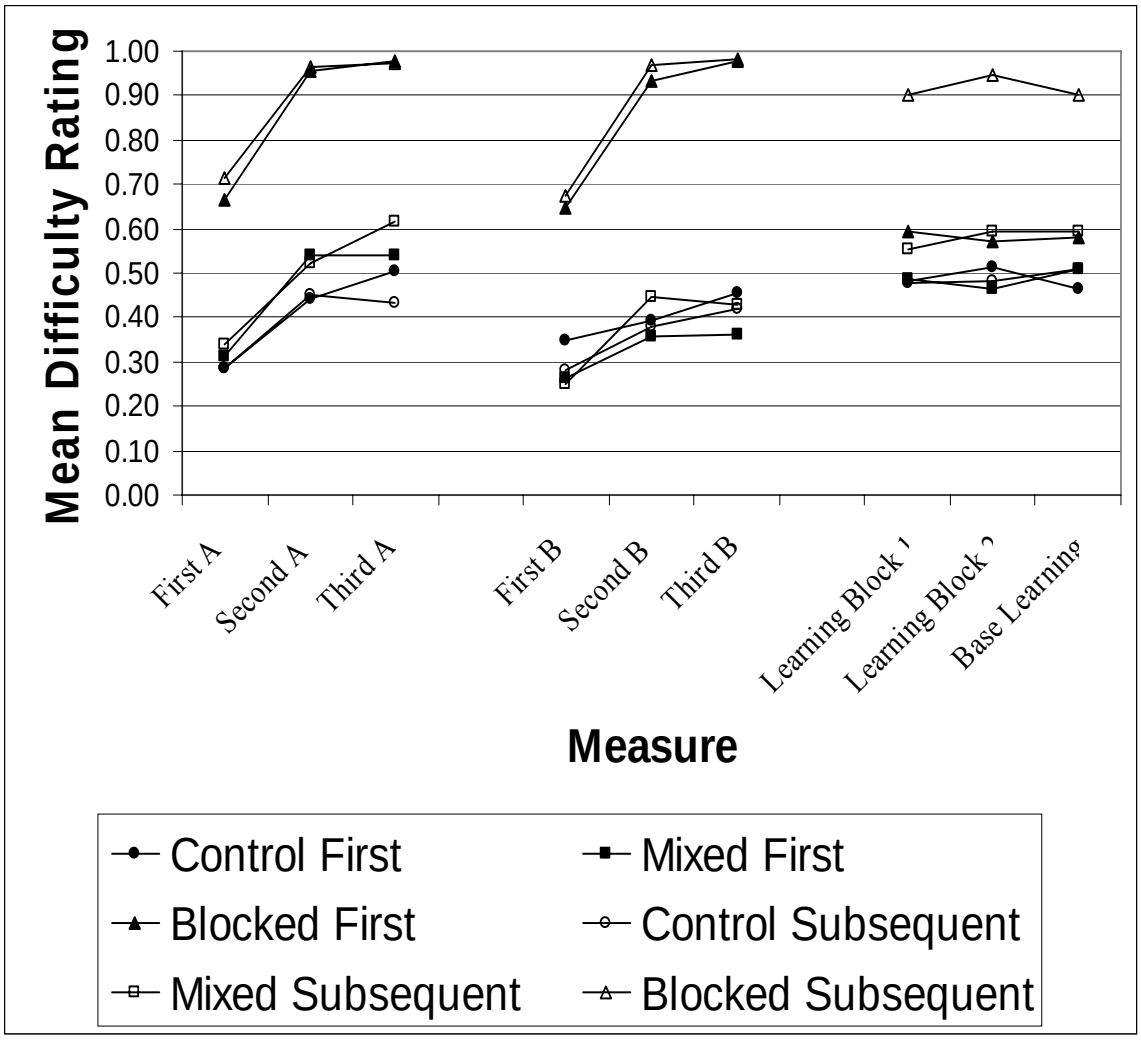

Figure 5. First and Subsequent Learning Accuracies for Base Items by Condition.

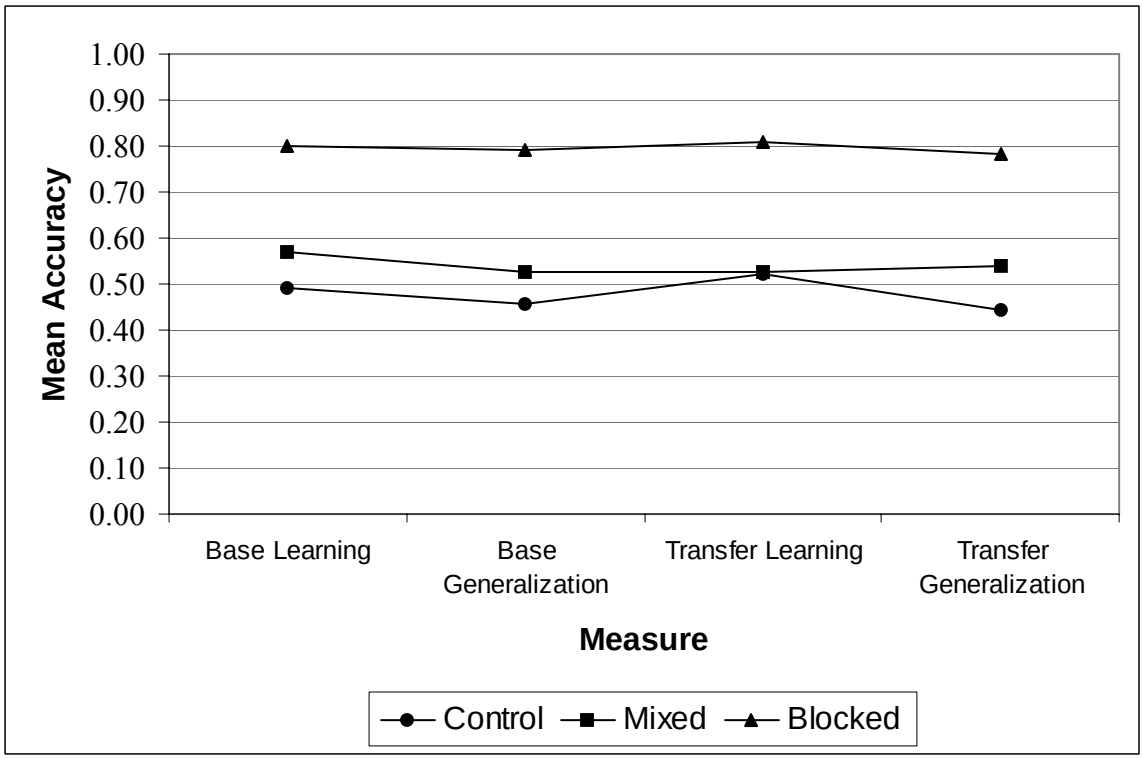

Figure 6. Overall Learning and Generalization Accuracies by Condition.

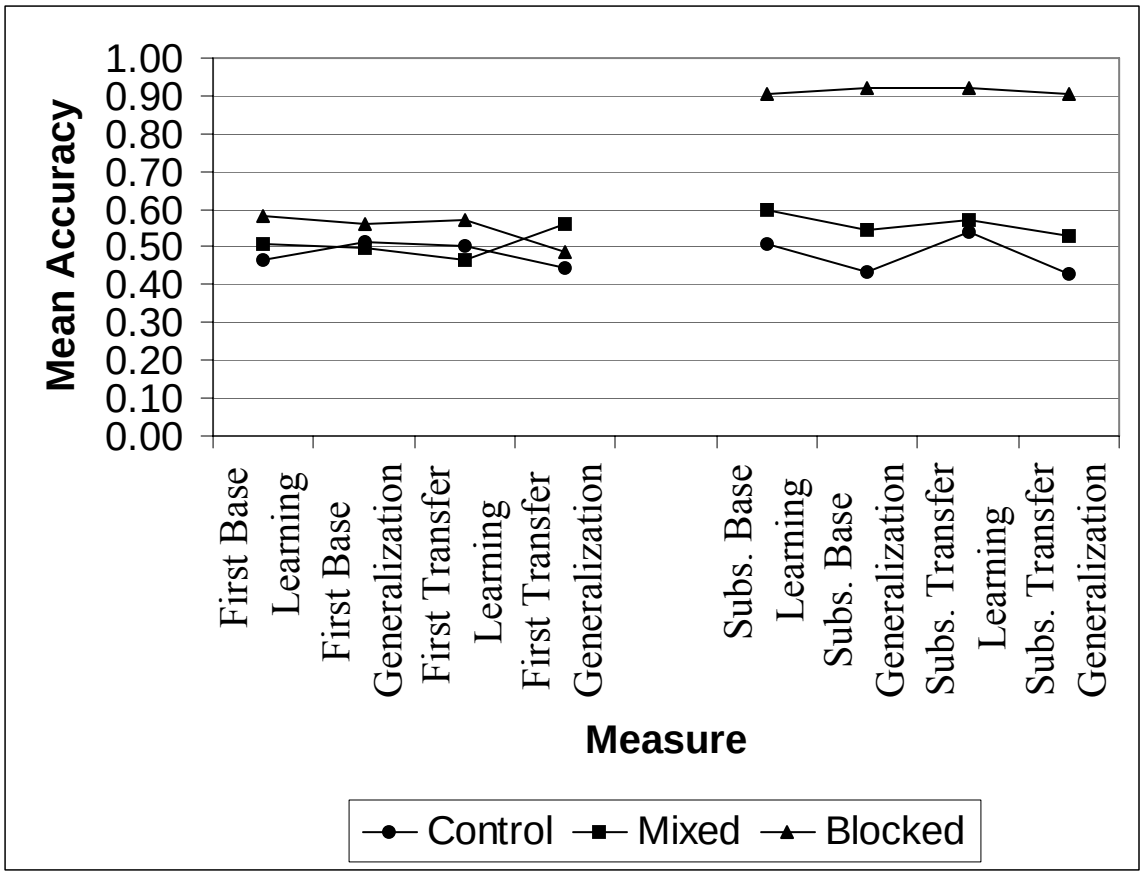

Figure 7. Learning and Generalization First and Subsequent Accuracies by Condition.

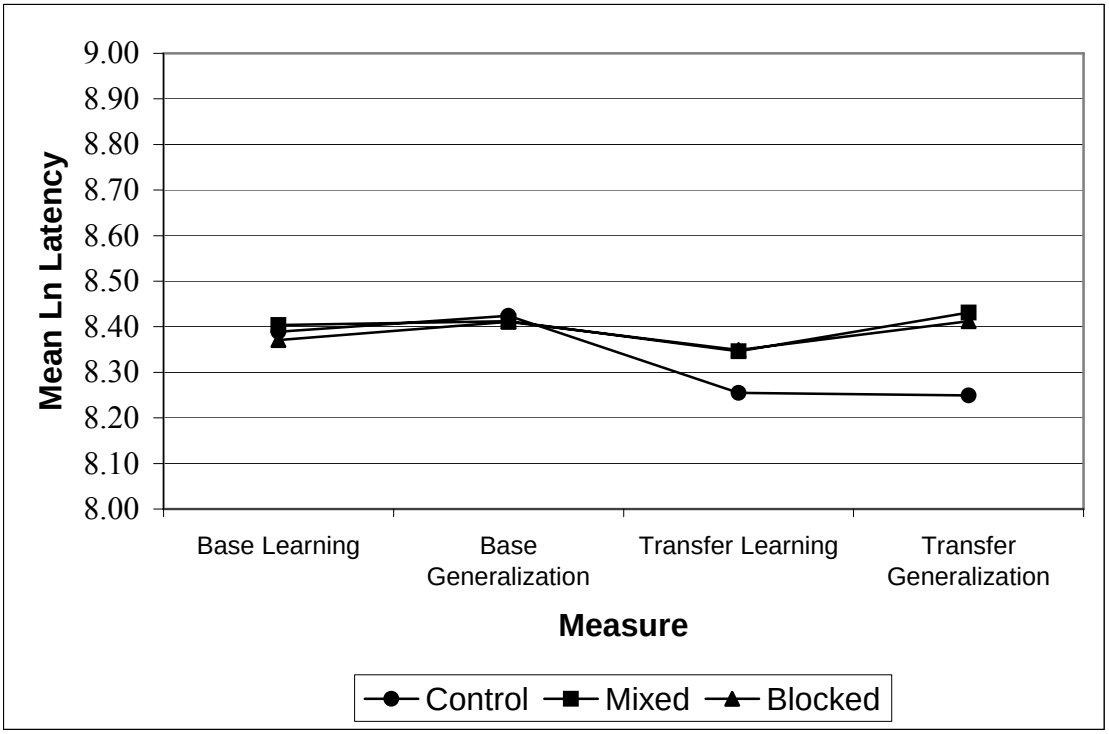

Figure 8. Overall Learning and Generalization Latencies by Condition.

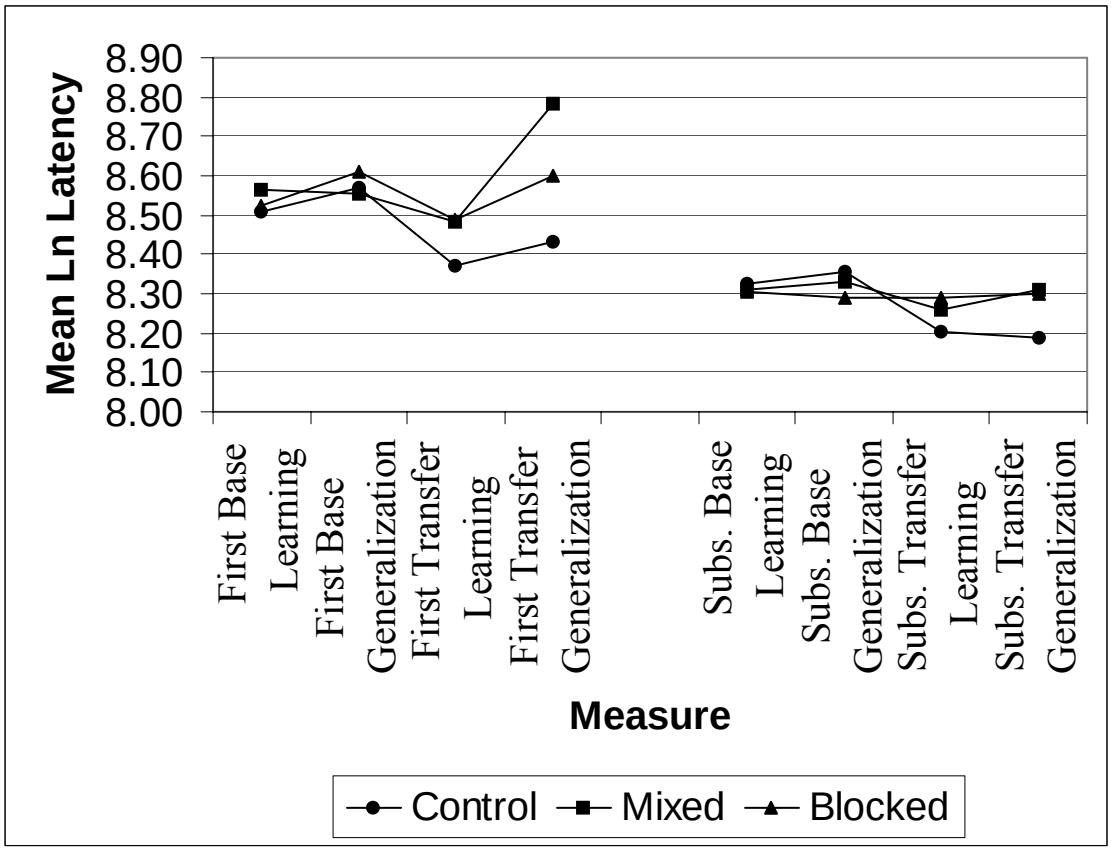

Figure 9. Learning and Generalization First and Subsequent Latencies by Condition.

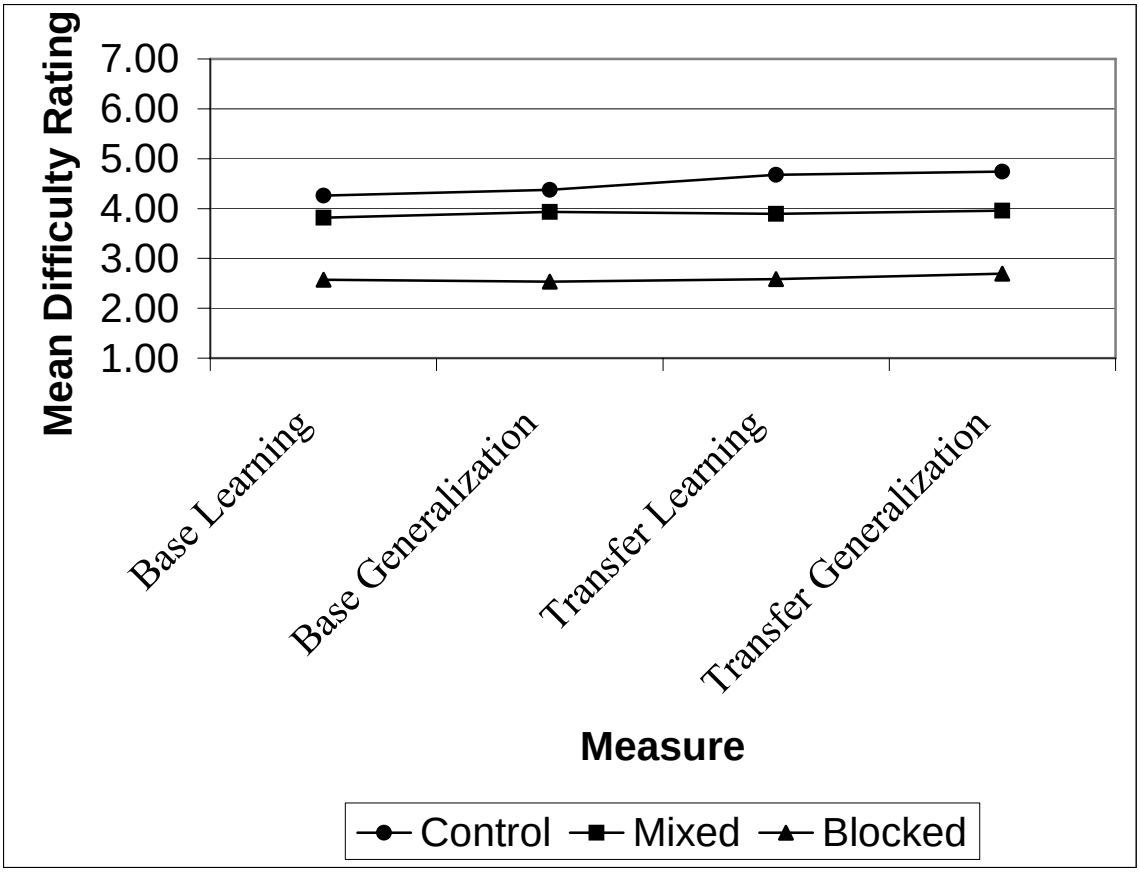

Figure 10. Overall Learning and Generalization Difficulty Ratings by Condition.

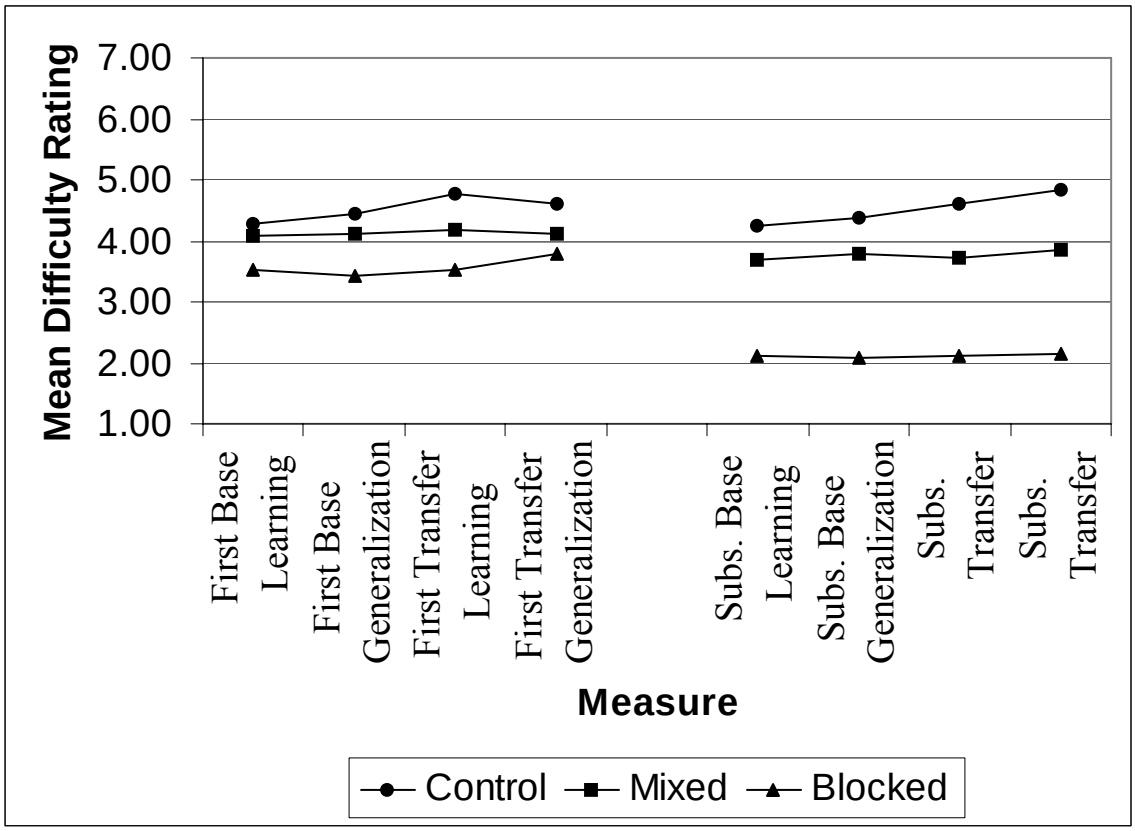

Figure 11. Learning and Generalization First and Subsequent Difficulty Ratings by Condition.

CHAPTER III

EXPERIMENT 2

Experiment 2 examined the interaction of individual differences in WM span and the pattern-sequence effect. The method developed in Experiment 1 was modified as described below and administered to a group of high and a group of low WM span participants in order to test the hypothesis that WM plays a role in the discovery and acquisition of abstract category knowledge (Clapper & Bower, 1994, 2002; Gureckis & Love, 2003b; Love et al., 2004)

There are several measures of WM capacity including reading span (Just & Carpenter, 1992; Engle, Tuholski, Laughlin, & Conway, 1999), operation span (Turner & Engle, 1989; Engle et al., 1999), counting span (Engle et al., 1999), spatial span (Shah & Miyake, 1996), as well as others. These tasks vary in the processing task and the to-be-remembered items. Engle et al. (1999) demonstrated that, despite these differences in processing task and type of item to be remembered, the operation span, reading span, and counting span tasks all loaded on a common factor. Furthermore, single-task measures of short-term memory (storage only) and general fluid abilities loaded on separate factors. Aospan loads on the same factor as the reading span and operation span tasks, indicating that it is based on the same underlying construct (Unsworth et al., 2005). Given the relative equality of WM span tasks, Aospan was selected because it has the advantage of being delivered and scored automatically via computer.

Method

The experimental method was identical to that of Experiment 1 with three modifications. First, this experiment used only the blocked and mixed conditions from Experiment 1. The control condition was dropped as it did not differ significantly from the mixed condition. Second, a pre-experimental assessment of WM span was administered prior to the fact learning task to select high and low WM span participants for use in an extreme groups design. Third, participants completed an additional 24 trials in each of the transfer blocks (3 features for each of the 8 additional plant names not in Experiment 1) in order to equalize the number of base and transfer items and to increase the difficulty of the transfer task by adding additional plant names.

Based on the hypothesized role of WM in both Clapper and Bower's (2002) formulation of the rational model and SUSTAIN (Gureckis & Love, 2003b; Love et al., 2004), the WM factor was assumed to have at least a medium effect size. Power analysis using GPower (Faul & Erdfelder, 1992) indicated that 27 participants per cell would be needed to detect an effect of this size with a power of .8.

Participants

Participants were 114 undergraduate students at Texas Tech University who had not participated in Experiment 1. Participants either received course credit or $20. As in Experiment 1, participants achieving less than 34% correct on the base-learning items were excluded from the analysis. Six participants failed to meet this criterion. All six were low WM span participants, five in the mixed condition and one in the blocked

condition, resulting in 108 participants overall: 54 high WM spans and 54 low WM spans, each randomly assigned to either the mixed or blocked condition (27 in each cell).

Materials

The stimuli for Experiment 2 were identical to those used in Experiment 1 with eight additional exemplars in the transfer set (four in each group), bringing the total number of transfer items to 16. This was done to equalize the number of items in the learning and transfer phases of the experiment. These additional items were selected from the same source (Coombs, 1985) and matched between categories on average word length (5.688 letters overall, 5.688 for Category A, 5.688 for Category B), number of syllables (2.344 overall, 2.313 for Category A, 2.375 for Category B), first letter, last letter, and last syllable. The complete stimuli sets are presented in Appendix A.

Procedure

Pre-Experimental Measures

Following informed consent, participants completed the automated operation span measure of WM (Aospan; Unsworth, Heitz, Schrock, & Engle, 2005) and Ravens Progressive Matrices.

Participants evaluate a series of mathematical expressions and remember the letter following the expression (e.g., *is (1 * 2) + 4 = 5? P*). The number of expression-letter pairs in the series varies from three to seven. At the end of each series, participants are

prompted to recall the letters in the correct serial order. The number of correctly recalled letters across all trials was the span score.

Participants also completed a computerized version of Ravens Progressive Matrices (Raven, Court, & Raven, 1977), a widely-used test of general fluid intelligence. Participants view a 3 X 3 matrix of line drawings with one item missing. The drawings are arranged such that the transitions from row to row and column to column obey one or more rules. The participant must determine which of eight alternatives correctly completes the pattern. Participants had 25 minutes to complete as many of the 36 trials as they could. Instructions emphasized accuracy over speed.

The high and low WM span groups were primarily determined by scores on Aospan. Participants in the upper quartile reported in Unsworth et al. (2005) were considered high WM span. In an effort to ensure that participants included in the low WM span group were indeed low WM span, an additional selection requirement was imposed. Participants were included in the low WM span group only if they scored in the lower quartile on the Aospan task as well as scoring in the lower quartile on Ravens Progressive Matrices. WM and general fluid intelligence are correlated (Engle et al., 1999, Unsworth et al., 2005), and the Ravens score was included to exclude medium and high WM individuals who might not be putting forth much effort on the Aospan task and so appear to be low WM span when they are not.

Hypotheses

Hypothesis 1-5 were identical to those in Experiment 1, with the addition of a WM factor. The results predicted by the hypotheses below are presented in Figure 12, 13, and 14.

Hypothesis 1

There will be greater learning during the learning phase in the blocked than the mixed condition. This effect will depend on WM span. Greater WM capacity should result in greater differences between the conditions favoring high WM span. This hypothesis is based on the role of the memory parameters specified by Clapper and Bower (2002) and SUSTAIN. This prediction is illustrated in Figure 12 by the filled circles and squares. Experiment 1 established that the pattern-sequence effect resulted in the acquisition of category knowledge. Both the modified rational model used by Clapper and Bower (1994, 2002) to operationalize the category invention hypothesis and the implicit learning hypothesis predict greater learning in the blocked than in the mixed condition. The category invention hypothesis predicts greater acquisition of category knowledge for the high WM span participants than the low WM span participants. The implicit learning hypothesis predicts no differences between the high and low WM span participants because the category learning system is separate from the WM system.

Hypothesis 2

More learning will occur during the transfer-learning phase for the blocked condition than the mixed condition. This effect depends on WM span. The reasoning and predictions for this hypothesis is the same as that behind Hypothesis 1. The predicted pattern of results in presented in Figure 12 and the interpretation is identical to that for Hypothesis 1.

Hypothesis 3

Participants in the blocked condition will achieve above chance accuracy on the base-generalization items and the transfer-generalization items. This effect depends on WM span. As in Hypothesis 1, this is an extension of the pattern-sequence effect and the predicted role of WM consistent with the models posited by Clapper and Bower (2002) and SUSTAIN (Gureckis & Love, 2003b; Love et al., 2004). This prediction is illustrated in Figure 12 by the open circles and squares and will be supported by a main effect of WM span or greater WM span benefits for the generalization items than the learning items. If the category learning is due to implicit learning, there should be no WM span difference for the generalization items.

Hypothesis 4

Participants in the blocked condition should exhibit faster response latencies for the new items (base-generalization and transfer-generalization items) than participants in the mixed condition. This effect depends on WM span. As in Hypothesis 1, this

prediction is based on the role of memory in unsupervised learning described by Clapper and Bower (2002) and SUSTAIN (Gureckis & Love, 2003b; Love et al., 2004). Extending those models, I predict that greater WM capacity will result in greater differences between the blocked and mixed conditions. This interaction is presented in Figure 13 as the open circles and squares. If category learning results from an implicit system, there will be no difference in latencies by WM span. Again, WM differences may interact with performance in the mixed condition depending on the amount of learning and difficulty of the task, with WM differences in the same direction as for the blocked condition.

Hypothesis 5

Difficulty ratings for the base-learning and base-transfer items will be lower in the blocked condition than the mixed condition. This effect depends on WM span. Greater knowledge of the category structure will result in lower difficulty ratings. Extending the effect of memory specified by Clapper and Bower (2002), high WM span participants will demonstrate greater learning than low WM span participants and will therefore rate the items that conform to the category structure as less difficult than items that do not follow a known structure. The implicit learning hypothesis predicts that high WM span will benefit the learning items and not the transfer items. The predicted pattern of results is presented in Figure 14 by the open circles and squares.

Hypothesis 6

High WM span will be associated with faster learning. WM is consistently associated with faster learning of new information (Kyllonen & Stephens, 1990; Shute, 1991). High span participants will achieve higher accuracy than low span participants on the items they see more than once (the base-learning and transfer-learning items) regardless of condition. This effect does not discriminate between the category invention hypothesis (Clapper & Bower, 2002) and the implicit learning hypothesis (Knowlton & Squire, 1993), but serves as a manipulation check to ensure that the task is difficult enough to require some WM resources. This hypothesis is presented in Figure 12 as the difference between low WM (filled circles) and the high WM span participants spans (filled squares) on the learning items.

Data Screening

Latency scores were again normalized using natural log transformations of the raw response times for each of the response time dependent variables. Data were then examined for outliers and violations of normality using SPSS Explore. Twenty-nine extreme scores were identified on 11 of the dependent variables. Extreme scores were Winsorized separately for each dependent variable. A total of 36 replacements were made, accounting for no more than 5.56% on any single variable, and less than 3.5% of the data overall for these 11 variables.

The normal and detrended PP plots of the transformed variables were again visually evaluated for non-normality and exhibited no significant departures from

normality. The normal PP plots were very linear for all dependent variables and the detrended PP plots contained no points outside of .7 deviations across all dependent variables.

Levene's test of error variances revealed several significant violations of equality of error variances: in transfer-learning accuracy, base-learning accuracy on subsequent trials (as opposed to the first trial) associated with a plant name, transfer-learning accuracy on subsequent trials, base-transfer accuracy, transfer-generalization accuracy on subsequent trials, transfer-learning difficulty, and transfer-generalization difficulty. These violations were not considered serious, and MANOVA is robust to this type of violation especially when equal cell sizes and two-tailed significance tests are used for the analysis (Tabachnick & Fidell, 2001).

Questions 4 and 5 on the post-experimental questionnaire were again examined to ensure that participants did not have substantial knowledge of the plant words. The number of Latin plant names listed as familiar by condition and WM span are presented in Table 8. The Latin plant names listed as familiar, and the number of times they were mentioned by participants, are presented in Table 9. There were no significant differences between conditions in the number of participants identifying one or more of the Latin plant names as familiar, X^2 (3) = 3.536, p = .316, and the number of plants rated as familiar was deemed acceptable.

Statistical Analyses

The analyses from Experiment 1 were replicated with the addition of WM span as a between-subjects factor. The doubly multivariate MANOVA was replicated with the WM factor as an additional between-subjects factor. This analysis was followed by two profile analyses, one on the accuracy measures and the other on the latency measures. Each profile analysis was implemented through a 2 (Condition: blocked, mixed) X 2 (WM span: high, low) X 4 (Measure: base-learning, base-generalization, transfer-learning, transfer-generalization) mixed design MANOVA. Condition and WM span were between-subjects factors, and measure was the within-subjects factor. An additional 2 (Condition: blocked, mixed) X 2 (WM span: high, low) X 4 (Measure: base-learning, base-generalization, transfer-learning, transfer-generalization) mixed design MANOVA was conducted on the difficulty ratings. All MANOVAs were conducted using SPSS GLM using the Wilk's criteria. As in Experiment 1, follow-up tests for the within-subjects factor were conducted using SPSS repeated contrasts. The same three results from the profile analysis will be evaluated in the same manner with the following modifications.

The addition of the WM span factor creates the possibility of three levels effects: Condition, WM span, and the Condition X WM span interaction. Significant results on a test of levels indicate that there is a difference between the groups across the four dependent measures. Hypothesis 1 predicts a significant WM span levels effect. Hypothesis 1 would also be supported by a Condition X WM span interaction due to high WM span participants performing differentially better in the blocked than the mixed

condition. As long as the high WM span participants perform better than the low WM span participants, an interaction that shows greater differences between high and low WM span participants in the blocked than the mixed condition could provide support for Hypothesis 1, Hypothesis 2, and Hypothesis 3, subject to the outcomes of the flatness and parallelism analyses.

The test of flatness will be interpreted as in Experiment 1. Subject to showing that learning has taken place, a non-significant effect supports Hypothesis 3 because performance on the learning items would not differ from the generalization items. A significant flatness effect that results from lower performance on the generalization items would support the implicit learning hypothesis because greater performance on the learning items than the generalization items would indicate that participants are learning item-feature pairs as opposed to category knowledge.

The addition of the WM span factor creates three possible parallelism effects: Condition X Measure, Span X Measure, and Condition X Span X Measure. The three tests of parallelism are interpreted in the same fashion. A significant test of parallelism indicates that the differences between the groups differ for the four dependent measures. As in Experiment 1, a significant parallelism effect could support either the discovery of abstract category information (Clapper & Bower, 2002) or the implicit learning hypothesis (Ashby et al., 1998; Knowlton & Squire, 1993). The implicit learning hypothesis predicts a significant WM span X Measure test of parallelism due to WM span differences on the learning items, but not the generalization items. Hypothesis 3 would be supported by either a non-significant WM span X Measure test of parallelism or an

interaction due to greater WM span differences on the generalization items than the learning items. Similarly, a significant three-way interaction due to differential benefits of WM span on the generalization items in either the blocked or the mixed condition would support Hypothesis 3.

Significant effects were again followed by an additional analysis that compared the first items with the mean of the second and third items (subsequent items). This comparison was achieved by adding a within-subjects sequence factor to each profile analysis. A doubly multivariate design was again used to evaluate these results. The means and standard deviations for the four measures of accuracy, latency, and difficulty rating are presented in Table 10, Table 12, and Table 14, respectively. The means and standard deviations for the overall accuracy, latency, and difficulty rating measures are presented in Table 11, Table 13, and Table 15, respectively. The post-experimental questionnaire was analyzed separately and is reported below. Significant results for follow-up tests and additional analyses are at the $p < .05$ level unless otherwise reported.

Results

Manipulation Check

As in Experiment 1, learning of the response options was assessed by submitting the first six blocks of the experiment (the base items) to a 2 (WM span: high, low) X 6 (Block) X 2 (Condition: blocked, mixed) mixed design MANOVA, with block as a within subjects variable and mean accuracy for each of the six base-item blocks were the dependent variables. The results are presented in Figure 14. As in Experiment 1, the

blocked and mixed condition presented the same items in the pretraining phase, but in a different temporal order (see Table 1 for a summary of the experimental phases and data). For the purpose of this analysis, block 1 was composed of the First A and First B values, block 2 was composed of the Second A and Second B values, and block 3 was composed of the Third A and Third B values. Blocks 4, 5, and 6 followed the same temporal presentation sequence in all three conditions. Block 4 and 5 were composed of the Learning Block 1 and Learning Block 2 data, respectively. Block 6 was composed of the base-learning items from the test phase.

Like Experiment 1, there was a main effect of condition, $F(1, 104) = 293.478$, $p < .001$, $\eta^2 = .738$, a main effect of block, $F(3.899, 405.478) = 115.834$, $p < .001$, $\eta^2 = .527$, and a significant condition X block interaction, $F(3.899, 405.478) = 58.375$, $p < .001$, $\eta^2 = .360$. The three-way interaction was not significant, $F < 1$. As predicted by Hypothesis 6, there a main effect of WM span, $F(1, 104) = 11.863$, $p = .001$, $\eta^2 = .102$. Overall, the high WM span participants had higher accuracies than low WM span participants. Separate univariate analyses were conducted for each condition to locate the source of the interaction. Both were significant, $F(3.822, 194.000) = 138.915$, $p < .001$, $\eta^2 = .728$, and $F(3.496, 179.716) = 59.467$, $p < .001$, $\eta^2 = .533$, for the blocked and mixed conditions, respectively. Evaluation of the blocks using SPSS repeated contrasts revealed a pattern of results similar to Experiment 1 (see Figure 15). The blocked condition demonstrated significant improvement from the first to second block and second to third block, only to perform significantly worse on the fourth block, (again coinciding with the start of the intermixed learning trials), only to become significantly better on the fifth block. The

final block did not differ from the fifth block. The mixed condition improved significantly from the first to second block, rose a non-significant amount on the third block, then improved significantly from the third to fourth block. The improvement from the fourth to fifth block was almost significant, p = .051, then scores leveled out on the final block. This pattern is the same as the one reported by Clapper and Bower (2002) and found in Experiment 1, which consisted of significant learning across the pretraining trials for participants in the blocked condition only. A oneway ANOVA confirmed that the difference between the conditions at the end of the pretraining phase (blocks 1-3) was significant, $F(2, 63) = 187.346$, $p < .001$; participants in the blocked condition achieved higher accuracy ($M = .9622$, $SD = .0644$) than participants in the mixed condition ($M = .5104$, $SD = .1384$).

First Versus Subsequent Item Analysis

The learning trends were again separated into performance on the first items and performance on the subsequent items for the six blocks of the base-learning items. The high WM span participants achieved higher accuracies ($M = .667$, $SD = .088$) than the low WM span participants ($M = .608$, $SD = .088$) overall, $F(1, 104) = 12.824$, $p < .001$, $\eta^2 = .110$. None of the WM span interactions were significant, all *ps* > .12. The Block X Condition X Time test of parallelism was significant, Wilk's $\Lambda = .488$, $F(5, 100) = 20.982$, $p < .001$, $\eta^2 = .512$. Separate tests were conducted for the first and subsequent items for the blocked and mixed conditions to locate the source of the interaction.

The follow-up analyses revealed a different pattern of results for the first-items, but the same pattern of results for the subsequent items. The results are presented in Figure 16. For the first-items, the Condition X Block interaction was significant, $F(5, 102) = 64.135$, $p < .001$, $\eta^2 = .759$. Univariate ANOVAs were then conducted to locate the source of the interaction. The effect of Block was significant for the first items in the blocked, $F(3.679, 194.984) = 158.442$, $p < .001$, $\eta^2 = .749$, and mixed conditions, $F(4.292, 227.463) = 22.741$, $p < .001$, $\eta^2 = .300$. The results for the first-items were very different from the overall results. In the blocked condition, accuracy on the first items improved from the first to the second block, stayed the same between the second and third block, then declined significantly from the third block to the fourth block and remained the same for the fifth and sixth blocks. In the mixed condition, accuracy on the first items improved from the first to the second block then stayed the same for the rest of the learning blocks. This pattern is different from the overall results and the results of Experiment 1 because performance for participants in the blocked conditioned leveled off after the first block. A oneway ANOVA confirmed that the participants in the blocked condition achieved higher accuracy ($M = .9502$, $SD = .0701$) than participants in the mixed condition ($M = .4931$, $SD = .1641$) at the end of the pretraining phase, $F(1, 106) = 345.405$, $p < .001$.

For the subsequent-items, the Condition X Block interaction was significant, $F(5, 102) = 72.767$, $p < .001$, $\eta^2 = .781$. Univariate ANOVAs were again conducted to locate the source of the interaction. The effect of Block was significant for subsequent items in the blocked, $F(3.218, 186.118) = 100.010$, $p < .001$, $\eta^2 = .654$, and mixed conditions,

$F(3.261, 172.810) = 42.287$, $p < .001$, $\eta^2 = .444$. The results for the subsequent-items were very similar to the overall results. In the blocked condition, accuracy increased from the first to second block, and again from the second to third block. Accuracy then declined between the third and fourth block, then increased significantly on the fifth block and remained unchanged on the final block. The mixed condition improved significantly from the first to the second block, but remained stable from the second to the third block before improving significantly again on the fourth block, and between the fourth and fifth blocks. There was no further improvement on the final block. This pattern is identical to the overall results and the results of Experiment 1. A oneway ANOVA confirmed that the participants in the blocked condition achieved higher accuracy ($M = .9682$, $SD = .0691$) than participants in the mixed condition ($M = .5191$, $SD = .1694$) at the end of the pretraining phase, $F(1, 106) = 325.487$, $p < .001$.

The Surprise Effect

The surprise effect found in Experiment 1 was replicated in Experiment 2. This effect is visible in their results as a sharp drop in accuracy between then end of the Category A trials and the first Category B trials for the blocked condition, followed by a rapid increase in accuracy in the Category B items to the high levels obtained for the Category A items. This effect is present in the current results presented in Figure 14 as the significant drop between the Third A and the First B items for the blocked condition, $t(107) = 17.101$, $p < .001$. The effect is also true for both the first and subsequent items in Figure 5, $t(107) = 10.480$, $p < .001$ and $t(107) = 16.457$, $p < .001$, respectively.

Omnibus Test: Accuracy and Latency

The behavioral measures of category learning, accuracy and latency, were first evaluated in a 2(condition: blocked, mixed) X 4 (measure: base-learning, base-generalization, transfer-learning, transfer-generalization) X 2(WM span: high, low) doubly multivariate MANOVA conducted on the accuracy and latency dependent measures. As in Experiment 1, this analysis is an omnibus test for Hypotheses 1-4 and the means are presented in Tables 10 and 11 for the accuracy measures, and Table 12 and 13 for the latency measures. The three-way interaction was not significant, Wilk's Λ = .961, $F(6, 99) < 1$. The main effect of condition was significant, Wilk's $\Lambda = .630$, $F(2, 103) = 30.273$, $p < .001$, $\eta^2 = .370$, as was the main effect of measure, Wilk's $\Lambda = .575$, $F(6, 99) = 12.190$, $p < .001$, $\eta^2 = .425$. These two main effects were qualified by a significant condition X measure interaction, Wilk's $\Lambda = .880$, $F(6, 99) = 2.248$, $p = .045$, $\eta^2 = .120$, indicating that the composite accuracy and latency measures were not the same on the base-learning, base-generalization, transfer-learning, and transfer-generalization items in the two conditions. The main effect of WM span was also significant, Wilk's Λ = .917, $F(2, 103) = 4.674$, $p = .011$, $\eta^2 = .083$. The condition X WM span interaction was not significant, nor was the measure X WM span interaction, indicating that there were consistent differences between the high and low WM span participants on the accuracy and latency composite variables. Separate profile analyses on the accuracy and latency measures were performed to locate the source of the interactions and address the experimental hypotheses.

First Versus Subsequent Item Analysis

Examination of the first and subsequent items via the addition of time factor (first, subsequent) to the analysis revealed the same significant main effects of condition, Wilk's $\Lambda = .733$, $F(2, 103) = 18.757$, $p < .001$, $\eta^2 = .267$, and measure, Wilk's $\Lambda = .610$, $F(6, 99) = 10.529$, $p < .001$, $\eta^2 = .390$. These two main effects were again qualified by a significant condition X measure interaction, Wilk's $\Lambda = .883$, $F(6, 99) = 2.189$, $p = .050$, $\eta^2 = .117$. The main effect of WM span was significant, Wilk's $\Lambda = .903$, $F(2, 103) = 5.535$, $p = .005$, $\eta^2 = .097$. The main effect of time was significant, Wilk's $\Lambda = .259$, $F(2, 103) = 147.077$, $p < .001$, $\eta^2 = .741$, but was qualified by a significant condition X time interaction, Wilk's $\Lambda = .606$, $F(2, 103) = 33.424$, $p < .001$, $\eta^2 = .394$. Follow-up profile analyses were conducted on the accuracy and latency measures separately and are reported below.

Hypotheses 1, 2, and 3

Hypotheses 1, 2, and 3 were based on the accuracy measures and were evaluated together in the same profile analysis as in Experiment 1 with the addition of the WM factor. The results are presented in Figure 17. The mean accuracies for the experimental cells are presented in Table 10 and the mean accuracies of the experimental conditions collapsed across WM span are presented in Table 11. None of the tests of parallelism were significant, all *ps* > .12, permitting a clean interpretation of the two test of levels

(Condition and WM span) and the test of flatness, which are discussed below as they apply to Hypotheses 1-3.

Hypothesis 1

There will be greater learning during the learning phase in the blocked than the mixed condition. This effect will depend on WM span. Greater WM capacity should result in greater differences between the conditions favoring high WM span. This hypothesis was not supported. The results are presented in Figure 17 as the difference between the blocked (squares) and mixed (triangles) conditions and between the high WM span participants (filled shapes) and low WM span participants (open shapes) on the base-learning items. The means are presented in Table 10 and Table 11. The predicted interaction of condition and WM span was not significant based on the non-significant Measure X Condition X WM span parallelism effect, $F(3, 102) = 1.485$, $p = .223$. As in Experiment 1, participants in the blocked condition achieved higher accuracies than participants in the mixed condition; the test of levels for the condition factor was significant, $F(1, 104) = 53.828$, $p < .001$, $\eta^2 = .341$. This replicates the results of Experiment 1 and the findings of Clapper and Bower (2002). The high WM span participants achieved higher accuracies than the low WM span participants, the levels effect of WM span was significant, $F(1, 104) = 4.310$, $p = .040$, $\eta^2 = .040$. Single sample t-tests revealed that the high and low WM span participants in both the blocked and mixed conditions performed above chance on the base-learning items. The above chance performance by participants in the blocked condition replicates the findings of

Experiment 1 and Clapper and Bower (2002). These results also support a role for WM in unsupervised category learning but do not support Clapper and Bower's (1994, 2002) use of the rational model to explain the pattern-sequence effect.

Hypothesis 2

More learning will occur during the transfer-learning phase for the blocked condition than the mixed condition. This effect depends on WM span. Greater WM capacity should result in greater differences between the conditions favoring high WM span. This hypothesis was not supported. The results are presented in Figure 17 as for Hypothesis 1 but on the transfer-learning items. The predicted interaction of condition and WM span was not significant based on the non-significant Measure X Condition X WM span parallelism effect, $F(3, 102) = 1.485$, $p = .223$. As in Experiment 1, participants in the blocked condition achieved higher accuracies than participants in the mixed condition on the transfer-learning items, $F(1, 104) = 53.828$, $p < .001$, $\eta^2 = .341$. This replicates the results of Experiment 1. The high WM span participants achieved higher accuracies than the low WM span participants, $F(1, 104) = 4.310$, $p = .040$, $\eta^2 = .040$. As in Experiment 1, all participants in the blocked condition performed above chance, as did the high WM span participants in the mixed condition. These results also support a role for WM in unsupervised category learning but do not support Clapper and Bower's (1994, 2002) use of the rational model to explain the pattern-sequence effect.

Hypothesis 3

Participants in the blocked condition will achieve above chance accuracy on the base-generalization items and the transfer-generalization items. This effect depends on WM span. This is the critical hypothesis of the experiment that tests whether there is a role for WM in the acquisition of abstract category knowledge in the pattern-sequence effect. This hypothesis was not supported based on the non-significant Measure X Condition X WM span parallelism effect, $F(3, 102) = 1.485$, $p = .223$. As in Experiment 1, participants in the blocked condition achieved higher accuracies than participants in the mixed condition on the base-generalization and transfer-generalization items, $F(1, 104) = 53.828$, $p < .001$, $\eta^2 = .341$. This replicates the results of Experiment 1 and provides strong support for the acquisition of abstract category knowledge by participants in the blocked condition. The high WM span participants achieved higher accuracies than the low WM span participants on the base-generalization and transfer-generalization items, $F(1, 104) = 4.310$, $p = .040$, $\eta^2 = .040$. Only participants in the blocked condition performed above chance on the base-generalization items. Only the low WM span participants in the mixed condition failed to score above-chance on the transfer-generalization items (see Table 10 for means). These results again support a role for WM in unsupervised category learning and provide the strongest evidence against Clapper and Bower's (1994, 2002) use of the rational model to explain the pattern-sequence effect.

There were no significant differences between the learning accuracies and the generalization accuracies, Wilk's $\Lambda = .96$, $F(3, 102) = 1.343$, $p = .265$. This effect replicates Experiment 1 and provides strong support for the hypothesis that the pattern-sequence effect results in the acquisition of abstract category knowledge. In the absence

of an interaction, this effect creates further difficulties for Clapper and Bower's (2002) use of the rational model when combined with the main effect of Condition and the main effect of WM span. The modified rational model of Clapper and Bower (2002) has no way of explaining an effect of WM in the absence of an interaction with the blocking manipulation, and WM had a consistent benefit for both the learning and generalization items.

Accuracy on the first block on transfer items was examined as in Experiment 1 as an additional test of Hypothesis 3. As in Experiment 1, participants in the blocked condition achieved higher accuracy than did participants in the mixed condition on the first block of transfer items, $F(1, 104) = 44.486$, $p < .001$, $\eta^2 = .300$. The high WM span participants achieved higher accuracy than did the low WM span participants on the first block of transfer items, $F(1, 104) = 4.778$, $p = .031$, $\eta^2 = .044$. Single sample t-tests against chance performance revealed that both the high ($M = .785$, $SD = .092$) and low WM span participants ($M = .709$, $SD = .141$) in the blocked condition performed significantly above chance, $t(26) = 7.693$, $p < .001$, and $t(26) = 16.161$, $p < .001$, respectively. The high WM span participants in the mixed condition also performed significantly above chance ($M = .586$, $SD = .169$), $t(26) = 2.638$, $p = .014$. The WM span X condition interaction was not significant, $F < 1$. These results support a role for WM in the acquisition of abstract category knowledge, but do not support Clapper and Bower's (1994, 2002) modification of the rational model because their model predicts greater benefit for the high WM span participants in the blocked condition.

First Versus Subsequent Item Analysis

The addition of the time factor revealed a more complex story for the relationship between WM and the pattern-sequence effect. As in Experiment 1, the Condition X Measure parallelism effect was significant, Wilk's $\Lambda = .606$, $F(1, 104) = 67.496$, $p < .001$, $\eta^2 = .394$. Separate follow-up analyses revealed that the blocked condition was indistinguishable from the mixed condition on the first items, $F(2, 59) = 1.613$, $p = .208$, but performed significantly better on the subsequent items, $F(2, 59) = 62.224$, $p < .001$, $\eta^2 = .678$. The levels effect of WM was also significant, Wilk's $\Lambda = .606$, $F(1, 104) = 6.218$, $p = .014$, $\eta^2 = .056$. The results are presented in Figure 17 and are elaborated upon below as they apply to Hypotheses 1-3.

Hypothesis 1. These results replicate the results of Experiment 1, but fail to support the hypothesized role of WM. As in Experiment 1, participants in both the blocked and mixed condition performed equally well on the first items, $F(2, 59) = 1.613$, $p = .208$, but only the participants in the blocked condition were able to use the information in the feedback from the first item to improve on the subsequent items, $F(2, 59) = 62.224$, $p < .001$, $\eta^2 = .678$. Single sample t-tests revealed that all participants had better than chance accuracy on the subsequent base-learning items, but only the high WM span participants in the blocked and mixed conditions performed significantly better than chance on the base-learning first-items, all *ps* < .05. Hypothesis 1 predicts that high WM will result in greater benefit in the blocked condition than the mixed condition. The present results support a role for WM in the acquisition of category knowledge but do not support Hypothesis 1 because high WM span provided an equal benefit in the blocked

and mixed conditions for the first items, $F(1, 104) = 9.250$, $p = .003$, $\eta^2 = .082$, but not for the subsequent items, $F(1, 104) = 1.656$, $p = .201$.

Hypothesis 2. These results fail to support Hypothesis 2 for the same reason given for Hypothesis 1. Replicating the results of Experiment 1, participants in the blocked condition performed no differently than participants in the mixed condition on the first transfer-learning items, $F(2, 59) = 1.613$, $p = .208$, but did significantly better on the subsequent transfer-learning items, $F(2, 59) = 62.224$, $p < .001$, $\eta^2 = .678$. Single sample t-tests revealed that all participants performed significantly better than chance on the transfer-learning subsequent-item accuracy measures, as did all participants except the low WM span participants in the mixed condition, all *ps* < .05 (see Table 10 for mean accuracies). Again, the results support a general role for WM but not in the manner predicted by Clapper and Bower (2002). The high WM span participants had higher accuracies on the first items, $F(1, 104) = 9.250$, $p = .003$, $\eta^2 = .082$, but not on the subsequent items, $F(1, 104) = 1.656$, $p = .201$.

Hypothesis 3. Hypothesis 3 was not supported. As in Experiment 1, participants in the blocked condition performed no differently than participants on the base-generalization and transfer-generalization first-item accuracies, $F(2, 59) = 1.613$, $p = .208$, but were the only ones able to use the feedback to do significantly better on the subsequent items, $F(2, 59) = 62.224$, $p < .001$, $\eta^2 = .678$. Single sample t-tests revealed that both high and low WM span participants in the blocked condition performed significantly better than chance on the base-generalization and transfer-generalization subsequent-item accuracy measures, all *ps* < .001. The high WM span participants also

achieved above-chance accuracy on the base-transfer first-items, $t(26) = 2.580$, $p = .016$. The means are presented in Table 10. The non-significant flatness effect for the first and subsequent accuracies replicated the findings of Experiment 1 and supported the hypothesis that the pattern-sequence produces abstract category knowledge. Hypothesis 3 was not supported because the results indicated that WM span provided equal benefit to those in the blocked condition as to those in the mixed condition on the first items, $F(1, 104) = 9.250$, $p = .003$, $\eta^2 = .082$, but did not result in higher accuracies on the subsequent items in either the blocked or the mixed conditions, $F(1, 104) = 1.656$, $p = .201$.

Examination of the first and subsequent items revealed different patterns of results for the first block of transfer items. Two additional 2 (WM span: high, low) X 2 (Condition: blocked, mixed) between subjects ANOVA were conducted; one had first-item accuracy on the first transfer block as the dependent variable, the other had subsequent-item accuracy on the first transfer block as the dependent variable. The high WM span participants achieved higher accuracy ($M = .523$, $SD = .126$) than did the low WM span participants ($M = .453$, $SD = .154$) on the first-items in the first block of transfer items, $F(1, 104) = 6.750$, $p = .011$, $\eta^2 = .061$, because the low WM span participants scored below chance, $t(53) = 2.271$, $p = .027$, and the high WM span participants scored no different than chance, $t(53) = 1.353$, $p = .182$. The blocked and mixed conditions were not significantly different, and the interaction was not significant, both $Fs < 1$. In the analysis of the subsequent items, participants in the blocked condition had higher accuracies ($M = .872$, $SD = .152$) than participants in the mixed condition (M

= .605, *SD* = .210), $F(1, 104) = 57.537$, $p < .001$, $\eta^2 = .356$. The main effect of WM span and the interaction were not significant, both *ps* > .12.

In the blocked condition, both high and low WM span participants performed significantly better than chance on all four subsequent-item measures (see Table 10). The high WM span participants in the blocked condition also performed significantly better than chance on the base-learning and base-generalization first-item accuracies. The low WM span participants did not exceed chance performance on any of the first-item accuracies in the blocked condition. There were differences in the mixed condition. The high WM span participants in the mixed condition performed significantly better than chance on subsequent-item accuracies for all four measures, indicating that they acquired some knowledge of the category structure. The low WM span participants in the blocked condition performed significantly above chance on the subsequent-item base-learning, transfer-learning, and transfer-generalization items. The high WM span participants in the mixed condition performed significantly better than chance on the first-item base-learning trials. The low WM span participants in the mixed condition did not perform above chance on any of the first-item measures (see Figure 18).

Hypothesis 4

Participants in the blocked condition should exhibit faster response latencies for the new items (base-generalization and transfer-generalization items) than participants in the mixed condition. This effect depends on WM span. This hypothesis was not supported. The results are presented in Figure 19. The results from Experiment 1 were

replicated with one small difference. The blocked condition did not answer any more quickly than the mixed condition on any of the four measures; neither the Measure X Condition test of parallelism nor the Condition test of levels were significant, Wilk's Λ = .939, $F(3, 102) = 2.204$, $p = .092$, and $F < 1$, respectively. As in Experiment 1, the test of flatness was significant, Wilk's $\Lambda = .598$, $F(3, 102) = 22.830$, $p < .001$, $\eta^2 = .267$, indicating that the four latency measures were not the same. Contrasts performed using SPSS Contrasts revealed that participants responded to the transfer-generalization items faster than the transfer-learning items, $F(1, 104) = 18.443$, $p < .001$, $\eta^2 = .151$, which were answered faster than the base-generalization items, $F(1, 104) = 31.046$, $p < .001$, $\eta^2 = .230$, which were answered just as quickly as to the base learning items, $F(1, 104) = 2.941$, $p = .089$. These results are similar to those from Experiment 1 except that the transfer-generalization items were answered faster than the transfer-learning items in Experiment 2 only. Even though the flatness effect was significant, the results support an overall learning effect rather than an effect due to the type of measure. The critical hypothesis, that high WM span participants will respond even faster in the blocked condition than in the mixed condition, was not supported. There were no response latency differences between the high and low WM span participants, $F(1, 104) = 2.729$, $p = .102$, or differentially by span between conditions, $F(1, 104) = 2.515$, $p = .116$.

First Versus Subsequent Item Analysis

The first and subsequent latencies did not change the conclusions from the overall data; Hypothesis 4 was not supported. The results are presented in Figure 20. Wilk's Λ

= .479, $F(1, 104) = 113.345$, $p < .001$, $\eta^2 = .521$, which was qualified by a significant condition X span X time three way interaction, Wilk's $\Lambda = .956$, $F(1, 104) = 4.770$, $p = .031$, $\eta^2 = .044$. No other interactions were significant, though the condition X span levels interaction was marginally significant, $F(1, 104) = 3.679$, $p = .058$, $\eta^2 = .034$ (see Figure 20). A Cicchetti correction was applied to a comparison of the cell means to locate the source of the interaction. This correction divides alpha based on the number of comparisons of cell means made (Toothaker, 1993). Four comparisons were made, comparing high versus low WM span participants by condition separately for the first-item and subsequent-item latencies. The significant interaction was due to faster response times for the first items by the high WM span participants than low WM span participants in the mixed condition only; the remaining cell means comparisons were not significant.

Hypothesis 5

Difficulty ratings for the base-learning and base-transfer items will be lower in the blocked condition than the mixed condition. This effect depends on WM span. This hypothesis was not supported. The results are presented in Figure 21. The results from Experiment 1 were replicated. The blocked condition reported lower difficulty ratings than did the mixed condition, the test of levels was significant, $F(1, 104) = 39.251$, $p < .001$, $\eta^2 = .274$. The test of flatness was not significant, Wilk's $\Lambda = .964$, $F(3, 102) = 1.287$, $p = .283$. Contrary to the WM span prediction in Hypothesis 5, the task was no easier for high WM span participants than it was for low WM span participants; the tests

of parallelism involving WM span were not significant, all *ps* > .12, nor was the test of levels for WM span, $F(1, 104) = 2.225$, $p = .139$.

First Versus Subsequent Item Analysis

The difficulty data were reexamined with the addition of the time factor. The results are presented in Figure 21. The first items were rated as significantly more difficult than the subsequent items, the main effect of Time was significant, Wilk's $\Lambda = .545$, $F(1, 104) = 86.735$, $p < .001$, $\eta^2 = .455$. This effect was qualified by a significant Condition X Time interaction, Wilk's $\Lambda = .765$, $F(1, 104) = 31.948$, $p < .001$. A Cicchetti correction was applied to a comparison of the cell means to locate the source of the interaction. This correction divides alpha based on the number of comparisons of cell means made (Toothaker, 1993). The interaction was due to lower subsequent difficulty ratings for participants in the blocked condition, but this difference was not significant for participants in the mixed condition. The effect of Condition on the subsequent but not the first items replicates the condition effect found in Experiment 1 for Hypothesis 5. The high and low WM span participants rated the task as equally difficult, $F(1, 104) = 2.179$, $p = .143$, which does not support a role for WM in perceived difficulty of the task. None of the other parallelism effect were significant, though the measure X time test of parallelism was marginally significant, Wilk's $\Lambda = .935$, $F(3, 102) = 2.351$, $p = .077$.

Post-Experimental Questionnaire

The post-experimental questionnaires were coded as described in Appendix C. The number of participants in each condition who reported actively searching for patterns and who exhibited awareness of the category structure are presented in Table 16. As in Experiment 1, participants in the blocked condition did not engage in significantly more active hypothesis testing than did participants in the mixed condition, $X^2 < 1$, but did report significantly more awareness of the category structure, X^2 (1) = 9.990, p = .002. There were differences in awareness due to WM span, but not in active hypothesis testing. Two Mantel-Haenszel tests were performed to compare the pattern of data for high versus low WM span participants, one compared high and low WM spans by condition on awareness, the other compared high and low WM spans by condition on their self-reported search for patterns. This analysis compares the pattern of data in one 2 X 2 contingency table with the pattern of data in another 2 X 2 contingency table and produces a statistic which is evaluated as a z-score (Conover, 1999). The test of awareness revealed a significant difference in the patterns of awareness by condition for the high span participants than the low span participants, Mantel-Haenszel statistic = 3.395, $p < .001$. The high WM span participants reported significantly less awareness in the mixed condition than did the low WM span participants, X^2 (1) = 8.206, p = .004. The test of searching revealed no significant differences in the patterns of active hypothesis testing by condition for the high and low WM span participant, Mantel-Haenszel statistic = .194, p = .349.

Discovery Versus Acquisition of Category Information

There are two hypotheses regarding the role of WM in the acquisition of category information. The category invention hypothesis (Clapper & Bower, 1994, 2002) posits that WM is involved in the discovery of category information. Another possibility is that both low and high WM span participants intentionally search for information at the same rates, but the high WM span participants are more successful at learning the structure once it is found (Ashby et al., 1998; Ashby et al., 1999; Ashby & Waldron, 1999). According to this position, high WM span participants who look for patterns should demonstrate greater knowledge of the category structure because hypothesis testing is an explicit process and is optimized to acquire the unidimensional rule structure employed here (Ashby et al., 1999). This should appear both on the post-experimental questionnaire assessment of awareness, and in the behavioral accuracy data.

A comparison of the high and low WM span participants who reported actively searching for patterns on their post-experimental questionnaires revealed that high WM span participants were no more aware of the category structure than were low WM span participants, X^2 (1) = 1.478, p = .224. This pattern was true regardless of experimental condition, Mantel-Haenszel statistic = 1.264, p = .844. This result does not support the two-system position because there was not increased awareness for high WM spans based on active searching for patterns.

When the analysis on the accuracy dependent variables was replicated using only those participants who actively searched for patterns and were classified as having some or full awareness of the category structure on the post-experimental questionnaire, there

were no higher order interactions involving WM span, but there was a significant effect of WM span, $F(1, 35) = 11.373$, $p = .002$, $\eta^2 = .245$, with high spans ($M = .768$, $SE = .021$) performing better than low spans ($M = .671$, $SE = .021$). At first, this would seem to support explicit hypothesis testing that having greater WM resources improved learning the same amount regardless of the condition. However, according to this position, there should be no differences between the high and low WM span participants who did not search for patterns in either awareness or performance as WM benefits would be accompanied by explicit intention and awareness. Therefore, the results of participants who did not search for patterns but who indicated awareness must also be examined.

The analysis on the accuracy dependent variables was replicated using only those participants who did not actively search for patterns and indicated some or full awareness of the category structure on their post-experimental questionnaires. There were no higher order interactions involving WM span, but there was a marginally significant effect of WM span, $F(1, 44) = 3.845$, $p = .056$, $\eta^2 = .080$, with high spans ($M = .714$, $SE = .033$) performing better than low spans ($M = .628$, $SE = .023$). This result suggests a role for WM in the acquisition of category knowledge even when the participant is *not* looking for patterns. Either there is a single category learning system, or there is a less rigid separation of the hypothesis testing and implicit learning systems. The results of the present experiment cannot differentiate between these two explanations.

Discussion

Experiment 2 replicated the findings of Experiment 1 and provided support for a role for WM in the acquisition of category knowledge in unsupervised category learning but did not support the hypotheses based on Clapper and Bower's (2002) modification of the rational model.

Replication of Experiment 1

The five extensions of Clapper and Bower (1994, 2002) found in Experiment 1 were replicated. As in Experiment 1, participants in the blocked condition achieved better than chance performance not only on the learning items, but also on the novel item-feature pairs that had not been seen during learning. Furthermore, these generalization items were no different from the learning items. This was true for both the base items and the transfer items. Participants in the blocked condition were again able to apply their category knowledge to new items and achieve better than chance performance on the first block of the transfer items. The different patterns of results found for the first and subsequent items also support the conclusion that the pattern-sequence effect gives rise to abstract category knowledge. The second accuracies greatly exceeded the first accuracies for the four main accuracy measures: base-learning, transfer-learning, base-generalization, and transfer-generalization. Finally, the analysis of the difficulty ratings supports the conclusions drawn from the accuracy analyses. The first items were rated as more difficult in the blocked condition than the subsequent items, supporting the conclusion that participants had to remember the first items, but used category knowledge

to answer the subsequent items. The behavioral data indicate that participants often guessed on the first item to discover which group the plant belonged to, then used that information to get the next two items for that plant correct.

Experiment 2 also supports the conclusion that participants found learning the feature-feature relationships easier than learning the item-feature relationships. The pattern-sequence manipulation again resulted in robust differences between the blocked and the mixed conditions on the subsequent items, often having guessed on the first feature for each Latin plant name. These results indicate that participants learned the category knowledge quickly and accurately.

The post-experimental questionnaire results again mirror the behavioral results. There was greater awareness in the blocked condition than in the mixed condition but participants in the blocked condition did not report significantly more engagement in active search for patterns in the experimental materials, which indicates that awareness of the categories resulted spontaneously as Clapper and Bower (1994, 2002) proposed, and not as the result of deliberate hypothesis testing (e.g., Ashby et al., 1998; Billman & Knutson, 1996).

WM in Category Learning

There are three main ways that WM could play a role in category learning. WM could result in faster item learning due to greater item-feature memory, but not result in the acquisition of abstract category information. This is the position advocated by the implicit learning hypothesis (e.g., Knowlton & Squire, 1993), and it supported by several

studies that have found that higher WM is associated with faster learning or greater learning gains (e.g., Kyllonen & Stephens, 1990; Shute, 1991). WM could allow participants to discover the category structure more quickly and reliably as proposed by Clapper & Bower's (2002) adaptation of the rational model. Finally, WM could allow participants to acquire the category structure more quickly through active hypothesis testing (Ashby et al., 1998, Billman & Knutson, 1996) and superior ability to relate items to the categories to which they belong (exemplar-category memory). There is some evidence for each position in the present study.

WM and Item Learning

There was evidence that higher WM resulted in faster learning of the learning item-feature pairs. This finding was supported in the current study in several ways. First, Hypothesis 6 was supported; the high WM span participants had higher learning accuracies than did low WM span participants across the six learning blocks.

Second, the WM levels effect means that the high WM span participants had significantly higher accuracy on the first items during the learning trials (both base-learning and transfer-learning) than did the low WM span participants. The high WM span participants achieved above-chance accuracy on these items in both the blocked and mixed conditions. These items could have been memorized as independent item-feature pairs, which would allow participants to perform significantly above chance. This finding can also support the exemplar-category learning hypothesis, which is discussed below.

WM in the Discovery and Acquisition of Category Knowledge

The critical hypothesis of Experiment 2, that WM played a role in learning category level information, received some support. The levels effect of WM span coupled with the non-significant flatness effect for the accuracy measures supports Hypothesis 3 that WM plays a part in acquiring category knowledge. This is because feature-item memory could not have been used on the generalization items because they were not encountered during learning. Only greater knowledge of the category structure would permit the high WM span participants to achieve significantly higher accuracy on these new items. There are two possibilities for how knowledge of the category structure could allow the high WM span participants to perform better than the low WM span participants on the new items. One possibility is that WM aids the discovery of the category structure. This is the position advocated by Clapper and Bower's (2002) modification of Anderson's (1991) rational model and is the basis of the hypotheses explored in the present study. The other possibility is that WM aids the learning of category information by engaging in a repeated process of hypothesis testing and is also associated with intentional learning of exemplar-category associations. This position is consistent with the explicit hypothesis testing role for WM (Ashby et al., 1998, Ashby et al., 1999; Billman & Knutson, 1996). The present study was not designed to differentiate these two alternatives, but it does provide some insight as to which option is most likely.

The first and subsequent-item data paint a mixed picture for the role of WM in category learning. For the accuracy measures, the span differences were due to high WM

span participants performing significantly better than low WM span participants and significantly above chance on the first-item accuracy measures, but not the subsequent versions of these items. This was true for high WM span participants in both the blocked and mixed conditions. On the one hand, this supports a role for WM span in the acquisition of category level information because if their performance was due strictly to superior item-feature memory, then the high WM span participants should not have exceeded chance performance on the first generalization items. On the other hand, the modified rational model used by Clapper and Bower (2002) predicts that the high WM span participants should benefit more than the low WM span participants in the mixed condition, which did not occur. It is also possible that high WM span was not responsible for the acquisition of the category knowledge per se, but for learning exemplar-category information. This alternative is undermined by the fact that the high WM span participants had higher accuracy on the first block of transfer items. If the WM span advantage was due to exemplar-category memory, then high WM span participants should not have exceeded the performance of low WM span participants on these trials with unfamiliar items. This result is in accordance with the theoretical description of the category invention effect proposed by Clapper and Bower (2002).

The present framework predicts that WM span would result in greater subsequent-item generalization accuracy, both base-generalization and transfer-generalization. It is not clear whether this hypothesis is wrong, or whether it is confounded by a likely ceiling effect. High and low WM span participants in the blocked condition exhibited extremely high accuracy on the subsequent-item base-learning, base-generalization, transfer-

learning, and transfer-generalization (see Table 10). These accuracies were within one standard deviation of perfect accuracy and were significantly negatively skewed (all *ps* < .05). In either case, there is evidence for WM involvement in unsupervised category learning, but not based on the model proposed by Clapper and Bower (1994, 2002).

That the high WM span participants had significantly higher accuracy than did low WM span participants on the first block of transfer trials supports a role for WM in the acquisition of category level information. This conclusion is based on a closer examination of the first versus the subsequent items, which revealed that the difference between high and low WM span participants was due to higher accuracy on the subsequent items, not the first items, by the high WM span participants. Again, if higher WM span only improved memory and not category level knowledge, then the high WM span participants should not have outperformed the low WM span participants on the subsequent items because they would not have any item-feature memories from which to base their answers.

Hypothesis 4 predicted that the high WM span participants would respond faster than the low WM span participants on the generalization items. There were no significant differences in response latencies by WM span, but WM span did matter for the accuracy measures. This pattern of results indicates that response latency may not be a sensitive or appropriate measure of category knowledge in this paradigm, but does eliminate the possibility that participants in the blocked condition engaged in a speed-accuracy trade-off.

Hypothesis 5 was not supported. The high WM spans did not differ in their difficulty ratings from the low WM span participants. This was true overall, and for both the first and subsequent items. The only differences in difficulty ratings were due to lower ratings provided by participants in the blocked condition than the mixed condition.

Support for a role for WM in the acquisition of abstract category knowledge comes from the differences between the high and low WM span participants on the analysis of the post-experimental questionnaire. High and low WM spans did not report different amounts of intentional hypothesis testing, but the high WM span participants reported greater awareness of at least some of the category structure than the low WM span participants in the mixed condition. While this is theoretically consistent with the category invention hypothesis of Clapper and Bower (1994, 2002), the behavioral data did not follow the pattern predicted by their model. There is a role for WM in unsupervised category learning, but not as Clapper and Bower propose.

Table 8. Number of Familiar Plants by Condition by WM Span for Experiment 2.

Number of Familiar Plants	Blocked Condition		Mixed Condition		
	High	Low	High	Low	Total
0	20	23	19	17	79
1	5	1	5	3	14
2	1	3	3	6	13
3	1	0	0	1	2
Total	27	27	27	27	108

Table 9. Familiar Plant Names by Condition by WM Span for Experiment 2.

	Blocked Condition		Mixed Condition		
Plant Name	High	Low	High	Low	Total
Artex				1	1
Carex		1	1		2
Citax		1			1
Cyria				1	1
Hedera			1		1
Hordeum	1				1
Mentha			3	3	6
Nepeta	1	1			2
Nolana				1	1
Prunus		1			1
Rumex				1	1
Salix	1				1
Salvia	4	1	3	4	12
Sedum		1		1	2
Taxus	1		1		2
Verbena	2	1	2	3	8
Vertix			1		1
Vicia			1		1
Vitex				1	1

Table 10. Accuracy Measures by Condition by WM Span for Experiment 2.

	Blocked				Mixed			
	High WM Span		Low WM Span		High WM Span		Low WM Span	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
			Base-Learning					
Overall	0.7963*	0.1049	0.7384*	0.1425	0.6262*	0.1507	0.5683*	0.1494
First	0.5784*	0.1858	0.5091	0.1759	0.5868*	0.1979	0.5380	0.1875
Second	0.8980*	0.1020	0.8512*	0.1603	0.6447*	0.1992	0.5880*	0.2137
			Base-Generalization					
Overall	0.7963*	0.1456	0.7454*	0.1273	0.5602	0.2223	0.5370	0.1638
First	0.6207*	0.2431	0.4635	0.1936	0.5046	0.3071	0.4360	0.2265
Second	0.8850*	0.1567	0.8797*	0.1482	0.5996*	0.2161	0.5877*	0.2074
			Transfer-Learning					
Overall	0.7836*	0.1453	0.7419*	0.1190	0.5891*	0.1934	0.5532	0.1740
First	0.5614	0.1649	0.5237	0.1626	0.5321	0.1561	0.4725	0.1899

Table 10 Continued.

	Blocked				Mixed			
	High WM Span		Low WM Span		High WM Span		Low WM Span	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Second	0.8988*	0.1683	0.8418*	0.1297	0.6139*	0.2567	0.5962*	0.2234
Transfer-Generalization								
Overall	0.7940*	0.1448	0.7176*	0.1525	0.6296*	0.2101	0.5440	0.2035
First	0.5127	0.2867	0.4390	0.2845	0.5953	0.2620	0.4528	0.2741
Second	0.9327*	0.1186	0.8807*	0.1406	0.6618*	0.2575	0.5853	0.2690
Total								
Overall	0.7925	0.1345	0.7358	0.1345	0.6013	0.1954	0.5518	0.1688
First	0.5683	0.2255	0.4838	0.2095	0.5547	0.2377	0.4748	0.2226
Second	0.9036	0.1383	0.8633	0.1441	0.6300	0.2318	0.5893	0.2264

Note. Items marked with an asterisk were significantly different than chance (.5000).

Table 11. Combined Accuracy by Condition by WM Span for Experiment 2.

WM Span	Overall		First		Second	
	Mean	SD	Mean	SD	Mean	SD
			Blocked			
High	0.7925	0.1345	0.5683	0.2255	0.9036	0.1383
Low	0.7358	0.1345	0.4838	0.2095	0.8633	0.1441
Total	0.7642	0.1371	0.5261	0.2212	0.8835	0.1423
			Mixed			
High	0.6013	0.1954	0.5547	0.2377	0.6300	0.2318
Low	0.5518	0.1688	0.4748	0.2226	0.5893	0.2264
Total	0.5765	0.1838	0.5148	0.2332	0.6096	0.2295
			Total			
High	0.6969	0.1928	0.5615	0.2312	0.7668	0.2346
Low	0.6438	0.1780	0.4793	0.2157	0.7263	0.2339
Total	0.6704	0.1872	0.5204	0.2271	0.7466	0.2349

Note. Chance = .500.

Table 12. Latency Measures by Condition by WM Span for Experiment 2.

	Blocked				Mixed			
	High WM Span		Low WM Span		High WM Span		Low WM Span	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
			Base-Learning					
Overall	8.3813	0.1736	8.4020	0.2200	8.3535	0.2073	8.4722	0.2661
First	8.5653	0.2674	8.5339	0.2821	8.4701	0.2644	8.6688	0.3432
Second	8.2937	0.2024	8.3362	0.2143	8.2926	0.2232	8.3776	0.2484
			Base-Generalization					
Overall	8.3703	0.1860	8.3982	0.2409	8.3861	0.2558	8.5471	0.2977
First	8.5336	0.2589	8.4833	0.3884	8.4524	0.3579	8.7084	0.3757
Second	8.2640	0.2052	8.3532	0.2221	8.3288	0.2831	8.4514	0.3024

Table 12 Continued.

	Blocked				Mixed			
	High WM Span		Low WM Span		High WM Span		Low WM Span	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Transfer-Learning								
Overall	8.3423	0.1901	8.3147	0.1911	8.2573	0.2360	8.4065	0.2444
First	8.5217	0.2576	8.4279	0.2614	8.3882	0.3200	8.5223	0.3192
Second	8.2382	0.2336	8.2626	0.1924	8.1841	0.2411	8.3489	0.2427
Transfer-Generalization								
Overall	8.2729	0.2348	8.2627	0.2409	8.2282	0.2491	8.3270	0.2838
First	8.4739	0.3373	8.3899	0.3021	8.3435	0.3653	8.4902	0.3306
Second	8.2059	0.2777	8.1817	0.2548	8.1833	0.2747	8.2323	0.3117

Table 13. Combined Latency by Condition by WM Span for Experiment 2

WM Span	Overall		First		Second	
	Mean	SD	Mean	SD	Mean	SD
			Blocked			
High	8.342	0.199	8.524	0.280	8.250	0.231
Low	8.344	0.229	8.459	0.313	8.283	0.229
Total	8.343	0.214	8.491	0.298	8.267	0.230
			Mixed			
High	8.306	0.243	8.414	0.329	8.247	0.261
Low	8.438	0.282	8.597	0.351	8.353	0.285
Total	8.372	0.271	8.505	0.351	8.300	0.278
			Total			
High	8.342	0.199	8.524	0.280	8.250	0.231
Low	8.344	0.229	8.459	0.313	8.283	0.229
Total	8.358	0.244	8.498	0.325	8.283	0.255

Table 14. Difficulty Rating Measures by Condition by WM Span for Experiment 2.

	Blocked				Mixed			
	High WM Span		Low WM Span		High WM Span		Low WM Span	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
			Base-Learning					
Overall	2.3472	1.1072	2.6956	1.0016	3.7083	1.2622	4.0949	1.4569
First	3.2449	1.2899	3.7484	1.5294	3.9175	1.3475	4.2501	1.2571
Second	1.9188	1.1781	2.1723	1.0510	3.5977	1.4655	4.0170	1.6502
			Base-Generalization					
Overall	2.2940	1.1377	2.6875	1.0093	3.9282	1.4364	4.1921	1.4234
First	3.1482	1.4320	3.8439	1.5570	4.1441	1.6062	4.4598	1.4038
Second	1.8302	1.1474	2.1386	1.1377	3.8049	1.5580	4.0426	1.5799

Table 14 Continued.

	Blocked				Mixed			
	High WM Span		Low WM Span		High WM Span		Low WM Span	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Transfer-Learning								
Overall	2.3553	0.9352	2.8588	1.1248	3.9942	1.6032	4.0787	1.5158
First	3.4913	1.1975	3.9363	1.4386	4.4106	1.5358	4.2900	1.3543
Second	1.7350	1.0068	2.4021	1.2403	3.8004	1.8028	3.9717	1.6423
Transfer-Generalization								
Overall	2.2083	0.9402	2.9468	1.2690	4.0116	1.6618	4.0764	1.6355
First	3.5302	1.1774	3.8930	1.4768	4.3114	1.6600	4.3426	1.4192
Second	1.6227	0.9690	2.2731	1.3019	3.8234	1.8307	3.9513	1.7692

Table 15. Combined Difficulty Ratings by Condition by WM Span for Experiment 2

WM Span	Overall		First		Second	
	Mean	SD	Mean	SD	Mean	SD
			Blocked			
High	2.301	1.021	3.354	1.271	1.777	1.070
Low	2.797	1.096	3.855	1.482	2.247	1.174
Total	2.549	1.086	3.605	1.400	2.012	1.145
			Mixed			
High	3.911	1.483	4.196	1.532	3.757	1.651
Low	4.111	1.490	4.336	1.343	3.996	1.639
Total	4.011	1.486	4.266	1.439	3.876	1.645
			Total			
High	3.106	1.505	3.775	1.466	2.767	1.706
Low	3.454	1.461	4.096	1.431	3.121	1.671
Total	3.280	1.492	3.935	1.456	2.944	1.696

Table 16. Searching and Awareness by Condition by Span for Experiment 2.

WM Span	No Search		Search		
	Not aware	Aware	No	Yes/Maybe	Total
			Blocked		
High	2	25	18	9	27
Low	2	25	13	14	27
Total	4	50	31	23	54
			Mixed		
High	11	16	16	11	27
Low	6	21	16	11	27
Total	17	37	32	22	54
			Total		
High	13	41	34	20	54
Low	8	46	29	25	54
Total	21	87	63	45	108

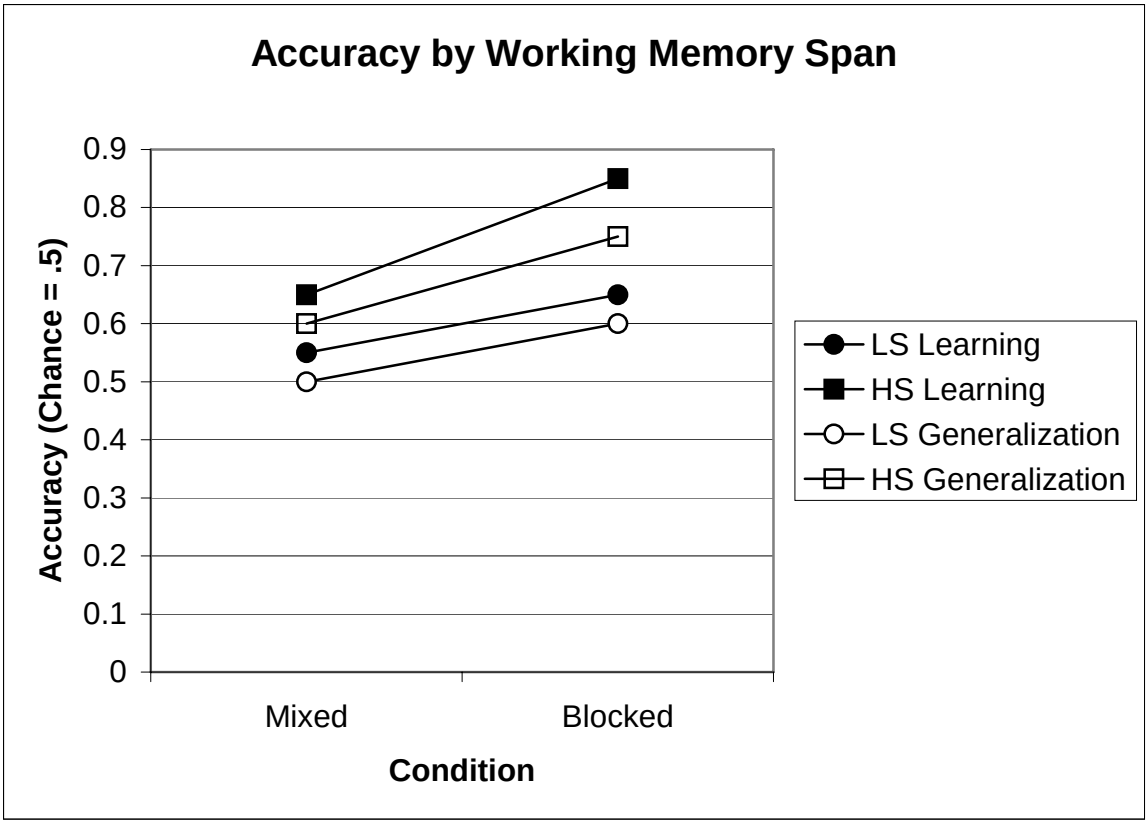

Figure 12 Predicted Accuracy Trends for Base and Transfer Items for High (HS) and Low (LS) Working Memory Span Participants in Experiment 2.

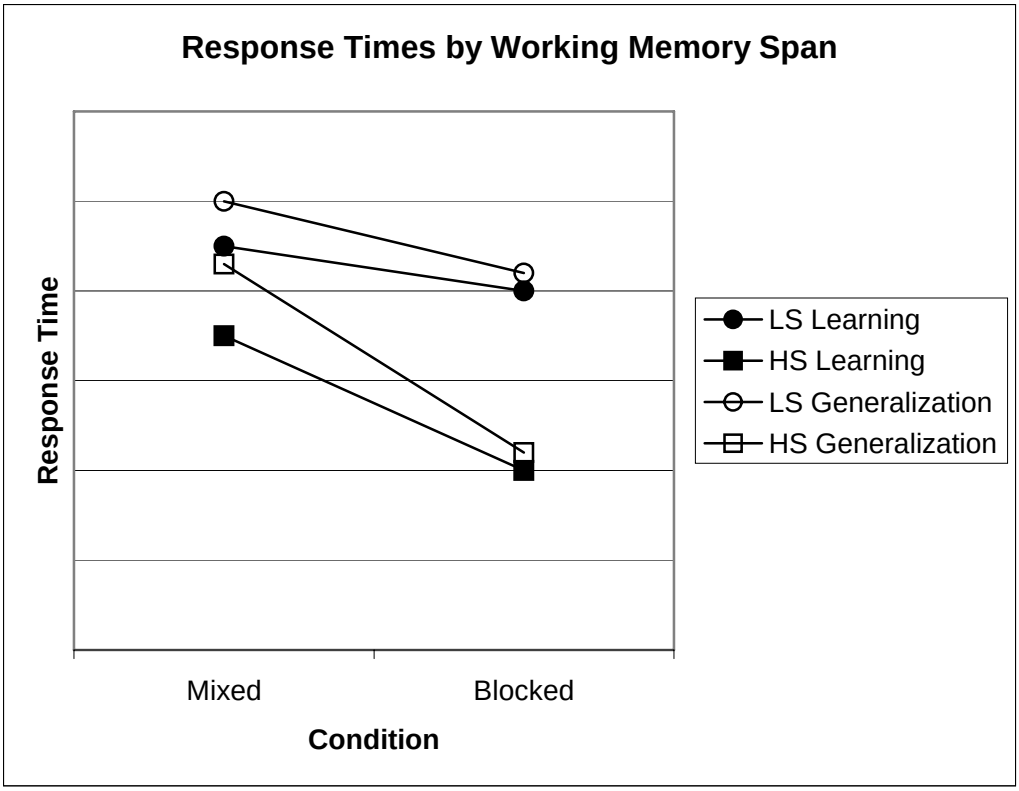

Figure 13. Predicted Latency Trends for Base and Transfer Items for High (HS) and Low (LS) Working Memory Span Participants in Experiment 2.

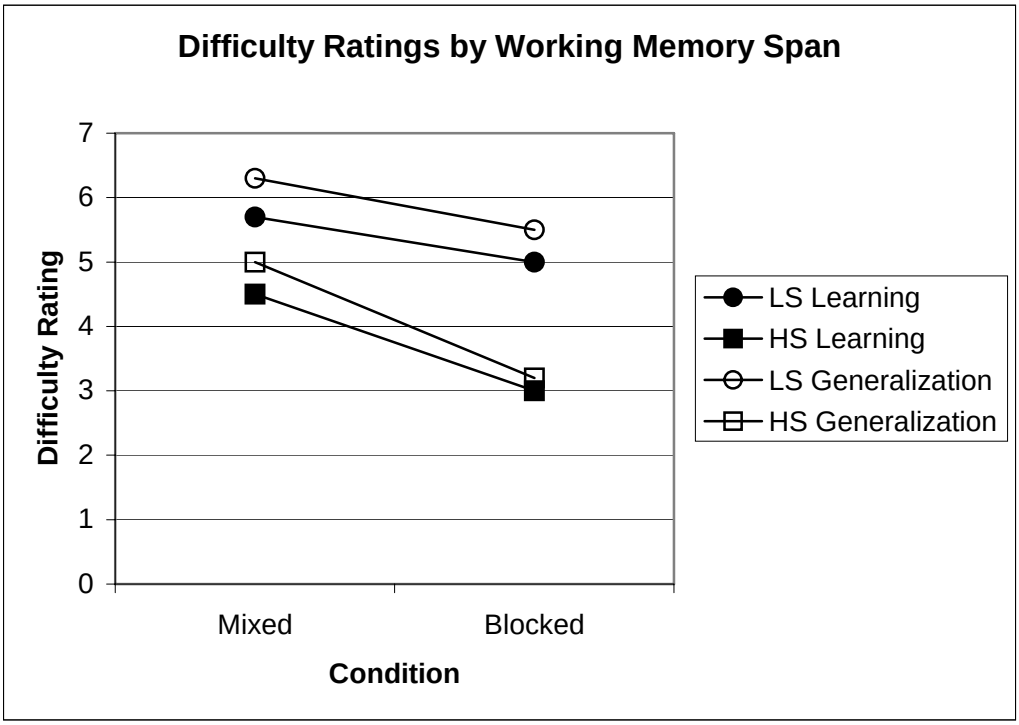

Figure 14. Predicted Difficulty Rating Trends for Base and Transfer Items for High (HS) and Low (LS) Working Memory Span Participants in Experiment 2.

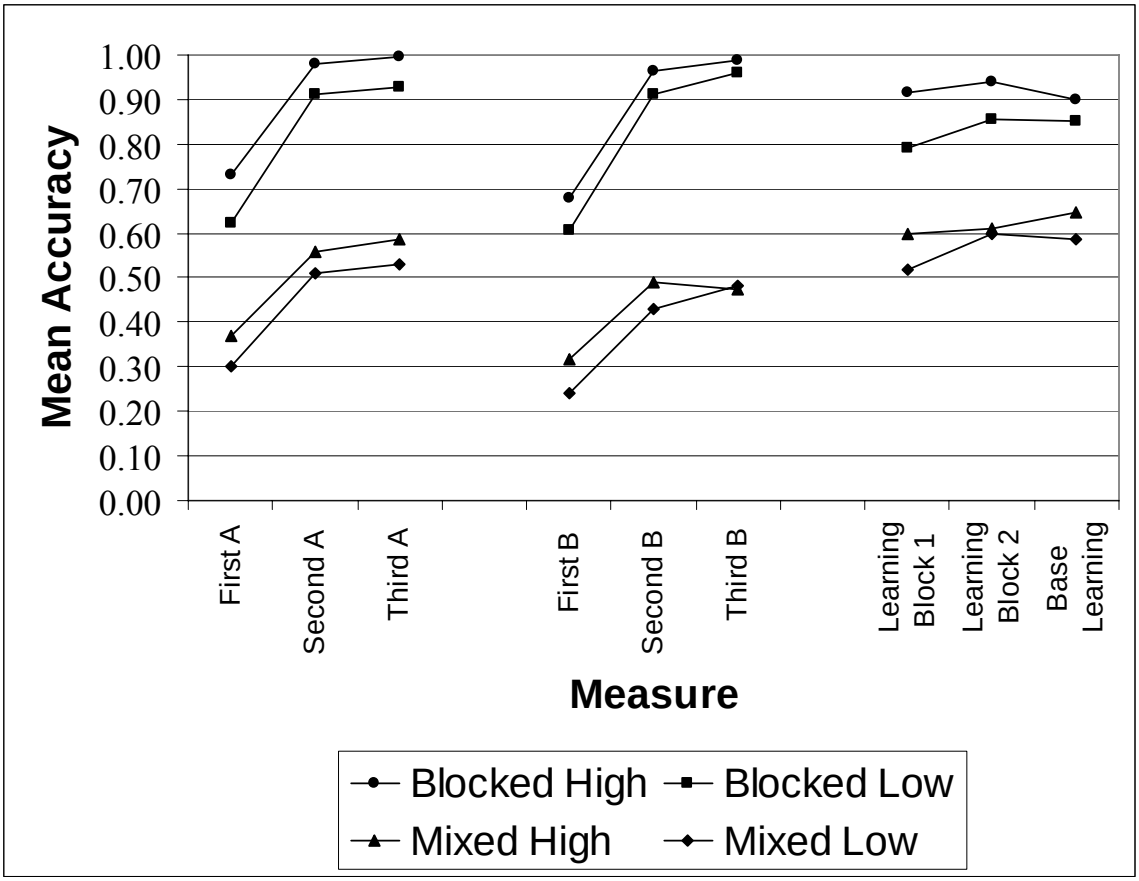

Figure 15. Overall Learning Accuracies for Base Items by Condition by WM span.

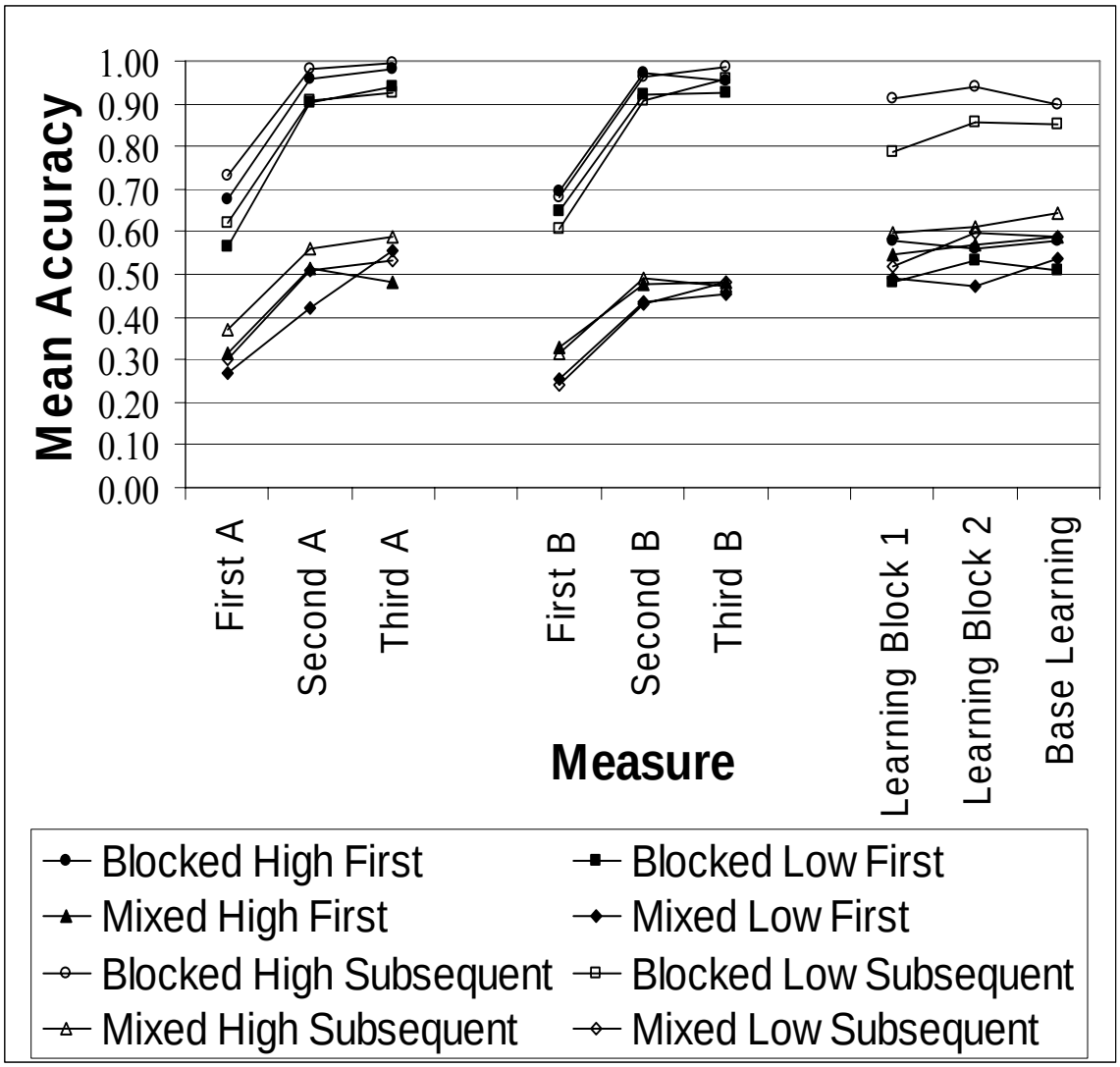

Figure 16. First and Subsequent Learning Accuracies for Base Items by Condition by WM span.

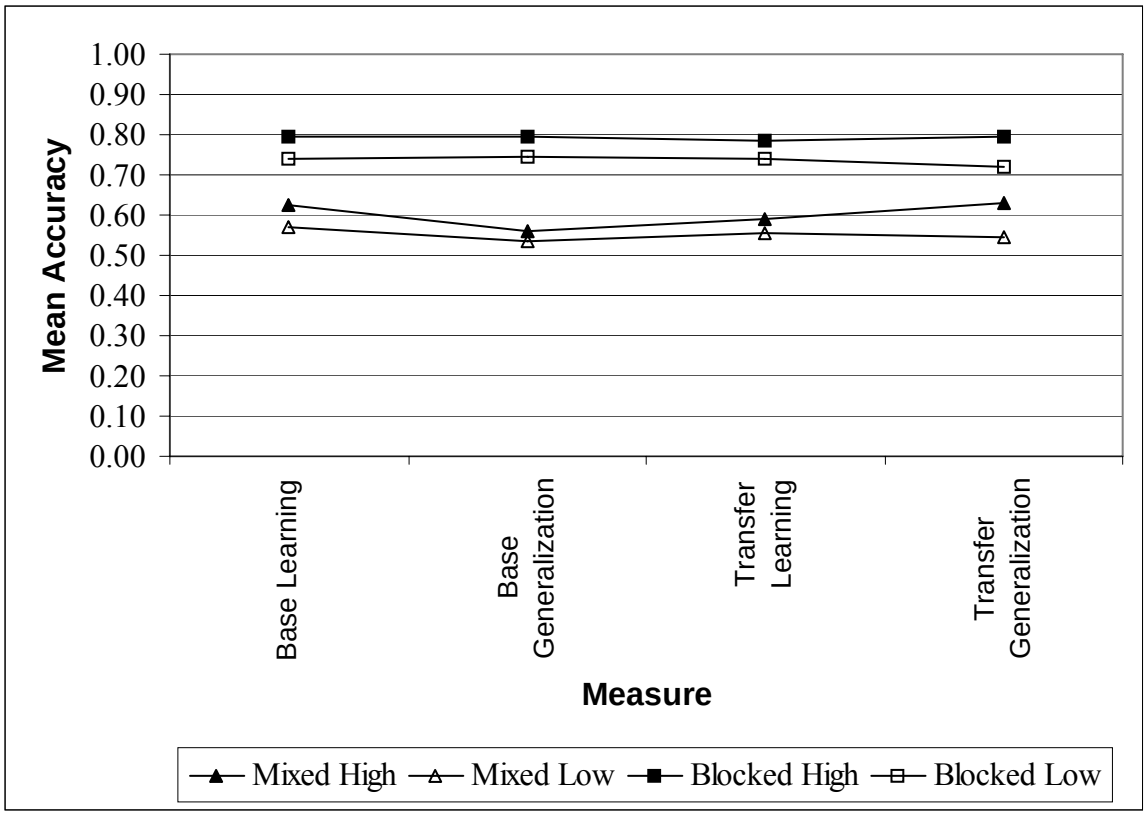

Figure 17. Overall Learning and Generalization Accuracies by Condition and WM span.

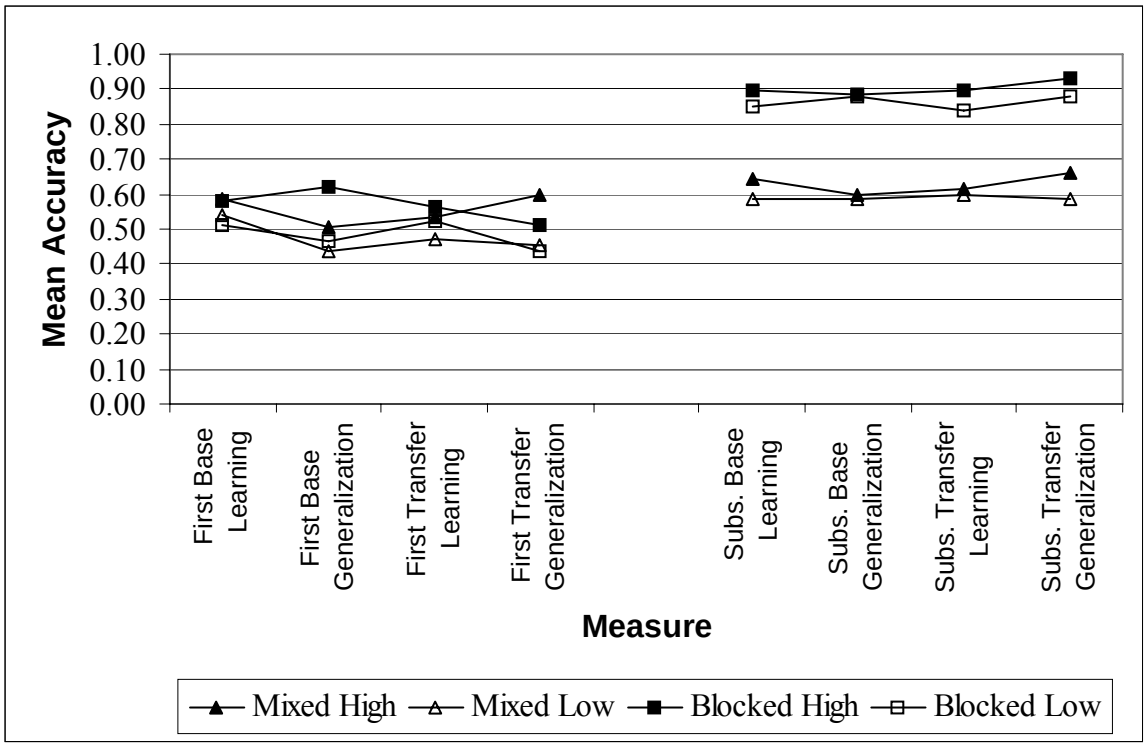

Figure 18. Learning and Generalization First and Subsequent Accuracies by Condition and WM span.

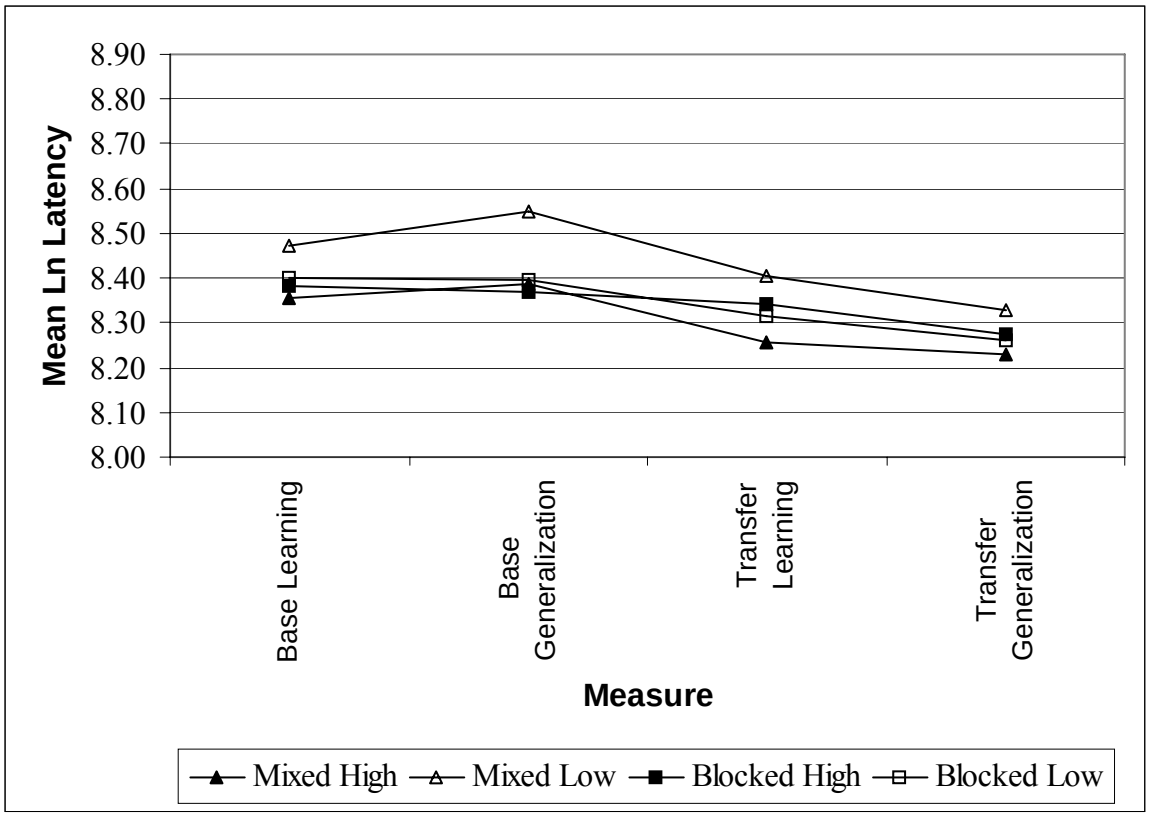

Figure 19. Overall Learning and Generalization Latencies by Condition and WM span.

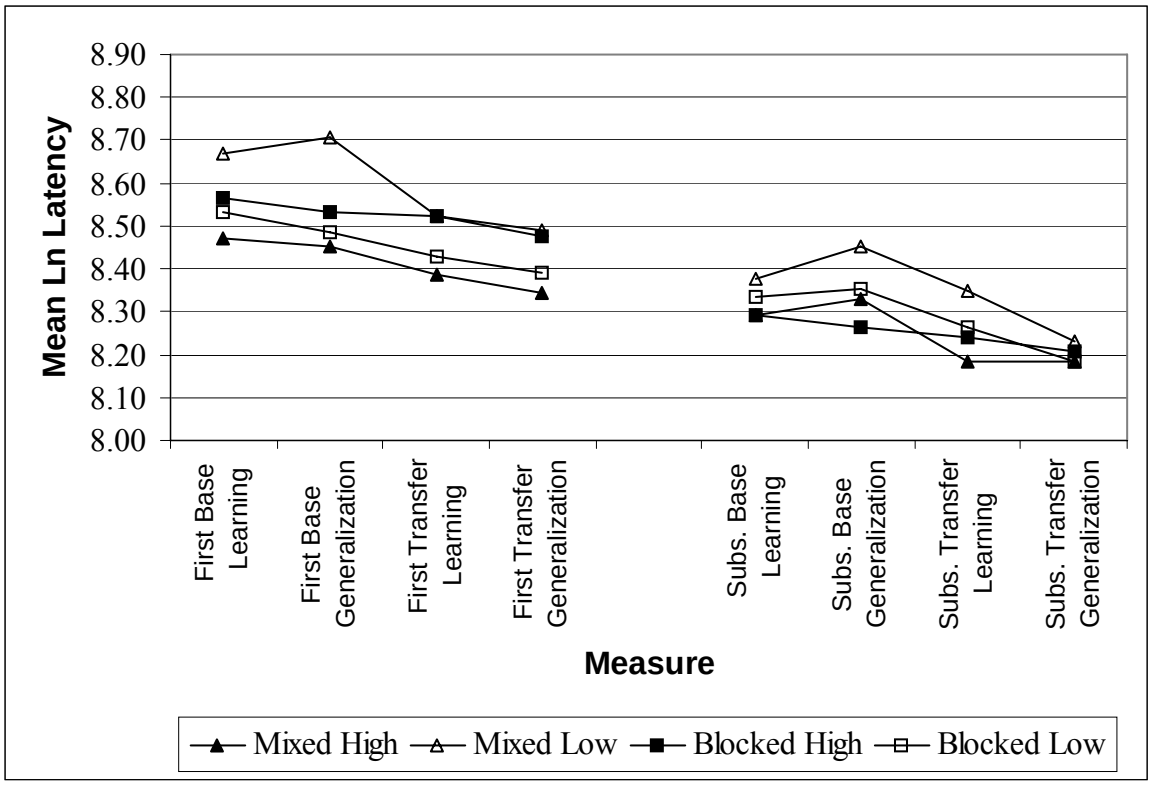

Figure 20. Learning and Generalization First and Subsequent Latencies by Condition and WM span.

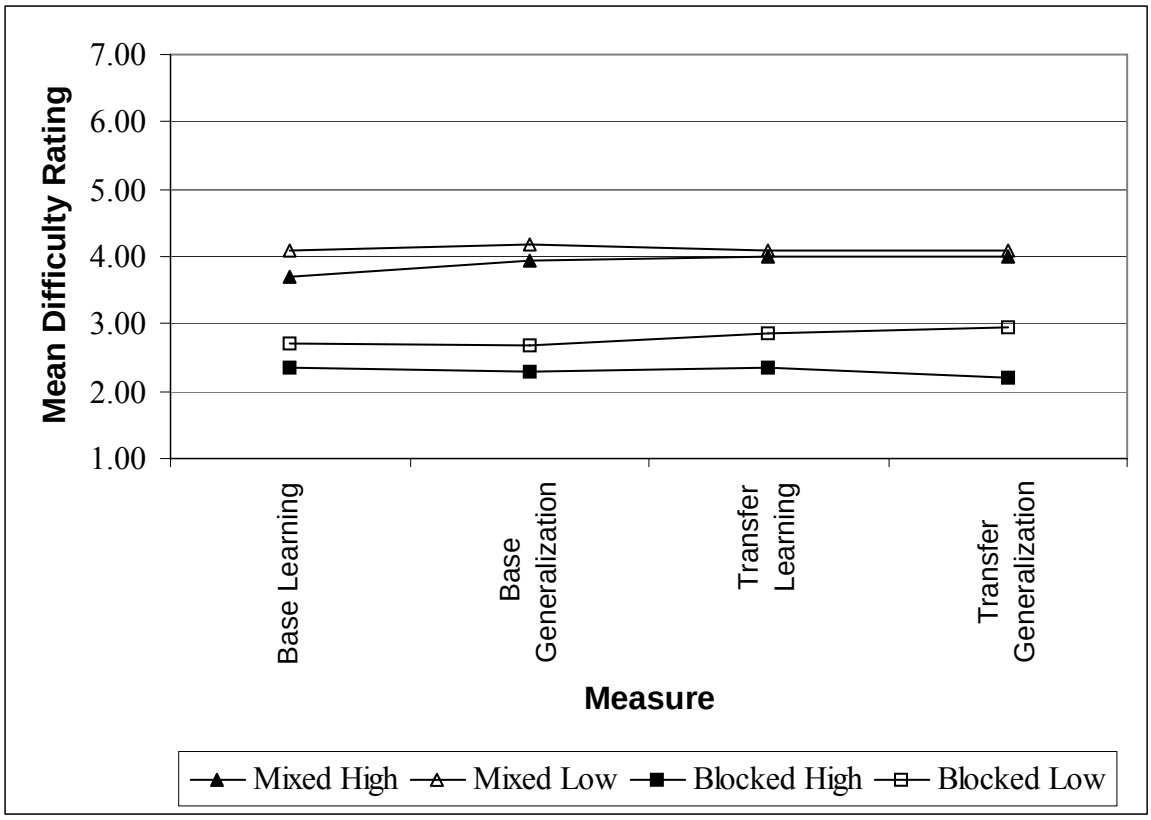

Figure 21. Overall Learning and Generalization Difficulty Ratings by Condition and WM span.

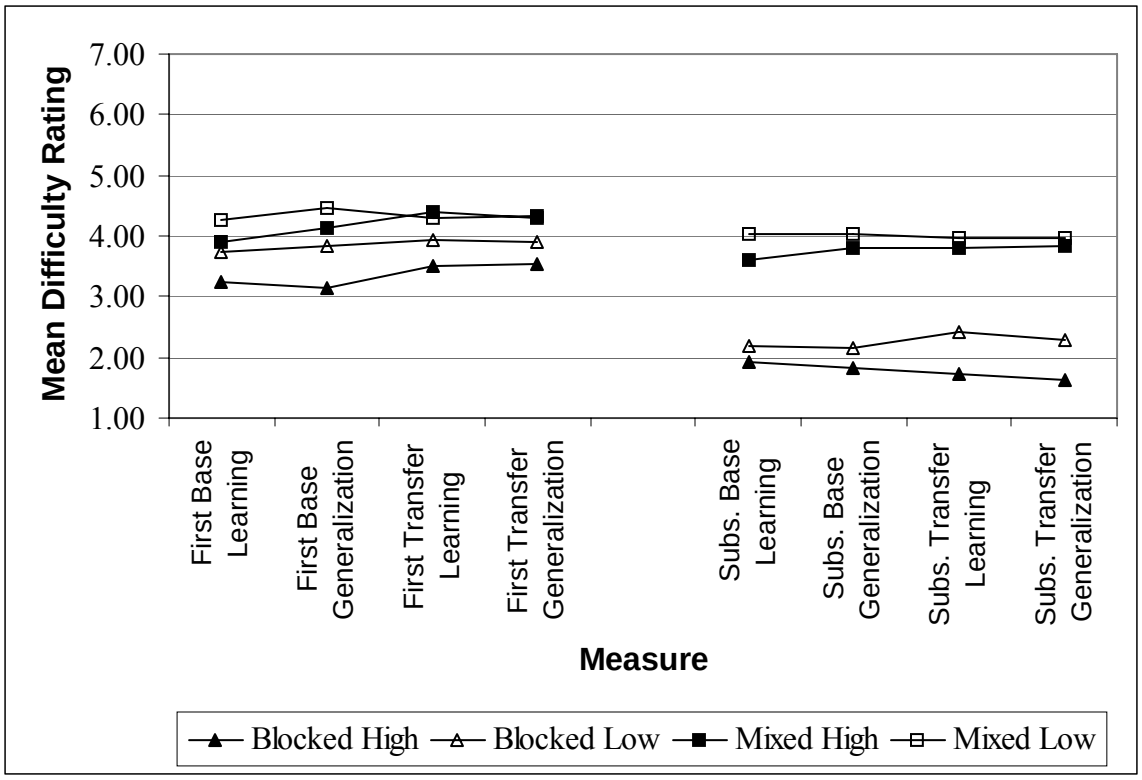

Figure 22. Learning and Generalization First and Subsequent Difficulty Ratings by Condition and WM span.

CHAPTER IV

GENERAL DISCUSSION

The results of Experiments 1 and 2 showed that the way in which identical information is presented to the learner has an effect on what is learned. The two experiments support the specific hypothesis that a blocked presentation format, here labeled a pattern-sequence presentation, leads to the acquisition of abstract category knowledge. Without a pattern-sequence presentation, participants were oblivious to the inherent structure in the learning materials (the mixed vs. control conditions in Experiment 1), and in both experiments applied a guessing strategy when required to learn without the structured presentation format implemented through the pattern-sequence format.

Abstract Category Knowledge Versus Item Memory

A major goal of these studies was to determine whether participants acquired abstract category knowledge or simply rote learning of item-feature associations through a pattern-sequence presentation. The strongest support for the existence of abstract category knowledge comes from the above chance performance on the base-generalization and transfer-generalization items. Participants in the blocked condition in both Experiment 1 and Experiment 2 performed at above-chance levels on these two measures, demonstrating that the pattern-sequence effect enabled participants to organize their knowledge of the categories to answer unseen item-feature pairs effectively. Furthermore, participants did not rate the generalization items as any more difficult than

the learning items. Even when the same structure was present, participants in Experiment 1 and low WM span participants in Experiment 2 were unable to learn the category information in the mixed condition sufficiently well to exceed chance performance on the generalization items.

There was other support for abstract category learning. When participants learned in the blocked condition, they were more accurate at associating the correct features to plants. Interestingly, this was true for subsequent-item accuracy, but not first-item accuracy. These results indicate that participants used the feedback from the first item to determine which category the plant belonged to, then applied their category knowledge to answer the following two questions about that plant.

A Role for WM in Category Learning

Two main theories of unsupervised category learning, SUSTAIN (Gureckis & Love, 2003b; Love et al., 2004) and Clapper and Bower (2002) have appealed to WM as a possible mechanism for the acquisition of category knowledge, but none have empirically demonstrated a role for WM in healthy adults. Experiment 2 provides empirical support for the role of WM in the acquisition of abstract category knowledge, but failed to support the specific predictions based on the model used by Clapper and Bower (2002). Even though the main experimental hypotheses were not supported, there were several pieces of converging evidence that supports a role for WM in the discovery and acquisition of abstract category information.

Evidence

The strongest evidence indicating a role for WM comes from the superior performance of the high WM span participants on the base-generalization and transfer-generalization items in Experiment 2. High WM span participants performed better than low WM span participants on the first items for these generalization trials, showing that high span participants were better able to connect feature information to Latin plant names. This was not due to a speed-accuracy trade-off because not only did high span participants have higher accuracy on the base-generalization items, they also responded faster on these items than did the low WM span participants. Other evidence supported a role for WM in the learning data as well, specifically, high WM span participants had higher accuracy than did the low WM span participants when learning plant features.

Implications for Theories of Unsupervised Category Learning

The Implicit Theory

The findings in favor of WM playing a role in the acquisition of abstract category knowledge runs counter to the implicit learning hypothesis that category knowledge and item memory are mediated by different physical and cognitive systems (Ashby et al., 1998, Knowlton & Squire, 1993). Knowlton and Squire's results with amnesic and normal controls on a dot distortion task found no differences in prototype classification, a measure of abstract category knowledge, between the two groups, but the amnesics performed much more poorly on a new/old discrimination task, a measure of item-feature memory. When compared to the present study, the amnesics would have lower WM span

than the low WM span participants used in Experiment 2. The differences between high and low WM span participants in Experiment 2 exhibited an analogous item-feature memory effect; high WM span participants achieved higher accuracy on the first base-learning items which, arguably, depends on item memory than did the low WM span participants.

The implicit memory hypothesis would predict no differences between high and low WM span participants on the generalization items as these items would not enjoy the item-memory benefit conferred by greater WM capacity. This was not the case. The high WM span participants had significantly higher generalization accuracies than the low WM span participants, which suggests that either the category learning system is not isolated from WM, or there is just a single system that processes item and category information. Indeed, this is one of the central debates in the study of category learning and is explored in greater detail in the next section. Alternatively, proponents of implicit category acquisition could argue that the uniform benefits of WM across learning and testing conditions in the present study argues for a general-resource interpretation for the role of WM or a less rigid separation of the implicit and explicit systems.

Hypothesis Testing versus Discovery

Several researchers (Ashby et al., 1998; Ashby, Queller, & Berretty, 1999; Ashby & Speiring, 2004; Ashby & Waldron, 1999, 2000; Maddox & Ashby, 2004; Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005) propose that there are two different systems responsible for category learning; an intentional, rule-based system that relies heavily on

WM resources and an unintentional, similarity-based system that operates via implicit learning principles and is similar to the category learning system proposed by Knowlton and Squire (1993). The rule based system relies on WM resources to actively search the environment for relationships amongst features, the more complex the relationships or category definitions, the greater demand placed on WM. The implicit system does not rely on WM and is largely intact in amnesics and those with hippocampal and frontal impairments (Ashby et al., 1998; Knowlton & Squire, 1993; Reber, 1967, 1989). This system builds category representations based on perceptual similarity or by associating regions of feature space with a particular response (Ashby et al., 1998; Ashby & Gott, 1988; Maddox & Ashby, 1993). The results of the present study do not preclude the existence of two separate systems, nor do they preclude the simultaneous contributions of implicit and intentional rule-learning systems. The results do create difficulties for the rigid separation of the two systems. According to the two-system model, category learning either proceeds via an explicit hypothesis testing process, or via a perceptual or precategorical association mechanism (Ashby et al., 1998; Ashby & Gott, 1988; Knowlton & Squire, 1993; Maddox & Ashby, 1993). There is no provision in these theories for communication between the two systems that might be capable of explaining how WM would mediate the unintentional discovery and emergence to consciousness of category information.

The main findings in the present experiments showing a role for WM in the acquisition of abstract category information cannot differentiate between the category invention hypothesis and the two-system hypothesis that explicitly learns category rules

and exemplar-category associations. The lack of a difference between the learning and generalization items coupled with the main effect of WM span supports the two-system hypothesis. The category invention hypothesis predicts an interaction between WM span and the ordering manipulation. These results, represented in the present work as Hypotheses 1-5 in Experiment 2, were not supported. There are a few findings that support the theoretical specification of the category invention effect, that there is a role for WM in the discovery of category information.

The two-system model predicts that the high WM span participants are more successful at learning the structure once it is found and then learning which exemplars belong to which category (exemplar-category information). According to this position, high WM span participants who look for patterns should demonstrate greater knowledge of the category structure because hypothesis testing is an explicit process. When the high and low WM span participants who reported actively searching for patterns on their post-experimental questionnaires were compared, however, the high WM span participants were no more aware of the category structure than were low WM span participants. This result is consistent with the category-invention process, which does not require that the learner actively search for structure.

The two-system model predicts greater learning for high WM span participants who actively searched for patterns because the WM system is best at searching for and finding rule-based categories like the ones used in the present study (Ashby et al., 1998, Ashby et al., 1999). This was indeed the case; greater WM resources improved learning in both conditions. However, the two-system model also predicts no differences between

the high and low WM span participants who did not search for patterns in either awareness or performance as WM benefits would be accompanied only by explicit intention and awareness. This prediction was not supported. High WM span participants who did not actively search for patterns and indicated some or full awareness of the category structure on their post-experimental questionnaires performed better than their low WM span counterparts, suggesting that WM acts to improve learning without intention, a result consistent only with the category invention hypothesis. These results support a single category learning system or a less rigid separation of the hypothesis testing and implicit learning systems.

Implications for Theories of Working Memory

The purpose of this research was to determine if there is a role for WM in the discovery and acquisition of category information in unsupervised category learning. The many theories of WM attribute observed effects to various aspects and functions of a WM system. The present study was not set up to differentiate among these many theories, but it does illustrate a few key findings relevant to theories of WM. The connection between WM and the discovery of structure or regularities in the environment creates problems for theories that propose that WM is a strictly explicit system, such as Engle's (2002) controlled attention model. These models propose that the central aspect of WM relevant to the present findings is the control of attention. The present study demonstrated no differences between high and low WM span participants in the amount of explicit search for patterns, the process that should demonstrate differences due to WM

capacity and the resultant category knowledge, at least according to this type of theory.

An alternative interpretation of the findings in Experiment 2 suggests a more general role for WM in general cognitive processes than is currently captured by theories of WM. The present results are consistent with the hypothesis that WM acts as a general resource for processing. According to this view, WM capacity determines what we are capable of encoding and retrieving. Furthermore, this function of WM is not tied to traditional WM mechanisms such as the control of attention or the manipulation of the contents of memory. It describes a continuous, unintentional, and critical role for WM in the acquisition of category information. Theories of WM that propose general capacity differences (e.g., Cowan, 1995) without limiting WM differences to differences in explicit processes are supported by the current findings. Cowan's (1995) model is an embedded process model that describes WM as the active part of long-term memory. In the present data, WM supported learning, but surprisingly, did not interact with the category invention effect.

Methodological Limitations and Future Directions

The results of Experiments 1 and 2 support the methodology used here as being capable of distinguishing item memory from abstract category knowledge. The accuracy measures were the most sensitive indices of category learning and should be used in future studies. The difficulty measures provided an alternate measure of awareness, but did not provide as much detail as the accuracy measures. The latency measures were not sensitive to accuracy differences, though they were valuable in eliminating a speed-

accuracy trade-off alternative hypothesis. The current method was not able to evaluate the long-term effects of the rapid learning due to the pattern-sequence effect. It is possible that, as the category invention effect posits a strong role for WM in the rapid acquisition of the category structures, that those effects would be relatively short-lived.

The evidence in support of WM involvement in acquiring abstract category level information was weaker than expected. Low power and small effect sizes for the WM factor is likely a part of the reason for the weaker than expected support for the experimental hypotheses. The initial power analyses were based on a medium effect size for the WM effect. The values of η^2 found for the WM effects in Experiment 2 indicate that WM has a small effect size in this paradigm, especially when compared to the large effect size of the blocking manipulation used to create the blocked and mixed conditions. Additional participants would likely convert the marginally significant WM awareness effect into a significant effect. The other option is to make the blocking manipulation smaller in magnitude and allow a greater role for WM by making the task more demanding on WM resources. The ceiling effect found in the blocked condition indicates that the task used in Experiment 1 and 2 was likely too easy to require substantial WM resources.

The present experiment hypothesized that WM is required for the discovery and acquisition of category structure. The slim support for a WM effect in the present study does not necessarily mean that WM is not a critical component in the discovery of category structure. It is possible that the present findings accurately capture the contribution of WM to unsupervised category learning, reflecting a small role for WM in

the discovery and acquisition of category knowledge. It is also possible that WM is critically important, but there exists a limited range of task difficulty within which human WM differences manifest. If the task difficulty exceeds WM resources there will be no category learning. If the task difficulty is within WM resources, then there will be category learning. In the context of the present study, the mixed condition imposes WM requirements that exceed the capacity of even the high WM span participants, whereas the blocked condition required scant WM resources possessed by even by the low WM span participants. Clear support of this hypothesis depends on careful calibration of the difficulty of the blocking manipulation.

There are several ways to increase the WM span resources required to discover and acquire the category information. The main way of requiring greater WM resources is to make the task more difficult. Clapper (2006a) found that using three categories reduced performance relative to two categories. This manipulation would also require greater WM resources in the prevention of interference; keeping the categories separated. Move the blocking manipulation to the middle of the experiment. Clapper (2006b) found that more blocked trials were required to acquire the category structure when the blocked trials were placed in the middle of the experiment than when they were placed at the beginning of the experiment. Using a dual task paradigm or imposing a memory load or a secondary task uses WM resources and reduces the amount available for the category learning task. Lower performance under dual-task conditions relative to single-task conditions would support a role for WM in the discovery of category structure.

Future studies should also investigate more deeply the difference between the

WM-aids-discovery theory and the WM-aids-rule-learning theory. This issue is less tractable than simply establishing that WM plays a role in the acquisition of category-level knowledge in unsupervised category learning. One way to do this would be to compare the performance of participants who were instructed to discover and use the category structure to learn the plants and their features against the performance of participants who were not instructed to look for category structure. When coupled with a sufficiently difficult and taxing (on WM) paradigm, differences between these two conditions on performance and awareness could reveal the relative contribution of WM to discovery of rules versus the learning of those rules.

Conclusions

The present work had two main aims: to investigate whether the pattern-sequence effect was due to item memory or abstract category knowledge, and whether greater WM resources amplify the discovery and acquisition of category knowledge when using a pattern-sequence presentation format, as suggested by Clapper and Bower's (2002) simulation model of category acquisition. Experiments 1 and 2 supported the hypothesis that the pattern-sequence presentation format resulted in the acquisition of abstract category knowledge, not in enhanced item-feature memory for the viewed items. The blocking manipulation resulted in the pattern-sequence effect for both the items seen during training and for the generalization items that were not seen during learning, a result that can only be explained by the presence of category knowledge independent of rote item-feature encoding and retrieval.

WM was associated with greater generalization performance, the primary indicator of category knowledge used in this study. Its primary benefit was due to enhanced performance on the first-items, and was not a significant factor for the subsequent-items. Therefore, strong support for the hypothesized WM involvement in the pattern-sequence effect was not found. However, the presence of a ceiling effect in generalization accuracies in the blocked condition represents a serious calibration confound that does not permit a clean rejection of the experimental hypothesis. It would seem that there may be a narrow window between a task that is so easy that everyone gets it and a task that no one gets it in which to find differences in WM span. The blocked condition appears to belong to the former, the mixed condition to the latter. Because of a possible ceiling effect, the present findings neither support nor discredit the alternative hypothesis, that item knowledge and category knowledge are mediated in separate encapsulated systems.

There was no evidence for WM involvement in the pattern-sequence effect as described by the category invention effect (Clapper & Bower, 2002). High WM span participants acquired the category structure just as well as the low WM span participants, which is counter to the category invention hypothesis of Clapper and Bower (2002). High WM span participants were no more likely to look for patterns than were low WM span participants, nor were they more successful at acquiring the category information once they had found it. The view that WM either results in improved item memory or is involved in explicit hypothesis testing would predict greater active search, and greater mastery of the information once found. Neither of these outcomes was observed. The

behavioral data also support a role for WM. When the analysis was repeated on those participants who did not report searching for patterns, there were no qualitative differences with those who did. The present study provides evidence that WM is involved in the discovery and acquisition of abstract category knowledge.

Insight into the process of category learning in the present studies resulted from an ad hoc analysis of first and subsequent trials associated with plant names. This analysis was not directly motivated by any of the theories reviewed for this research. The analyses consistently suggested a strategy applied by both high- and low- WM span learners whereby they used feedback from the first trial to respond nearly flawlessly on subsequent trials. This suggests a high level of metacognitive control and signals an effective adaptation by learners to the learning situation. The present findings also confirm that the structure of the learning situation determines whether the person will acquire this strategy.

REFERENCES

Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., & Carter, C. S. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience, 12,* 997-987.

Allen, R., & Reber, A. S. (1998). Unconscious intelligence. In W. Bechtel & G. Graham (Eds.), *A Companion to Cognitive Science* (314-323). Malden, MA: Blackwell Publishers.

Anderson, A. L., Ross, B. H., & Chin-Parker, S. (2002). A further investigation of category learning by inference. *Memory & Cognition, 30,* 119-128.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409-429.

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review*, *8*, 629-647.

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 259-277.

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

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

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

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

Ashby, G. F., O'Brien, J. B. (2005). Category learning and multiple memory systems. Trends in Cognitive Sciences, 9, 83-89.

Ashby, F. G., Queller, S., & Berretty, P. T. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61,* 1178-1199.

Ashby, F. G., & Spiering, B. J. (2004). The neurobiology of category learning. *Behavioral and Cognitive Neuroscience Reviews, 3,* 101-113.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review, 6,* 363-378.

Ashby, F. G., & Waldron, E. M. (2000). The neuropsychological bases of category learning. *Current Directions in Psychological Science, 9,* 10-14.

Baddeley, A. D. (1986). *Working memory.* New York: Oxford University Press.

Baddeley, A. & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in Research and Theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Berg, E. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15-22.

Billman, D., & Heit, E. (1988). Observational learning from internal feedback: A simulation of an adaptive learning method. *Cognitive Science, 12,* 587-625.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 458-475.

Bleckley, M. K., Durso, F. T., Crutchfield, J. M., Engle, R. W., & Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review, 10,* 884-889.

Bolte, A., & Goschke, T. (2004). Implicit learning of category sequences. Poster presented at the 45th Annual Meeting of the Psychonomic Society, Minneapolis.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics, 27,* 185-209.

Bunting, M. F., & Cowan, N. (2005). Working memory and flexibility in awareness and attention. *Psychological Research, 69,* 412-419.

Čech, C. G., & Shoben, E. J. (2001). Categorization processes in mental comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 800-816.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition, 30,* 353-362.

Clapper, J. P. (2006a). Category load, structural alignability, and category learning. Poster presented at the 47th Annual Meeting of the Psychonomic Society, Houston.

Clapper, J. P. (2006b). When more is less: Negative exposure effects in unsupervised learning. *Memory & Cognition, 34,* 890-902.

Clapper, J. P., & Bower, G. H. (1994). Category invention in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 443-460.

Clapper, J. P., & Bower, G. H. (2002). Adaptive categorization in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 908-923.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Conover, W. J. (1999). Practical nonparametric statistics (3rd edition). New York: John Wiley & Sons.

Conway, A., Cowan, N., Bunting, M., Therriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163-184.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354-373.

Coombes, A. J. (1985). *Dictionary of Plant Names.* Portland, OR: Timber Press.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 397-417.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19-23.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Human Learning and Memory, 18,* 972-992.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84,* 253-262.

Engle, R. W., & Kane, M. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation, 44,* 145-199.

Engle, R. W., Kane, M., & Tuholski, S. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake and P. Shah (eds.), *Models of Working Memory* (pp.102-134). Cambridge, UK: Cambridge University Press.

Engle, R. W., Tuholski, S., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309-331.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars n category learning. *Journal of Experimental Psychology: General, 127,* 107-140.

Estes, W. K. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115,* 155-174.

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analysis for MS_DOS [Computer software]. Bonn, FRG: Bonn University, Dep. Of Psychology.

Filoteo, J. V., Maddox, W. T., Ing, A. D., Zizak, V., & Song, D. D. (2005). The impact of irrelevant dimensional variation on rule-based category learning in patients with Parkinson's disease. *Journal of the international Neuropsychological Society, 11,* 503-513.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234-257.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227-247.

Gluck, M. A., & Myers, C. E. (2001). Gateway to memory: An introduction to neural network modeling of the hippocampus and learning. Cambridge, MA: MIT Press.

Goshen-Gottstein, Y., & Kempinsky, H. (2001). Probing memory with conceptual cues at multiple retention intervals: A comparison of forgetting rates on implicit and explicit tests. *Psychonomic Bulletin & Review*, *8*, 139-146.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1998). Abstract and effector-specific representations of motor sequences identified with PET. *The Journal of Neuroscience, 18,* 9420-9428.

Gureckis, T. M., & Love, B. C. (2003a). Human unsupervised and supervised learning as a quantitative distinction. *International Journal of Pattern Recognition, 17,* 885-901.

Gureckis, T. M., & Love, B. C. (2003b). Towards a unified account of supervised and unsupervised category learning. *Journal of Experimental & Theoretical Artificial Intelligence, 15,* 1-24.

Gureckis, T. M., & Love, B. C. (2004). Common mechanisms in infant and adult category learning. *Infancy, 5,* 173-198.

Hambrick, D. Z., & Engle, R. W. (2003). The role of working memory in problem solving. In *The psychology of problem solving* (Eds. Davidson, J. E. & Sternberg, R. J.), New York: Cambridge University Press.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation, (vol. 22),* 193-225.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher, & A. Koriat (Eds.*), Attention and Performance XVII Cognitive Regulation of Performance: Interaction of Theory and Application* (pp. 653-675). Cambridge, Massachusetts: MIT Press.

Hoffman, A. B., & Murphy, G. L. (2006). Category dimensionality and feature knowledge: When more features are learned as easily as fewer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 301-315.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representation of structure: A theory of analogical access and mapping. *Psychological Review, 104,* 427-466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220-264.

Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology, 69,* 596-608.

Jha, A., Fabian, S., & Aguirre, G. (2004). The role of prefrontal cortex in resolving distractor interference. *Cognitive, Affective & Behavioral Neuroscience*, *4*, 517-527.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology, 54A,* 1105-1124.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169-183.

Kaplan, A. S., & Murphy, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition, 27,* 699-712.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56,* 263-269.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107,* 397-402.

Kareev, Y, Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126,* 278-287.

Kelly, S. E., Burton, A. M., Kato, T., & Akamatsu, S. (2001). Incidental learning of real-world regularities. *Psychological Science, 12,* 86-89.

Kersten, A. W., & Billman, D. (1997). Event category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 638-658.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262,* 1747-1749.

Knowlton, B. K., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 169-181.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic category learning in amnesia. *Learning and Memory,19,* 106-120.

Kruschke, J. K. (1992). ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. *Psychological Review, 99,* 22-44.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1083-1119.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389-433.

Li, K. Z. H. (1999). Selection from working memory: On the relationship between processing and storage components. *Aging, Neuropsychology, and Cognition, 6,* 99-116.

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

Love, B. C., & Gureckis, T. M. (2004). The Hippocampus: Where a Cognitive Model meets Cognitive Neuroscience. *Proceedings of the Cognitive Science Society*.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309-332.

Love, B. C. (2005). Environment and goals jointly direct category acquisition. Current Directions in Psychological Science, 14, 195-199.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53,* 49-70.

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

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 650-662.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 100-107.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129,* 592-613.

McClelland, J. L. (1981). Retrieving general and specific information from stored knowledge of specifics. *Proceedings of the Third Annual Meeting of the Cognitive Science Society,* 170-172.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language, 38,* 225-252.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

Medin, D. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between Item and Category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 607-625.

Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189-223). Hillsdale, NJ: Erlbaum.

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

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in category learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355-368.

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, and Cognition, 27,* 775-799.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Models of active maintenance and executive control*. Cambridge, UK: Cambridge University Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1-32.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54-65.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282-304.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22,* 352-369.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomenon in perceptual categorization. *Psychonomic Bulletin & Review, 9,* 160-168.

Nosofsky, R. M., Kruschke, R. M., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 211-233.

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

Pavlik, P., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science, 29,* 559-586.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119,* 264-275.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353-363.

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

Price, A. L. (2005). Cortico-striatal contributions to category learning: Dissociating the verbal and implicit systems. *Behavioral Neuroscience, 119,* 1438-1447.

Price, A. L. (2006). Explicit category learning in Parkinson's Disease: Deficits related to impaired rule generation and selection processes. *Neuropsychology, 20,* 249-257.

Raven, J. C., Court, J. H., & Raven, J. (1977). *Standard progressive matrices.* London: H. K. Lewis.

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

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219-235.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382-407.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. Journal of Experimental Psychology: General, 126, 211-227.

Ross, B. H., & Makin, V. S. (1999). Prototype versus exemplar models in cognition. In R. J. Sternberg (Ed.), *The Nature of Cognition* (205-241). Cambridge, Massachusetts: The MIT Press.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General, 133,* 534-553.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 681-696.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*, 4-27.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65,* 167-196.

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 800-811.

Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1251-1263.

Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics (4th ed.). Boston: Allyn and Bacon.

Taraban, R. (2006). High-frequency phrases and word endings aid gender-like category induction. [manuscript submitted]

Taraban, R. (2004). Drawing learners' attention to syntactic context aids gender-like category induction. *Journal of Memory and Language, 51,* 202-216.

Taraban, R., & Hayes, M. (2000). Category induction for ordinary facts. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 936-941). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Taraban, R., & Hayes, M. (2001). Category induction for knowledge of facts. [unpublished manuscript]

Taraban, R., & Kempe, V. (1999). Gender processing in native and non-native Russian speakers. *Applied Psycholinguistics, 20,* 119-148.

Taraban, R., & Roark, B. (1996). Competition in learning language-based categories. *Applied Psycholinguistics, 17,* 125-148.

Toothaker, L. E. (1993). *Multiple comparison procedures.* Newbury Park, CA: Sage Publications.

Tracy, J. I., Mohamed, F., Faro, S., Pinus, A., Tiver, R., Harvan, J., Bloomer, C., Pyrros, A., & Madi, S. (2003). Differential brain responses when applying criterion attribute versus family resemblance rule learning. *Brain and Cognition, 51,* 276-286.

Tulving, E. (1991). Concepts of human memory. In L. R. Squire, N. M. Weinberger, G. Lynch, & J. L. McGough (Eds.), *Memory: Organization and locus of change* (3-32). New York: Oxford University Press.

Turner, M. & Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavioral Research Methods, 37,* 498-505.

Wagar, B., & Dixon, M. J. (2005). Past experience influences object representation in working memory. Brain & Cognition, 57, 248-256.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning. *Psychonomic Bulletin & Review, 8,* 168-176.

Wattenmaker, W. D. (1991). Learning modes, feature correlations, and memory-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 908-923.

Wattenmaker, W. D. (1992). Relational properties and memory-based category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282-304

Yamauchi, T., & Markman, A. B. (1998). Category-learning by inference and classification. *Journal of Memory and Language, 39,* 124-148.

Yamauchi, T., & Markman, A. B. (2000a). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 776-795.

Yamauchi, T., & Markman, A. B. (2000b). Learning categories composed of varying instances: The effect of classification, inference and structural alignment. *Memory & Cognition, 28,* 64-78.

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 585-593.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34,* 387-398.

APPENDIX A

EXPERIMENTAL STIMULI

Table 17 Structured Stimuli Used in the Blocked and Mixed Conditions.

Name	Root	Stem	Leaf	Flower
		Category A		
Learning Items				
Acorus	Taproot*	Smooth	Pointed	Headed
Cynara	Taproot*	Smooth	Pointed	Headed
Larix	Taproot	Smooth*	Pointed	Headed
Mentha	Taproot	Smooth*	Pointed	Headed
Nolana	Taproot	Smooth	Pointed*	Headed
Ruscus	Taproot	Smooth	Pointed*	Headed
Sedum	Taproot	Smooth	Pointed	Headed*
Vitis	Taproot	Smooth	Pointed	Headed*
Transfer Items				
Juglans	Taproot*	Smooth	Pointed	Headed
Punica	Taproot	Smooth*	Pointed	Headed
Rumex	Taproot	Smooth	Pointed*	Headed
Sorbus	Taproot	Smooth	Pointed	Headed*
Buxus ~	Taproot*	Smooth	Pointed	Headed
Hedera ~	Taproot	Smooth*	Pointed	Headed

Table 17 Continued.

Name	Root	Stem	Leaf	Flower
Trapa ~	Taproot	Smooth	Pointed*	Headed
Viscum ~	Taproot	Smooth	Pointed	Headed*
		Category B		
Learning Items				
Alnus	Fibrous*	Woody	Rounded	Spiked
Caltha	Fibrous*	Woody	Rounded	Spiked
Linum	Fibrous	Woody*	Rounded	Spiked
Morus	Fibrous	Woody*	Rounded	Spiked
Nigella	Fibrous	Woody	Rounded*	Spiked
Rubus	Fibrous	Woody	Rounded*	Spiked
Salvia	Fibrous	Woody	Rounded	Spiked*
Vitex	Fibrous	Woody	Rounded	Spiked*
Transfer Items				
Juncus	Fibrous*	Woody	Rounded	Spiked
Pilea	Fibrous	Woody*	Rounded	Spiked
Ricinus	Fibrous	Woody	Rounded*	Spiked
Salix	Fibrous	Woody	Rounded	Spiked*
Briza ~	Fibrous*	Woody	Rounded	Spiked
Hordeum ~	Fibrous	Woody*	Rounded	Spiked
Taxus ~	Fibrous	Woody	Rounded*	Spiked

Table 17 Continued.

Name	Root	Stem	Leaf	Flower
Verbena ~	Fibrous	Woody	Rounded	Spiked*

Note. Items marked with an asterisk were only presented during the test phase (base-generalization items). The transfer items marked with ~ were added in Experiment 2.

Table 18 Control Condition Stimuli.

Name	Root	Stem	Leaf	Flower
		Category A		
Learning Items				
Acorus	Taproot*	Woody	Rounded	Spiked
Cynara	Fibrous*	Woody	Pointed	Headed
Larix	Taproot	Smooth*	Rounded	Headed
Mentha	Fibrous	Smooth*	Pointed	Spiked
Nolana	Taproot	Woody	Rounded*	Headed
Ruscus	Fibrous	Woody	Pointed*	Spiked
Sedum	Taproot	Smooth	Rounded	Spiked*
Vitis	Fibrous	Smooth	Pointed	Headed*
Transfer Items				
Juglans	Taproot*	Smooth	Pointed	Headed
Punica	Taproot	Smooth*	Pointed	Headed
Rumex	Taproot	Smooth	Pointed*	Headed
Sorbus	Taproot	Smooth	Pointed	Headed*

Table 18 Continued.

Name	Root	Stem	Leaf	Flower
		Category B		
Learning Items				
Alnus	Taproot*	Smooth	Pointed	Headed
Caltha	Fibrous*	Smooth	Rounded	Spiked
Linum	Taproot	Woody*	Pointed	Spiked
Morus	Fibrous	Woody*	Rounded	Headed
Nigella	Taproot	Smooth	Pointed*	Spiked
Rubus	Fibrous	Smooth	Rounded*	Headed
Salvia	Taproot	Woody	Pointed	Headed*
Vitex	Fibrous	Woody	Rounded	Spiked*
Transfer Items				
Juncus	Fibrous*	Woody	Rounded	Spiked
Pilea	Fibrous	Woody*	Rounded	Spiked
Ricinus	Fibrous	Woody	Rounded*	Spiked
Salix	Fibrous	Woody	Rounded	Spiked*

Note. Items marked with an asterisk were only presented during the test phase (base-generalization items).

APPENDIX B

EXPERIMENTAL INSTRUCTIONS TO PARTICIPANTS IN EXPERIMENT 1 AND 2

*Learning Instructions. *(headers in italics were not read nor shown to participants)*

In this experiment, you will be evaluating the difficulty of items we plan to use in future studies. Future studies in this lab require a set of items that are comparable in their difficulty. We need you to rate how easy or hard it is to learn each of the items so we can pick which ones to use in the future.

In order to ensure that your ratings are as accurate as possible, you are going to see the items several times and rate how difficult it was to learn them several times. It is important that you try and learn the items so that your difficulty ratings are as accurate as possible.

In this experiment, you will be learning about plants.

You will see information about a series of plant samples, including the Latin names and characteristics of those plants

The characteristics include the roots, stems, leaves, and flowers.

Type in your answer, and the computer will let you know if you were correct.

For example, a trial would look like:

Azium, seed is:

There is a Latin plant name (Azium), and you need to type in the seed characteristic for it.

In this introduction, you will have to press the spacebar for the next screen.

At first, you will not know what the correct answer is.

Type your guess, and the computer will let you know if you were correct or incorrect, and it will show you the correct answer.

Continuing the example, if you entered 'hard', you would see:

Azium, seed is: hard

Incorrect

The correct answer is flat

Press the spacebar for the next trial.

You will only see each plant sample a few times, so use the feedback to learn the characteristics of the plants.

It is very important that you learn about these plants, because your evaluations of their difficulty will be best if you try to do what future participants will have to do.

After some items, the computer will prompt you to evaluate how difficult it was. Use the number keys and the scale on the screen to indicate how difficult you thought the item was.

The computer will present additional instructions later in the experiment.

Follow the instructions to the best of your ability, and ask the experimenter if you don't understand the instructions.

If you have any questions, please ask the experimenter now.

If you are ready to begin, press the spacebar to start.

Difficulty Task Instructions

Please use the following scale to rate the previous item for how difficult it was to learn:

1	- - - - -	2	- - - - -	3	- - - - -	4	- - - - -	5	- - - - -	6	- - - - -	7
not at all difficult												*extremely difficult*

Enter a number from 1 to 7:

Debriefing.

The purpose of this research is to learn how people learn about categories when they have not been told to look for them.

Do you have any questions?

This experiment depends on people not knowing that there are categories. Please don’t discuss this experiment with other students. Thank you

APPENDIX C

POST-EXPERIMENTAL QUESTIONNAIRE

AND ITS SCORING PROCEDURES

Post-Experimental Questionnaire

Page 1

Please answer all of the following questions in order. Answer all of the questions even if you feel that you are repeating information. Please show each page to the experimenter before turning to the next page.

1. How did you go about learning the plants in the first part? Try to include all the methods that you used.

2. How did you go about learning the *new plants* at the end of the experiment? If you used any of the same methods, please indicate which ones you used.

Page 2

3. New plant specimens have been discovered. Based on what you know about the plants in this study, fill in the blanks with the most likely features. Do the best you can even if you have to guess.

Carex, stem is: _____

Carex, leaf is: _____

Carex, root is: _____

Carex, flower is: _____

Nepeta, stem is: _____

Nepeta, leaf is: _____

Nepeta, root is: _____

Nepeta, flower is: _____

Prunus, stem is: _____

Prunus, leaf is: _____

Prunus, root is: _____

Prunus, flower is: _____

Vicia, stem is: _____

Vicia, leaf is: _____

Vicia, root is: _____

Vicia, flower is: _____

Page 3

4. Were any of the items used in this research familiar? (circle one) Yes No

5. If yes, which ones?

Page 4

6. Did you notice if any of the plants or their features had anything in common with each other? If so, what did they have in common?

Page 5

7. The plants belonged to groups. Describe how those groups were formed. Do the best you can and guess if you have to.

8. If you noticed that the plants belonged to groups, when did you notice?

Scoring Procedure for the Post-Experimental Questionnaire

The four questions on the post-experimental questionnaire will be scored for awareness of the category structure used in the study. The questions will be coded for no, partial, and full awareness as follows.

Question 1 and 2

If participants mention the complete category structure by listing all of the feature values for at least one of the categories, that participant will be assumed to have full awareness. Partial awareness will be assumed if the participant correctly lists part of the category structure by listing part of the feature values for one or both categories, or if participants make reference to using the first item to get the second and third items correct. If there was no mention of categories, or the only mention of categories is based

on false criteria (e.g., first or last letter or syllable), then that question will be recorded as no awareness.

Question 3

Knowledge of the category structure will be inferred if participants report feature values from the same category for an item. The following scoring procedure will be used in an effort to liberally assign category knowledge. No awareness will only be assumed if two of the four features are from each category. These items will be assigned a value of 0. Partial awareness, and a score of 1, will be assigned to Question 2 items that have three features from the same category. Full awareness will be inferred if all four feature values are from the same category. These items will be assigned a value of 2. The total will then be summed to produce an awareness score ranging from 0 to 8. Features with incorrect feature values (e.g., a value not associated with that feature or a value not used in this study) will be coded as 0.

Questions 4 and 5

Questions 4 and 5 were manipulation checks to ensure that the materials were initially meaningless. Question 4 was a count of the number of Latin plant names mentioned in Question 5. Question 5 was scored as the number of times each Latin plant name was mentioned.

Question 6 and 7

Questions 6 and 7 will be coded as Question 1.

Question 8

Question 8 will be scored to assign no awareness or awareness originating in one of the following epochs: during pretraining, during learning, during the test phase, during the transfer phase, during the transfer-test phase, or during the completion of the questionnaire.

APPENDIX D

EXTENDED REVIEW OF THE LITERATURE

Our knowledge of the world is not a random collection of unconnected objects, events, and principles. Rather, these concepts and examples are organized in a manner that reflects the regularity of our experience of the world. Categories represent one form of organized knowledge about the world. Knowing the category of an object or event allows us to infer characteristics and properties that we have not yet seen or cannot observe, based on those that we have (Ashby & Maddox, 2005; Medin 1989; Medin & Schaffer, 1978; Nosofsky, 1986, 1988; Posner & Keele, 1968; Rosch & Mervis, 1975). They allow us to extend our limited store of experience to a wide range of novel objects and events, guiding our interactions with a world that is full of new instances of familiar types. To have categorical knowledge is to have a plan for coping with a world defined by perpetual newness.

Category knowledge represents both the similarity of its members to one another, and the range of variability of members within a category. When something new is encountered, its similarity to existing categories is determined by its overall similarity to the group, either to group members (Nosofsky, 1986, 1988), or to a representation of the group as a whole (Minda & Smith, 2001; Posner & Keele, 1968; Smith & Minda, 2002). The more homogeneity in the category, the more closely new instances must resemble old instances to be included in the category (Čech & Shoben, 2001; Clapper & Bower, 1994, 2002).

Issues in Category Learning Research

Research on category learning is extensive and ongoing, a testament to both the complexity and the essential nature of categories in human cognition. Research in this area has illustrated several key issues and areas, many of which overlap, including theories of categorization, category learning tasks, the number of category learning systems, classification versus category structure learning, and supervised versus unsupervised category learning.

Theories

Exemplar and Prototype Theories

One of the earliest issues in the study of category learning focused on the type of representational process that was used to organize category information. The two main theoretical perspectives that emerged were based either on specific instances or prototypes (Ross & Makin, 1999).

Exemplar theories hold that categories are defined by the individual members they contain (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986, 1988). The decision whether a new instance is part of a known category depends on the similarity of that new instance to each of the stored exemplars (Nosofsky, 1986, 1988). Exemplar models are flexible categorization models in the sense that similarity can be determined based on any stimulus dimension of the stored exemplars, though prior experience will direct attention to more relevant features, biasing future categorization toward known diagnostic features (Kruschke, 1992).

According to prototype theories, categories are represented by an ideal (prototype) formed from the averaging of all category members (Fried & Holyoak, 1984; Minda & Smith, 2001; Posner & Keele, 1968; Reed, 1972; Smith & Minda, 2002). New instances belong to the category if they sufficiently resemble the prototype. Prototype theories are less flexible than exemplar theories in that the prototype becomes more or less well-defined as previous instances are incorporated into the prototype representation (the individual exemplars are lost from memory and/or not used to make the categorization decision). This means that prototypes are excellent for capturing linear category relationships, where the categories are separated based on a weighted sum of their feature dimensions, but have difficulty explaining how people learn non-linear category structures (Medin & Schwanenflugel, 1981).

Theory-Based and Rule-Based Theories

A shift in basic reasoning about category learning occurred when experience models (exemplar and prototype models) collided with the higher-order cognitive reasoning about categories. Theory-based category learning theories noted that natural categories, such as trees or birds, have an internal consistency that relates features to one another in a coherent and informative manner (e.g., Murphy & Medin, 1985). Theory-based categories are defined by more than feature similarity. Categories reflect a causal understanding of the functional relationships between features and their interaction with other world knowledge. For example, knowing that robins, geese, and pelicans can fly

and that all three have feathers also reflects our understanding that feathers are related to flying.

Rule-based category learning argues that people either construct ad hoc rules that determine category membership (Billman & Heit, 1988), or use given rules to classify instances, as in the Wisconsin Card Sorting Task (Berg, 1948). The Wisconsin Card Sorting Task requires participants to sort a set of geometric icons into different piles based on a given rule (e.g., by color or by shape). Rule-based categories are more rigid than theory-based categories. Whereas theory-based categories are tolerant of exceptional members (e.g., ostriches as a non-flying member of the bird category), rule-based categories are more rigid, often having to distinctly encode the exception separately. Rule-based categories are sometimes at odds with exemplar and prototype theories (Nosofsky, Clark, & Shin, 1989; Smith, Patalano, & Jonides, 1998). Some have even argued that rules may be little more than descriptions of the output of exemplar models (Nosofsky et al., 1989). More recent models of category learning include provisions for rules and exemplar categorization (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994).

Decision Bound Theories

A new class of category learning theory has recently been advanced. According to decision bound theories, category learning is learning where the boundaries between categories fall, associating regions of the feature space with category labels (Ashby et al., 1998; Ashby & Gott, 1988; Maddox & Ashby, 1993). New exemplars belong to a

category if they fall within the decision boundary for a category, where decision boundaries are defined by the exemplar's features and the response associated with those features. In other words, the decision boundary that separates categories is defined by the combination of the features possessed by the learning items and the responses to those learning items (i.e., which category the item belongs to). These theories have an advantage over prototype theories in that they can explain how participants learn non-linear categories (Ashby et al., 1998), and provide a better fit to the learning data for nonlinear and non-normally distributed categories (Maddox & Ashby, 1993).

Connectionist Theories

Connectionist theories come in several forms and have been used as theoretical models in their own right (e.g., the adaptive network model of Gluck & Bower, 1988), and as implementations of prototype, exemplar (Kruschke, 1992; Kruschke & Johansen, 1999; McClelland, 1981; Nosofsky, Kruschke, & McKinley, 1992), decision bound (Ashby & Gott, 1988; Ashby et al., 1998), and hybrid models (Anderson & Betz, 2001; Love, Medin, & Gureckis, 2004). Typically, connectionist theories rely on networks to calculate the category or feature values based on the pattern of inputs (features and/or category label) of the training stimuli. Rules are typically not encoded explicitly, but are an emergent property of the weighting of feature values by the network (but see Ashby, et al., 1998, for a connectionist implementation of rules and Kruschke, 1992, and Nosofsky, Palmeri, & McKinley, 1994, for hybrid connectionist-plus-rules systems). Connectionist networks are also capable of representing the feature correlations that define the structure

of a category. Because the inputs to a connectionist network comprise the features of the item, they are capable of modeling human performance under conditions of incomplete information, reflecting the fact that human learners do not always have perfectly reliable memory or access to all features of a stimulus.

Category Learning Tasks

Rule-Based Tasks

Category learning experiments have employed a variety of experimental tasks that provide insight into different aspects of category learning. Several tasks have been used in category learning research, but they fall into several broad categories. The first, and most common, are rule-based category learning experiments (Ashby & Maddox, 2005; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Shepard, Hovland, & Jenkins, 1961; Smith, Patalano, & Jonides, 1998). Rule-based category learning tasks are those that are based on one or more rules that defines which features or combination of features define members of a category. According to Ashby and Maddox (2005), rules are relatively simple and verbalizable statements that describe category membership based on perceptual similarity among features of items. Rules can be based on a single dimension (the red things in one category, blue things in another) or on the conjunction of two or more dimensions (the red triangles in one category, blue squares in a second category, and everything else in a third category). Rule-based tasks have been used to support exemplar, prototype, and rule-based category learning theories.

Information-Integration Tasks

Information-integration tasks require simultaneous evaluation of two or more feature dimensions to determine category membership. Information-integration differs from rule-based category learning because the former occurs before evaluations are made on independent stimulus dimensions (Ashby & Gott, 1988), thereby preventing the formulation of a verbalizable description of how the feature values define the category or categories. For example, a task might require participants to learn which category a set of geometric figures belong to, where the figures vary in size and in shade. A rule-based category structure would be sorting the items based on size or shade or some combination of size and shade (e.g., small items and/or dark items belong to category A, large items and/or light items belong to category B). Information-integration category structure would also use values on both stimulus dimensions but the combination of size and shade defining category membership wouldn't be discrete; rather category membership would require either the weighted integration of two or more different dimensions, or the evaluation and classification of the item as a whole. This process cannot be verbalized because the contribution of a feature dimension is combined with other feature dimensions predecisionally; the individual contribution of a stimulus dimension cannot be stated either alone or in terms of discrete values of another stimulus dimension. Information-integration tasks provide the most support for decision bound theories (Ashby & Gott, 1988) because their results are problematic for prototype theories (Smith & Minda, 2002), as they require non-linear decision bounds, and human performance on

these tasks is over-predicted by exemplar theories (e.g., those of Nosofsky, 1986, 1988; Smith, et al., 1998).

Prototype Distortion Tasks

Prototype distortion tasks (Posner & Keele, 1968; Smith & Minda, 2002) typically present participants with a selection of dot patterns that are distortions of an underlying prototype. Typical results show equal or greater learning of the unseen prototype than for seen exemplars. These tasks are often cited in support of prototype theories (Posner & Keele, 1968; Smith & Minda, 2002; but see Nosofsky et al., 1992, for an exemplar-based explanation).

Number of Category Learning Systems

Recent investigations have focused on the question of whether there is a single category learning system, or multiple systems that mediate category learning. This issue is intertwined with the issue of explicit versus implicit category learning, based on the dissociation of human memory systems (e.g., Tulving, 1991). Knowlton and Squire (1993) argued that it is the implicit memory system that is responsible for category learning whereas the explicit system is responsible for item memory. These researchers found that amnesic participants were not significantly different from normal participants in tests of category learning, even though they were significantly worse at recognizing the stimulus items used in the experiment. Explicit category learning draws on the working memory (WM) system and is based on item memory (Knowlton & Squire, 1993) and

hypothesis testing (Ashby et al., 1998). As such, it is best suited to learning individual items and rule-based category distinctions. Implicit category learning is based on perceptual priming and/or the procedural memory system. This system is best suited to learning associations between stimuli and responses (i.e., information-integration category learning) and, though it is slower to train than the explicit system, it is capable of faster performance and a higher level of ultimate mastery (Ashby et al., 1998). A general finding suggests that rule-based hypothesis testing is generally faster than incidental or implicit category learning when the rule defining category membership is simple and the stimuli easily discriminable, but not when the rule is complex or the stimuli easily confusable (Ashby et al., 1998, Ashby & Waldron, 1999; Gureckis & Love, 2003b). It is possible that the dot distortion stimuli used in Knowlton and Squire (1993) were not easily captured by a simple rule and therefore derived no benefit from the application of explicit memory resources by the non-amnesic participants.

Classification Learning Versus Category Structure Learning

Classification Learning

The difference between classification learning and category structure learning hinges on what is learned and what aspect of the category representation people emphasize and use. Most work on category learning has examined classification learning, or how people learn which items belong to which groups. These studies typically employ a sorting or category prediction methodology where participants are given instances and are instructed to sort them into groups or to predict which category

they belong to (Berg, 1948; Medin & Schaffer, 1978). These experiments focus on how people use feature information to determine category membership and are best described as investigating discrimination ability.

Category Structure Learning

Category structure learning studies have been described under many different names including schema abstraction (Elio & Anderson, 1981), comparison-based learning (Spalding & Ross, 1994), inference learning (Anderson, Ross, & Chin-Parker, 2002; Yamauchi, Love, & Markman, 2002; Yamauchi & Markman, 1998), similarity-based learning (Love, 2002), and category induction (Taraban, 2004, 2006; Taraban & Hayes, 2000, 2001), but all are concerned with understanding how people acquire the internal structure of category representations: knowledge of the relationships between feature within a category. These studies often focus attention onto the feature relationships within categories either through experimental instructions (e.g., to compare similar items; Estes, 1986, Spalding & Ross, 1994) or via experimental manipulations (e.g., to make predictions of the feature values; Anderson et al., 2002). A missing feature methodology is often used, querying a missing feature either when the category affiliation is known (Anderson et al., 2002) or unknown (Elio & Anderson, 1981; Spalding & Ross, 1994), or limits access to feature information (Billman & Knutson, 1996). Billman and Knutson (1996) presented a series of imaginary animals that varied on several feature dimensions. In their test of category learning, they obscured features of novel stimuli to investigate participants' internal category representation of the

relationships between features and their values. Category structure learning experiments demonstrate that *how* we learn a category determines *what* we know about the category. The preponderance of evidence suggests that category structure learning creates a category representation that has a much richer internal structure (Chin-Parker & Ross, 2002; Yamauchi & Markman, 1998; Yamauchi et al., 1998), but may not help people learn which features are predictive of category membership (Anderson et al., 2002; Wattenmaker, 1991, 1992). However, there is also evidence that knowledge of within-category structure is acquired as a byproduct of classification learning (Anderson & Fincham, 1996; Elio and Anderson, 1981; Schyns & Rodet, 1997). Anderson and Fincham (1996) found that the type of classification learning task (using experimenter provided or self-generated categories) had no effect on category structure learning, but both types of classification learning lead to learning of the category structure (acquisition of within category feature correlations) whereas observation without classification did not. Schyns and Rodet (1997) found that classification learning on a set of training stimuli altered the representations of within category feature correlations, indicating that category structure learning is affected by which features are present during learning and the order of feature exposure.

The feature correlations of the experimental stimuli affect the amount of within-category learning that occurs. Richer internal correlational structure aids category structure learning. Billman and colleagues (Billman & Heit, 1988; Billman & Knutson, 1996; Kersten & Billman, 1997) found that more overlapping intercorrelations among features (termed *systematicity* by the authors) resulted in greater learning of the category

structure. When the number of intercorrelated features is lower (less systematicity), learning is also slower. Learning is also slower when learning the same number of feature correlations is distributed among different features rather than the same features. Billman and colleagues (Billman & Heit, 1988; Billman & Knutson, 1996; Kersten & Billman, 1997) found that participants who experienced items whose feature correlations overlapped (e.g., A-B, B-C, A-C, D, E, F) learned the category structure better than a group who saw a different pattern of correlations in the same features (e.g., A-B, C-D, E-F).

Supervised Versus Unsupervised Category Learning

Supervised Category Learning

Research on category learning has primarily investigated supervised category learning; learning about categories when the learner knows that categories exist and receives feedback regarding their performance. Prototypical explicit category learning tasks, either classification or category structure learning and including information-integration (Ashby & Gott, 1988) and prototype distortion tasks (Smith & Minda, 2002), provide feedback about one's performance after each item. Certainly this type of category learning occurs, but it cannot account for all of our categorical knowledge (Billman & Knutson, 1996; Love, 2002; Love et al., 2004; Allen & Reber, 1998). Much less effort has been directed toward investigating unsupervised category learning; acquiring category knowledge without external corrective feedback.

Unsupervised Category Learning

Category learning in the real world often occurs without the benefit of receiving feedback that the choice of category was correct (Fried & Holyoak, 1984; Gureckis & Love, 2003a, 2003b; Imai & Garner, 1965; Kaplan & Murphy, 1999; Pothos & Chater, 2005; Spalding & Ross, 1994), or even that categories exist (Billman & Heit, 1988; Billman & Knutson, 1996; Brooks et al., 2006; Clapper & Bower, 1994, 2002; Gureckis & Love, 2002, 2003a, 2003b; Kersten & Billman, 1997; Reber, 1967, 1998; Taraban, 2004, 2006; Taraban & Hayes, 2000, 2001). Much of our knowledge of linguistic rules and classes, social rules and norms, and object classes and properties is acquired without the benefit of a tutor noting the category of the object or instance in question (Allen & Reber, 1998; Kelly, Burton, Kato, & Akamatsu, 2001). Rather, our knowledge of the categories of the world is a product of learning about the world. We may, for example, instruct a child to give a bone to Fluffy (the family dog) without instructing the child that dogs like bones. It is up to the child to infer that what is true for Fluffy is also true for dogs in general.

Unsupervised learning tasks have demonstrated that category learning is possible without an external tutor to provide feedback about category membership by allowing participants to freely construct categories however they choose (Fried & Holyoak, 1984; Imai & Garner, 1965), presenting category features simultaneously (Billman & Heit, 1988; Clapper & Bower, 1994; 2002; Kersten & Billman, 1997), directing attention to the features of the exemplars (Clapper & Bower, 1994; Spalding & Ross, 1994), as a result of explicit learning of individual instances (Brooks, Kempe, & Sionov, 2006; Taraban,

2004, 2006; Taraban & Hayes, 2000, 2001), and even when participants are supplied with a cover story that biases them to process the exemplars in a way that distracts them from category processing (Love, 2002).

Most investigations of unsupervised category learning occurred with explicit knowledge of categories, either that items belong to categories or via explicit instructions to use an experimenter-determined or self-determined number of categories. Typically, these investigations used classification learning; participants are instructed to sort stimuli into groups, but are not given feedback regarding their classification accuracy (e.g., Ashby, Queller, & Berretty, 1999; Fried & Holyoak, 1984; Imai & Garner, 1965; Pothos & Chater, 2005; Spalding & Ross, 1994). These investigations have lead to some general conclusions regarding unsupervised category learning. First, unsupervised classification learning is possible, both when the number of categories is given and when the sorting task is unconstrained. Second, and more importantly, unsupervised classification learning is more limited than supervised classification learning, possibly due to the type of underlying category representation formed (Ashby et al., 1999).

Unsupervised classification learning is characterized as favoring simple rules (Ashby et al., 1999) or simple prototypes (Fried & Holyoak, 1984; Spalding & Ross, 1994). Ashby et al., (1999) found that participants utilized unidimensional sorting strategies in the absence of feedback, even when the unidimensional rule was suboptimal and perfect accuracy was possible. Imai and Garner (1965) found that participants focused on attribute values that were easily discriminated as the basis for classification. Spalding and Ross (1994) found that participants focused on subsets of features and

formed abstractions by comparing features across the set of learning instances, a process essential for prototype formation. Pothos and Chater (2005) found that focusing attention on relevant feature dimensions through supervised category learning produced greater gains in unsupervised classification learning if that prior learning promoted learning of separable dimensions (i.e., linearly separable dimensions or prototypes).

Unsupervised inference learning has received little attention in the research literature. In these experiments, participants either view or learn a series of exemplars and are presented with a recognition task. Participants are usually not told to look for or use categories, and category use often is not necessary to complete the task. Knowledge of the category structure is assessed by asking participants to predict the missing feature values of new items, or discriminate which new instances follow the pattern from those that violated it.

There have been three main explanations for unsupervised learning: explicit rule testing, implicit (or incidental) learning, and surprise (or the capture of awareness). Each of these positions will be reviewed below.

Explicit Rule Testing

Several researchers (Billman & Heit, 1988; Billman & Knutson, 1996; Kersten & Billman, 1997; Spalding & Ross, 1994) explained people's ability to acquire knowledge of category structure (category structure learning) in unsupervised learning as a repeated process of hypothesis generation and testing. The hypothesis takes the form of a simple positive association between two features (A-B are present at the same time) and is

evaluated against the next relevant instance where relevance is determined by feature values (Billman & Heit, 1988; Billman & Knutson, 1996; Kersten & Billman, 1997) or a biasing feature of the instance (Spalding & Ross, 1994). If the hypothesis is incorrect, a new hypothesis is formed and tested. If the hypothesis is validated by subsequent items, then people are more likely to try to relate a third feature to the first two, expanding the relationship from two features (A-B) to three (A-B-C), rather than trying a completely unrelated hypothesis (C-D) after discovering the initial (A-B) relationship. Hypothesis testing in this context means conscious hypothesis testing; participants are supposed to be intentionally looking for relationships among features of the items (i.e., categories) as the basis of their experience of those items.

Implicit or Incidental Category Learning

While conscious hypothesis testing doubtless occurs, our knowledge of the world is not exclusively the product of explicit exploration (Allen & Reber, 1998). Research on implicit learning has revealed that people are capable of acquiring some knowledge of complex patterns and classes without intention or awareness (Nissen & Bullemer, 1987; Reber, 1967, 1989). This knowledge corresponds to both superficial features (i.e., perceptual features; Perruchet, & Pacteau, 1990) and abstract relationships (Bolte & Goschke, 2004; Grafton, Hazeltine, & Ivry, 1998), either of which could form the basis for forming categories. Traditionally, work in category learning assumes that category knowledge is a byproduct of experience of the items themselves (Gureckis & Love, 2003a, 2003b; Wattenmaker, 1991, 1992). However, it appears that the type of

interaction with learning items determines what the category representation comes to contain. Both the order of presentation (Clapper & Bower, 1994, 2002), the amount of idiosyncratic or irrelevant feature dimensions (Medin, Dewey, & Murphy, 1983), and the focus of the learning task (e.g., at the exemplar learning or category learning level, Medin et al., 1983; observation vs. classification, Anderson & Fincham, 1996; classification vs. pleasantness rating, Love, 2002) affect the nature of the category representation formed.

To date, only a handful of experiments in unsupervised category learning have investigated implicit category learning where participants are unaware that a category structure exists and remain unaware of the existence of categories. Ashby and colleagues (Ashby et al., 1998; Ashby & Speiring, 2004; Ashby & Waldron, 1999, 2000; Maddox & Ashby, 2004; Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005) have proposed an implicit learning system, but implicit to these researchers refers to the internal representational learning process, not the experimental parameters. Their COVIS (COmpetition between Verbal and Implicit Systems) model postulates that there are two category learning systems for perceptually-based category learning; an explicit verbal system and an implicit procedural system, based in two separate memory systems. The explicit system relies primarily on the WM system and learns via explicit rule testing, which results in a sudden improvement in performance when the correct rule is finally found. The implicit system relies on the basal ganglia and learns via a slow incremental process of adjusting response biases. "Implicit" to these researchers describes the fact that participants cannot articulate the basis for their categorizations, not with regard to the awareness of categories. This type of implicit category learning experiment is conducted

with full awareness that categories exist and requires immediate feedback to operate (i.e., supervised learning; Ashby et al., 1998; Ashby, Maddox, & Bohil, 2002; Ashby & Waldron, 1999; Maddox et al., 2003; Maddox & Ing, 2005). This is not the type of implicit or incidental category learning meant here.

Work in implicit learning demonstrates that completely implicit category learning, where implicit refers to both the processing done and the lack of awareness of the category structure, is possible (Nissen & Bullemer, 1987; Reber, 1967, 1989). Participants in artificial grammar learning experiments memorize a series of letter strings generated from a finite state grammar (Knowlton & Squire, 1996; Reber, 1967, 1989). Following this memorization task, participants are given a forced-choice task where they must identify which new exemplars follow the same rules. Despite denying conscious awareness of the rules governing the items and displaying no significant ability to describe the rules of the grammar, participants exhibit knowledge of the category structure (i.e., the structure of the finite state grammar) as operationalized by their above chance ability to sort the items into grammatical and ungrammatical sets (Reber, 1967, 1989).

Another type of implicit category learning would be better characterized as incidental category learning and refers instead to the process of acquiring knowledge of the categories and their structure. In other words, category knowledge is an unintended result of learning or interacting with the stimuli (Love, 2002; Taraban, 2004).

In a unique study, Love (2002) compared supervised classification learning with unsupervised classification and unsupervised category structure learning. The

experiment was based on Shepard et al.'s (1961) experiment on the effects of rule complexity on category learning. Love (2002) used four (Type I, II, IV, VI) of the original Shepard et al. (1961) six classification problem types, corresponding to single dimension categorical structure (Type I), two dimensional categorical structure (Type II), family resemblance categorical structure (Type IV), and three dimensional categorical structure requiring explicit memorization of each exemplar with its category label (Type VI). The dependent measure was an old/new forced choice test where the new items differed from the old items on a single feature dimension (the stimulus border color). Love (2002) found that unsupervised learning is sensitive to the type of processing and the category organization (i.e., problem type) present in the stimuli. Unsupervised learning did not produce the same order of learning difficulty found in supervised learning of the four problem types. Furthermore, differences were found between intentional (classification) and incidental (category structure) learning under unsupervised conditions. Unsupervised classification learning (intentional learning) was better suited to clearly defined rule-based category distinctions (Type II problem) whereas category structure learning (incidental learning), which biases the participant to focus on feature similarity rather than category membership, was better suited to learning linearly separable categories (i.e., forming prototypes; Type IV problem). These results strongly support the conclusion that the type of processing used to encode the stimuli determines what the category representation becomes more than whether or not feedback is provided. In this case, the difference between category structure and classification learning was the difference between learning the internal structure of the category feature

relationships via prototype formation (category structure learning) as opposed to learning which features discriminate category members (classification learning) even without feedback.

Taraban (2004, 2006; Taraban & Hayes, 2000, 2001) has investigated participants' ability to induce category structures in artificial language learning (Taraban, 2004, 2006) and general fact learning (Taraban & Hayes, 2000, 2001) paradigms. Participants in Taraban (2004) had to learn simple phrases in an artificial language consisting of a noun and a locative preposition that followed the noun (postposition; e.g., *to car = garth eef*). Participants saw one noun-preposition pair at a time and had to translate the pair into the artificial language. This process repeated until the participants achieved a high level of accuracy (100% on a block of trials). There were two forms of the locative postpositions in the artificial language and, unbeknownst to the participants, the postpositions were organized into two perfectly reliable categories, such that knowing the form of one postposition for a noun would allow the ideal learner to know the form of the other two postpositions for that noun with perfect accuracy. Represented abstractly, the two categories conformed to the feature value patterns *111* and *222*. Because the participants' task was to learn the phrases, they could master the training items without knowing the category structure, or even that the locative postpositions were organized into two categories. The nouns were either marked with long suffixes (*-oik, -oo*), short suffixes (*-a, -o*), or were unmarked (no consistent suffixes), where the suffixes were perfectly correlated with the form of the postpositions.

These were unsupervised category learning experiments because, even though participants received feedback regarding their memorization accuracy, they did not receive feedback regarding their knowledge of the category structure. These were category structure learning experiments because the learning test measured knowledge of the within category feature structures by utilizing test items containing unseen values of familiar items (probing knowledge of the category structure for familiar items) and novel items with familiar features (probing knowledge of the category structure for novel instances of the categories). Taraban (2004, 2006; Taraban & Hayes, 2000, 2001) investigated several factors that affect people's ability to acquire category structure within this paradigm that does not require knowledge of categories to perform the given task including feature marking, clustering, and frequency manipulations.

Taraban (2004) compared noun-cue learning to syntactic context learning utilizing an artificial language learning task. In noun-cue learning, the gender agreement between a noun and its article or adjective is acquired from the co-occurrence of the noun's suffix with the form of the article or other part of speech, as in Spanish where an –*a* ending indicates the feminine form of the definite article (*la*) and an –*o* ending indicates the masculine form (*el*). Syntactic context hypotheses suggest that people use the grammatical morphemes as part of the information used to construct syntactic categories, though information other than noun cues is used (e.g., word order or the co-occurrence with a limited set of other grammatical morphemes).

In a series of five experiments, Taraban (2004) demonstrated that participants were able to induce the structure of the two categories used in the experiment, both from

noun-cues and also from the syntactic context without noun-cues, though greater care was needed to draw the participants' attention to the relevant relationships. One effective means of focusing learners' attention to the relevant relationships was through blocking the learning items. Requiring participants to achieve perfect accuracy on both of the postpositions for a noun use in the learning task (the remaining postposition was reserved for a test of category knowledge) before moving on to the next noun, showing them the two groups formed by the postpositions, and providing a comparison "hint" item on the test trials were all manipulations that directed participants' attention to the relevant relationships and enabled them to acquire the category structure.

The materials in Taraban (2004) formed categories that were linearly separable and perfectly reliable, making it possible for participants to have represented the category either via a prototype or a set of rules. This is consistent with the assertion that unsupervised learning is most amenable to prototype formation (Fried & Holyoak, 1984; Gureckis & Love, 2003a, 2003b; Love, 2002) especially since the manipulations that resulted in improved category structure learning were derived from manipulations of attention, which are associated with prototype formation (Pothos & Chater, 2005; Spalding & Ross, 1994). However, it is not possible to eliminate the possibility that participants engaged in explicit hypothesis testing (e.g., Billman & Knutson, 1996) or that their knowledge of the category structures was a combination of knowing about the feature correlations (i.e., category structure) and knowing which items belonged to which category (i.e., classification).

Taraban & Hayes (2000, 2001) utilized a category structure similar to Taraban (2004) instantiated in a fact-learning paradigm. Participants were told that they would learn about a series of people, including their education, occupation, and gang membership. Each of these attributes had two possible values and, as in Taraban (2004), they were organized into two categories based on feature values. Taraban and Hayes (2001) found that clustering the presentation of a noun's phrases before presenting the next noun's phrases facilitated induction of the category structure (although Taraban & Hayes, 2000, did not), supporting the hypothesis that category structure learning results in prototype construction, but attention to the relevant feature dimensions is necessary to form the abstractions.

In sum, these few studies of incidental category structure learning suggest that participants can acquire knowledge of the category structure, though the type of category structure and stimulus exposure matter. Love (2002) found that participants learned family resemblance category structures (i.e., prototypes) more easily under incidental unsupervised inference conditions, whereas rule-based categories were more easily learned under intentional unsupervised conditions. Taraban (2004, 2006; Taraban & Hayes, 2001) found that attention is critical for inducing the category structure and that manipulating the order of presentation was effective in increasing learning of the category structure.

Surprise and the Capture of Awareness

Clapper and Bower (1994, 2002) proposed that unsupervised category learning results from discrepant items capturing awareness and spurring the creation of a new category to account for the difference between the previous items and the new, different, item (the category invention hypothesis). According to these researchers, the process of learning is not itself an explicit effort to discover category distinctions, nor is it the product of an implicit learning system; rather, it is the result of the mind noticing that this new thing is not like the old things and recruiting additional representational units to describe those differences (see also Anderson, 1991; Gureckis & Love, 2003a, 2003b; Love, 2002; Love et al., 2004; Spalding & Ross, 1994; Taraban, 2004).

The Pattern-sequence effect. In a series of three experiments, Clapper and Bower (1994) investigated unsupervised category learning using stimuli that presented all the category features simultaneously combined with an attribute-listing task. They presented line drawings of fictitious insects to participants. The participants were instructed to write a list of the distinctive features of each instance that would aid future participants in differentiating the insects. Participants were told to use only those features that were unique to each exemplar. Unbeknownst to the participants, the insects belonged to one of two groups defined by shared feature values. These shared features were prominent (e.g., fat vs. thin abdomens, wide vs. narrow wings), and the defining features were less noticeable (e.g., eye color). Greater reliance on the defining features in the descriptions was interpreted as evidence of greater category knowledge.

Clapper and Bower (1994, 2002) proposed that initial exposure to new exemplars affected the organization of exemplars into categories during unsupervised category learning (see also Spalding & Ross, 1994, for similar conclusions in explicit category learning). Being naturally parsimonious creatures, we organize our experience into no more categories than necessary to capture the variability in the natural world, assuming that new items come from the same category unless they are significantly different from one another or we are told otherwise. The amount of variability in the first few items we encounter defines the range of variability for that set or category. A new item outside this region would cause the person to invent a new category to represent the new items. By manipulating which category the initial exemplars belonged to, Clapper and Bower (1994) affected the amount of category learning. Presenting the items in a blocked design, where all of the items for Category A were presented before the items for Category B (Experiment 1), resulted in faster acquisition of the category structure than a mixed presentation of Category A and B items (the pattern-sequence effect), though the level of performance after 32 trials did not differ between these groups. Presenting exemplars from only one category for the first 16 trials (Experiment 2) followed by a mixed presentation of both categories also produced significantly greater category learning than presenting an equal mix of both categories, as indicated by insect descriptions containing significantly more defining features in the former experimental condition than the latter.

Clapper & Bower (2002) modified the learning task used in Clapper and Bower (1994), deemphasizing the instructions to attend to the unique features. The stimuli for

this series of three experiments consisted of descriptions of 32 trees belonging to one of two categories based on the values of their 12 descriptive features. Of the twelve features, nine were perfectly predictive of category membership, having one value for one category, and a different value for the other category. The categories were defined by the mutually exclusive values on the nine intercorrelated features, such that Category 1 would be represented as having feature values *111111111xxx* and Category 2 would be represented as having feature values *222222222xxx*, where x was balanced across all four possible feature values regardless of category. Unlike in Clapper and Bower (1994), participants were not instructed to focus on the informative or distinct features. They were simply instructed to study the features for one of the trees, and then answer multiple-choice questions about it. Participants had access to all 12 features of a tree for 24s during the study period, but could only select and view one feature at a time. The remaining features were masked. Following the study period, participants answered 12 multiple-choice questions about the tree they just saw. There was one question with four response options for each of the 12 features, presented in random order. Participants received feedback at the end of the question set indicating how many of the 12 items they answered correctly.

Clapper and Bower (2002) utilized two dependent measures of category learning. The first was based on the nominal task given to participants; their performance on the multiple-choice questions for the final 8 trees. The second dependent measure of learning was the proportion of time participants spent viewing the three uncorrelated features to the time spent viewing the nine correlated features. This was based on the

logic that participants would learn which features went together and, therefore, spend their time looking at and learning only those features that varied from item to item. If the ideal learner acquired the category structure, he/she would have perfect knowledge of the nine intercorrelated features after viewing only one of them, and would not need to dedicate time to learning them. Therefore, the more participants sought out and viewed the idiosyncratic features, the more they were assumed to have acquired the category structure (provided that their accuracy on the recognition questions for the intercorrelated features did not suffer relative to the idiosyncratic features).

Consistent with the prediction that initial exposure affected the formation of categories, both the behavioral data and the viewing time data supported the category invention hypothesis. Clapper and Bower (2002) found that participants spent significantly longer looking at the uncorrelated features than the correlated ones when, as in Clapper and Bower (1994), they had differential amounts of initial experience with the two categories (Experiments 1 and 2), but their accuracy on the recognition questions for the idiosyncratic features did not differ from the consistent features. The use of structured stimuli was not enough to produce category learning. Participants who saw items from the two categories intermixed since the start of the experiment demonstrated no greater category learning, either in answer to the multiple-choice questions or in viewing time, than participants in a control condition that used stimuli with random feature patterns (Experiment 1).

Clapper and Bower (1994, 2002) theorized that people initially encounter objects, they tend to group them together in a single category unless they have some reason to

suspect that the objects belong to multiple categories or the objects are very different from one another. Without expectations of multiple categories, the differences in feature values of newly encountered objects are often viewed as defining the range of variability to be found in any category of objects. In Clapper and Bower (1994, 2002), participants who saw both Category A and Category B objects from the beginning of the experiment were more likely to see the differences in feature values as representing the variability within a single group as opposed to representing two different groups. If, however, the first few encounters consist of a highly homogeneous set of objects, a new object with different feature values is seen as different from what has gone before and the person invents a new category to represent the new set of feature values. Clapper and Bower termed this the pattern-sequence effect. When participants in Clapper and Bower (1994, 2002) saw only Category A items for the first 12 or 16 trials before seeing the feature values associated with Category B items, they were more likely to create a new category to account for the new and different features presented rather than expanding the current category to include the new feature values.

Explaining the Pattern-Sequence Effect. Clapper and Bower (2002) explained the pattern-sequence effect with a modified version of Anderson's (1990, 1991) rational model that reflected imperfect memory.

The rational model specifies how categories are formed based on the premise that categorization is an adaptive means of relating to the world and capturing regularities in the environment. Clapper and Bower (2002) adapted Anderson's (1990, 1991) model to reflect imperfect memory. The resulting model depends on the following parameters; the

number of prior instances experienced (*n*), the prior belief that a specific value or set of values will occur ( ), the perceived likelihood that the instances belong to multiple groups (the coupling parameter; *c*), and the amount of prior information forgotten or remembered (the memory parameter;). The researcher controls *n* via the number of training instances and is determined primarily by the amount of noise in the feature values; the more consistent the feature values for a category, the higher the value of . Clapper and Bower (2002) focused on the relationship between the coupling parameter, *c*, and the (added) memory parameter, .

The expectation of multiple categories is captured by *c*, which represents the probability of adding a new instance to an existing category. The lower the value of *c*, the more likely the person will form a new category. The value of *c* is primarily determined by experimental instructions. Low values of *c* occur when participants are told to sort the items into categories (e.g., the Wisconsin Card sorting task) or are otherwise informed that the items belong to categories (e.g., prototype distortion tasks). Higher values of *c* occur when participants have not been told to look for and/or use categories and/or their natural inclination is not to look for a pattern of organization within the items (as in unsupervised category structure learning).

The parameter represents the accuracy of memory for the features of prior instances. The higher the value of , the more accurately the feature values of previously seen exemplars are recalled, and the more prior experience affects subsequent category assignment. In other words, the more accurately that the features of previous instances can be recalled and considered, the more those prior instances will affect the

assignment of a new instance to either an existing category or result in the formation of a new category.

Clapper and Bower (2002) demonstrated that sufficiently high values of *c* and low values of would explain the pattern-sequence effect because when items from the two categories were intermixed, people remembered too little of the previous items to differentiate the two category members, and combined them into the same category (i.e., people were prevented from forming separate categories). On the other hand, when several items from one category were presented in a row, participants were able to form a representation adequate for discriminating a new item. The present study is an empirical extension of the logic that low expectations of multiple categories coupled with poor memory inhibit the ability to acquire the structures of multiple categories. This requires operationalizing *c* and . I propose that *c* can be set through experimental design and the instructions to participants and that the value of depends on WM capacity and its contribution can be tested by comparing high and low WM span participants under the same learning conditions.

Adjusting *c* is difficult. Gross distinctions in *c* can be made through experimental instructions and methods. Telling participants to look for categories, or to look for how the items are similar or go together is a crude way of setting low values of *c* and are routinely employed in tests of supervised category learning (both classification and category structure learning). Hiding the experimental objectives behind instructions to learn the items or to evaluate them on some other, non-categorical, dimension leaves the value of *c* to be determined by individual variability. Fine distinctions are not possible

via instructions given to participants. It is not possible to tell people to only look for part of the category structure and ensure that they somehow limit themselves to only discovering part of it. A hint that structure exists is an open invitation to uncover the entire structure (Taraban & Hayes, 2001).

The SUSTAIN model (Love et al., 2004, Gureckis & Love, 2003a, 2003b) and Anderson's (1990, 1991) rational model both rely on incremental complexity to explain both inference and classification learning. There are several important similarities and differences between the two models, though both models provide a coherent means of interpreting the pattern-sequence effect as well as supervised classification and category structure learning.

Like Clapper and Bower's (2002) adaptation of the rational model, SUSTAIN adds representational complexity based on surprise; the model notices that the current item is not like the previous item and adds a new cluster to represent the new information. This illustrates one of the main differences between the two models. In the rational model, additional complexity is modeled in the form of additional categories, the formation of which is determined primarily by the coupling parameter (*c*; see above for a more complete description). When the rational model encounters a surprising item, one significantly different from its encoded categories, it creates a new representational cluster around the discrepant item. In SUSTAIN, additional complexity is represented by additional clusters, which can represent anything from individual exemplars to entire categories (i.e., prototypes). When SUSTAIN encounters a new item, it either joins an existing cluster or, if it is a surprising item (one significantly different from the current

set of applicable clusters), it recruits a new cluster to represent the new item. Each category is then represented by one or more clusters. For example, SUSTAIN might represent fish in one cluster and four-legged mammals in another. When it encounters a whale, it may initially place it in the fish cluster. If SUSTAIN learns (by external feedback, for example), that whales are mammals, then it will recruit an additional cluster centered on the whale exemplar to represent the added complexity.

SUSTAIN does not require external feedback to recruit additional clusters. Cluster recruitment depends on a parameter that represents the threshold for dissimilarity permissible before items recruit a new cluster. Indeed, additional clusters may be recruited even if they are not required to classify the stimuli under supervised conditions, but violate a rule or regularity in the stimuli themselves (Gureckis & Love, 2003b). The lower the threshold parameter, , the more easily a new cluster is recruited. In modeling the results of Love's (2002) comparison of supervised classification, unsupervised intentional (i.e., classification), and unsupervised incidental (i.e., category structure) learning, Gureckis and Love (2003a) found that the only parameter that needed to be altered to account for the difference in performance in the unsupervised classification and the unsupervised category structure learning conditions was .

<u>The Role of WM.</u> WM may provide an explanation for the pattern-sequence effects by operationalizing the limits of memory and awareness upon which the pattern-sequence effect relies according to both the rational model as defined by Clapper and Bower (2002) and SUSTAIN. WM was first differentiated from short-term memory by Baddeley and Hitch (1974) who added a processing dimension to the storage component

that was short-term memory. An explosion of research has subsequently linked WM to many cognitive processes relevant to unsupervised category learning including reasoning (Kyllonen & Christal, 1990), problem solving (Hambrick & Engle, 2003), learning (Baddeley, 1986), memory (Baddeley, 1986; Baddeley & Hitch, 1974; Cowan, 1995), comprehension (Daneman & Carpenter, 1980; Engle, Carullo, & Collins, 1992; Just & Carpenter, 1992), attention (Bleckley, Durso, Crutchfield, Engle, & Khanna, 2003; Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001), and awareness (Baddeley & Hitch, 1974; for overviews & reviews, see Baddeley, 1986; Cowan, 1995; Miyake & Shah, 1999).

For Billman and colleagues (Billman & Heit, 1988; Kersten & Billman, 1997), the process of unsupervised category structure learning occurs consciously. It is a rule-based online process that proposes that participants intentionally formulate a hypothesis, retain that hypothesis, and compare it against a new instance. All of these processes are WM processes. Furthermore, hypothesis testing is carried out by a focused sampling process that considers only a limited number of relationships due to limited learning resources (Billman & Heit, 1988). Neuropsychological studies of supervised category learning have linked explicit rule-based category learning to WM functioning (Ashby & Maddox, 2005; Ashby et al., 1998; Ashby & Waldron, 2000; Tracy et al., 2003; but see Aizenstein et al., 2000 for a different pattern of results). If continuous intentional hypothesis testing is behind the pattern-sequence effect, then WM is certainly involved as the locus of the hypothesis generation and testing process.

Clapper and Bower (2002) explained the pattern-sequence effect as the result of surprising features causing the creation of a new category. The most neurologically plausible locus of this process is the WM system. In other words, new categories are formed when the new exemplar in WM deviates significantly from the prior instances also held in WM and the discrepancy itself enters awareness (i.e., WM). The pattern-sequence effect depends on the nature of early exposure to exemplars; the presence of correlations among features alone is not sufficient. Early exposure to one category only sets the range of acceptable variation. When a new item appears that differs significantly from the previously encountered exemplars, a new category is spontaneously created if the discrepancy is noticed (i.e., in the rational model, if it is large enough). This is significantly different from Billman's (Billman & Heit, 1988; Billman & Knutson, 1996; Kerstan & Billman, 1997) proposal of how category knowledge is assembled in unsupervised learning. According to Billman and colleagues, the process is always explicit; an attempt on the part of the learner to tease apart the relationships via conscious effort. According to Clapper and Bower, the process doesn't have to start with, or even involve, conscious hypothesis testing. In terms of Anderson's (1990, 1991) rational model, awareness of the categories in unsupervised learning (with high values of *c*) is a byproduct of category learning, not the cause of category learning. Participants notice that the new things are not like the old things; they do not start by thinking "I have to look for how this one is different from the others."

There are several compelling reasons to assume that WM is essential for unsupervised category structure learning, both as the locus of in Clapper and Bower's (2002) implementation of the rational model and as the locus of in SUSTAIN.

1. WM is equated with awareness. To be aware of something means that it currently resides in WM (Cowan, 1995; Engle, 2002).
2. The amount of processing resources is limited, which limits the number of comparisons that can be made (Baddeley, 1986; Cowan, 1995). Spalding and Ross (1994) concluded that participants form abstractions from comparing instances, and that more features resulted in a narrowing of attention to a limited number of features as the basis of comparison. Abstractions are the basis for the formation of prototypes, which in turn are the basis of category structure learning (Yamauchi et al., 2002), and a critical component of SUSTAIN.
3. WM is associated with the control of attention (Bleckley et al., 2003; Engle, 2002; Engle & Kane, 2004; Engle, Kane, & Tuholsky, 1999; Kane et al., 2001). Brooks, Kempe, & Sionov (2006) found that executive attention was associated with learning syntactic rules in second language learning. Selective attention to stimulus features has an affect on category learning (Nosofsky, 1986; Nosofsky et al., 1994; Spalding & Ross, 1994). Several models of category learning posit an attentional mechanism that focuses on relevant feature dimensions (e.g., Kruschke, 1992; Love et al., 2004).
4. The number of instances or other pieces of information to be stored is limited. This means that a person can only have a very limited number of features,

relationships, or instances in awareness (Baddeley, 1986; Cowan, 1995). Clapper and Bower's (2002) describes the probability of recalling a given feature of a prior instances for comparison in the rational model without specifying whether the retrieval failure is due to loss from long-term memory or within WM (see 6, below).

5. WM places very real limits on the number of instances that can be sampled from the environment (Estes, 1986; Kareev, 1995, 2000; Kareev, Lieberman, & Lev, 1997). Kareev (1995, 2000; Kareev et al., 1997) found that the limited size of WM inflates the perceived size of correlations, leading to a greater estimation of relationships in the real world than actually exist and earlier detection of reliable relationships. This has the advantage of making actual correlations more salient than they really are, meaning that people are more likely to discover them (and discover them early), even if they over-emphasize the magnitude of those relationships.

6. Retrieval of instances is not always automatic and effortless. Bunting & Cowan (2005) showed that recalling the correct feature values in the face of interference or competition depends on WM resources (see also Conway & Engle, 1994). Accurate retrieval of feature information is another possible source of the parameter (see 4, above).

7. WM is implicated in resolving or preventing interference (Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999; Jha, Fabian, & Aguirre, 2004; Li, 1999), an

essential task for preserving the assignment of features to items and items to categories.

8. Wagar and Dixon (2005) demonstrated that knowledge of relevant category dimensions affects the encoding of new instances by directing the processing of the relevant features of those instances within WM.

9. Elio and Anderson (1981) argued that category generalizations are formed more easily when their instances are "simultaneously available in WM for patterns" (p. 402), yet knowledge of those generalizations does not have to be explicit or intentional.

10. McElree (1998) found a retrieval advantage for items belonging to the previously studied category in a list of categorically grouped items. For example, in a list containing three bird names followed by three flowers, followed by three cars, and so on, there would be a retrieval advantage for all three members of the car category. McElree concluded that the retrieval advantage was because the material still possessed residual activation making it more accessible. McElree concluded that the category representation was activated, not just the individual items. These results implicate WM in the retrieval of categorical information because members of the same category are more reliably retrieved if the category itself was recently active.

11. Rule-based classification learning is impaired under dual-task conditions. Waldron & Ashby (2001) demonstrated that a concurrent task delayed the time to acquire a simple verbalizable rule. Zeithamova and Ashby (2006) found that a

secondary task caused significant performance decrements when participants attempted to learn both unidimensional as well as conjunctive rules that defined categories. If participants are generating and testing hypotheses (cf., Billman & Knutson, 1996), those hypotheses are rule-based, and WM will be involved (Ashby et al., 1998).

12. Tracy et al. (2003) found prefrontal activation for both single attribute and family resemblance (i.e., prototype) classification using functional magnetic resonance imaging, indicating WM involvement in both types of classifications (but see Aizenstein et al., 2000, for different results on a prototype distortion task).

13. Several studies have found that clinical population with known impaired executive (i.e., WM) function were also impaired in rule-based supervised category learning (see also Filoteo, Maddox, Ing, Zizak, & Song, 2005; Price, 2005, 2006).

In all, any use of awareness or explicit processes depends on WM resources and WM has been shown to perform many of the functions attributed to in Clapper and Bower's (2002) implementation of Anderson's (1990, 1991) rational model and in SUSTAIN (Gureckis & Love, 2004).

Clapper and Bower (1994, 2002) demonstrated the pattern-sequence effect, manipulating presentation order, was sufficient to affect category invention by hypothesizing sufficiently low expectations of multiple categories combined with sufficiently poor memory. Extending the logic of Clapper and Bower (2002), persons with sufficiently low WM resources (i.e., low values of) should be impaired in learning

the category structure and may never form multiple categories when combined with low expectations of multiple categories (i.e., high values of *c*) because they lack the necessary ability to hold the current instance in memory, find previous examples (sorting those from the same category from those from another category), hold those previous examples in memory, and make comparisons to determine category similarity or difference, and form abstractions of categories. Therefore, we should observe an interaction of WM and the pattern-sequence effect; those with high WM capacity should outperform those with low WM capacity under conditions with high values of *c* (unsupervised category learning), learning more of the category structure when the stimuli are presented in a blocked sequence. Low WM span participants shouldn't learn the categories when they are presented in an intermixed sequence because they lack sufficient memory resources. It is possible that high WM capacity will result in significant learning when items are presented in an intermixed sequence. The exact nature of the interaction of WM and the pattern-sequence effect depends on the relative difficulty of the learning task and will be described in greater detail in Chapter III.

The SUSTAIN model also accounts for the pattern-sequence effect and relies on WM in a similar manner. SUSTAIN has two parameters that are theoretically consistent with WM function, and the attentional parameter. In addition to setting the tolerance for dissimilarity before a new cluster is created, Love and Gureckis (2004) described as the ability to bind clusters together and describe it as SUSTAIN's memory parameter (Gureckis & Love, 2004). The ability to bind features into units depends on healthy functioning of the hippocampus (Gluck & Myers, 2001). The hippocampus is part of the

WM system and is credited with the formation of new memories under the control of attention (Cowan, 1995). This is similar to Hummel and Holyoak's (1997, 2003) conception of the role of WM in analogy and schema use and formation. According to Hummel and Holyoak (1997, 2003), one of the main functions of WM is to bind elements together into coherent units, as in the formation of schemas. Furthermore, the capacity of WM is limited to the number of units that can be simultaneously active and mutually out of phase, where being in phase means that the elements are part of the same chunk, and being out of phase means that the elements belong to different chunks. Comparisons are made between the propositions contained in the current item with those stored as part of a known schema. Putting this in SUSTAIN's terms, WM is required for the formation of clusters, and limits the number of clusters that can be simultaneously available for consideration by representing the number of clusters that can be concurrently active. In their quantitative modeling of Love's (2002) replication of Shepard et al., (1961), Gureckis and Love (2003a) found that the only difference between unsupervised intentional (classification) learning and unsupervised incidental (category structure) learning, was in the magnitude of the threshold parameter. The magnitude of the parameter was lower in unsupervised category structure learning, which corresponds to a greater tendency to see things as the same. This suggests two possibilities. The first is that the task of learning within-category structure without external feedback sets a lower threshold for similarity, biasing the category learner to form fewer categories. Extending this logic, a second possibility is that people with very low WM capacity (i.e., very low values of) will have difficulty recruiting new clusters to represent different groups.

Though the mechanism is slightly different from the rational model, the prediction is the same: people with low WM capacity will have more difficulty differentiating the items and forming multiple groups under unsupervised category structure learning conditions than people with high WM capacity.

The other WM dependent parameter in the SUSTAIN model is the attentional parameter that functions to direct the system to emphasize (i.e., selectively attend to) more informative stimulus dimensions more heavily. For example, when faced with learning the difference between dogs and cats, an irrelevant dimension might be the presence of a tail, whereas a relevant dimension would be the sound the animal makes. The attentional parameter is responsible for making inputs on the sound feature more salient than inputs on the fur feature. This attentional parameter is associated with frontal lobe functioning (Sakamoto & Love, 2004), a primary area of executive control in WM. Clapper and Bower (2002) concluded that participants learned the category structure because they were controlling their attention, directing it to the relevant stimulus dimensions (in this case, the idiosyncratic dimensions). The attentional parameter is not important in the present study, as there are no idiosyncratic dimensions; all features are equally predictive of category membership and are perfectly reliable.

In sum, both of the models that best explain the pattern-sequence effect in unsupervised category structure learning, the rational model and SUSTAIN, have significant overlap with WM function. Especially relevant is the overlap of WM with critical parameters demonstrated for unsupervised category structure learning; the memory parameter, , in Clapper and Bower's (2002) implementation of the rational

model (Anderson, 1990, 1991), and the threshold parameter, , in SUSTAIN (Love et al., 2004). The present work explores the role of WM in unsupervised category structure learning as an attempt to link the theorized functions of WM and model-implied differences with behavioral data from healthy adults. To date, no studies have compared differences in WM on unsupervised category structure learning or attempted to operationalize or by examining differences in WM capacity in healthy adults.

While the present study hypothesizes that higher WM resources will provide a learning advantage in unsupervised category structure learning, the alternative is that the pattern-sequence effect is a byproduct of implicit processing, and will be unaffected by differences in WM capacity. According to this hypothesis, WM will affect the rate of item learning and memory, but will not result in greater category knowledge (Knowlton & Squire, 1993). High WM span participants should perform better on the learning items that they receive feedback on, but should not enjoy any advantage over low WM span participants on the generalization tests of category knowledge using items they have not received feedback on.

APPENDIX E

PILOT STUDY 1

The goal of the present study was to attempt to replicate and extend the pattern-sequence effect using stimuli that contained only one category feature value at a time for categories defined by fewer features than Clapper and Bower (1994, 2002) used. A greater number of intercorrelated features has been shown to aid learning the category structure (Billman & Knutson, 1996; Hoffman & Murphy, 2006). Using fewer features reduced the amount of correlational structure in the stimuli and prevented participants from controlling which feature to view (participants could only view the features in a fixed sequence, the order being determined at random). If the pattern-sequence effect results from unexpected feature values, then the results would show a learning benefit for seeing features of one category exclusively for the first several items (the number being chosen based on the work of Clapper and Bower, 1994, 2002) before seeing other values for the feature (i.e., the second category).

Pilot study 1 combined the category induction methodology of Taraban and Hayes (2000, 2001) and the pattern-sequence effect of Clapper and Bower (1994, 2002) to investigate the effects of initial exemplar exposure on unsupervised category learning when participants view one feature of an item at a time (Taraban & Hayes, 2000, 2001). The three experimental conditions were adapted from Clapper and Bower (1994, 2002) and included a control condition with randomized feature values (the control condition), and two experimental conditions, the contrast and the mixed conditions. The contrast and

mixed conditions contained items that belonged to two categories based on their feature values. In the mixed condition, the items from the two categories were presented pseudorandomly, such that no more than three exemplars from the same category appeared in a row. The contrast condition presented only items from one category for the first 12 trials before the presentation of the first item from the second category. Clapper and Bower (1994, 2002) found that this blocking manipulation led to significantly more knowledge of the category structure in the contrast condition than in the mixed condition (the pattern-sequence effect) and the control condition.

Category learning would be evident from four effects in the present study. First, participants exposed to only one category in the pretraining phase (the contrast condition) should exhibit above chance accuracy on item-feature pairs they have not seen before, but were composed of known items and known features (base-transfer items), which should be higher than participants in the mixed and control conditions. If the participants have acquired the category structure, then they should perform above chance on those items that conform to the pattern, even if they have not seen that particular feature of an item before. Second, participants in the contrast condition should exhibit above chance performance on the final block of new items during a transfer phase containing novel items, and their performance should be higher than those in the mixed and control conditions. If the category structure is learned in the pretraining and learning phases, then learning new items should be facilitated when the category structure of those new items is the same as the old items. Third, reaction time (RT) for the base transfer items should be shorter for participants in the contrast condition than for participants in the

mixed and control conditions. If participants have acquired the category structure, then they will require less time to evaluate a response to items they have not seen before but conform to a learned structure. Fourth, responses on a retrospective post-experimental questionnaire will demonstrate greater category knowledge for participants in the contrast condition than for participants in the mixed and control conditions. The category invention hypothesis proposes that the outcome of category creation is explicit and accessible to awareness. Their performance on the questionnaire questions should show greater category awareness for those with greater category knowledge.

No specific hypotheses are made regarding the performance of participants in the mixed condition because their performance will depend on the difficulty of the task. Clapper and Bower (1994) found some evidence of unsupervised category learning in participants in the mixed condition, though less than participants in the contrast condition, but Clapper and Bower (2002) did not. The easier the task, the more participants in the mixed condition will perform like participants in the contrast condition and may differ significantly from participants in the control condition. The harder the task, the more participants in the mixed condition will perform like participants in the control condition, and may not perform any better than chance. Under no circumstances should the mixed condition participants perform better than participants in the contrast condition. As the control condition has the lowest systematicity (i.e., intercorrelated features) in its stimuli set, participants in the control condition should never perform better than participants in the mixed or contrast conditions regardless of the presentation order manipulation. Better performance in the mixed and contrast conditions than the

control condition on the base-learning and base-transfer items would indicate that correlated feature values aid unsupervised category learning (Billman & Knutson, 1996). Better performance in the contrast than the mixed and control conditions would support the pattern-sequence effect (Clapper & Bower, 1994, 2002).

Methods

The present study conceptually replicated Clapper and Bower (2002) using a serial presentation format that displayed only a single feature of an exemplar at a time. Participants in Clapper and Bower (2002) had access to all 12 features of items, and could freely select which feature to view. Nine of the features were perfectly informative of category membership.

The present study used a more restrictive serial presentation format than Clapper and Bower (2002) and added a test of item knowledge that included unseen item features that tested category knowledge, not just rote memorization. The items (muscle names, e.g., *Anconius*) had three features (position, movement, function), only two of which appeared in the pretraining or learning phases for each item. The remaining feature was reserved for a test of category learning. Participants viewed each muscle-feature pair, one at a time, and had to enter a response before the next item was presented. For example, participants would see the sequence of *Anconius, position is:* followed by *Anconius, function is:*, before seeing the next muscle (*Anconius, movement is:* would only appear on the test).

Just like Clapper and Bower (2002), participants saw all of the features for an item before moving onto the next item, but unlike Clapper and Bower, they could not revisit an item. The test of category knowledge was a recognition test in both Clapper and Bower (2002) and the present experiment. Participants in both studies saw the name of the item and the feature being queried and selected a feature value from amongst those presented.

The purpose of this experiment was to replicate the pattern of results observed by Clapper and Bower using three conditions; contrast, mixed, and control. Participants in the contrast condition were expected to exhibit significantly more category learning than participants in the mixed and control conditions. Category learning was measured by accuracy and RT for items during the test phase. Specifically, category learning would be inferred from above chance accuracy on the features that were not seen during learning (new items). RTs for new items should be longer than items seen during learning, but should be significantly faster if category learning has occurred. Participants in the contrast condition should have higher accuracy on the new items than participants in the mixed and control conditions, and should have shorter RTs for new items as well. Participants in the mixed condition were expected to be slower and less accurate than participants in the contrast condition as Clapper and Bower (1994) found that participants in their mixed presentation condition did acquire the category structure, but at a significantly slower pace than participants in the contrast condition. The differences in presentation methodology could prevent participants in the mixed condition from acquiring the category structure, in which case they will not differ significantly from

control participants. However, it is also possible that three repetitions of the learning materials is sufficient for participants exposed to mixed presentation of structured stimuli to acquire the category structure, in which case their accuracy on the new items would be higher than controls and their RT would be shorter than controls for the new items.

Participants

Participants were 41 volunteers who participated for course credit in a General Psychology course. They were randomly assigned to one of three experimental conditions: contrast (n = 13), mixed (n = 14), and control (n = 14).

Materials

Stimuli. The stimuli consisted of 36 Latin muscle names along with three features for each of those muscles (position, function, movement). Each feature had two possible values (position: posterior, anterior; function: abduction, flexion; movement: extension, contraction). The Latin names are actual human muscle names paired with feature values so that the two categories were as physically similar as possible. The Latin muscle names were matched on average word length (8.47 overall, 8.56 for Category A, 8.39 for Category B), number of syllables (3.58 overall, 3.61 for Category A, 3.56 for Category B), first letter, last letter, and last syllable. See Table 19, 17, and 18 for the complete stimuli set.

All three conditions used the same 36 Latin muscle names and six feature values, but differed in how those feature values were combined and the order in which they were

presented. In the contrast and mixed conditions, a tacit categorical organization was formed by which feature values were associated with a muscle. In these two conditions, feature values were perfectly correlated and completely described all members of the category. Category A muscles were always posterior position, abduction function, and extension movement. Category B muscles were always anterior position, flexion function, and contraction movement. Participants were not shown the category labels. The feature values in the control condition were intentionally uncorrelated. There were eight possible feature value combinations (2 positions X 2 functions X 2 movements). These were assigned to items to minimize the amount of feature correlation. As there were 8 feature value combinations and the learning stimuli required stimuli to be multiples of six, it was not possible to have perfectly uncorrelated learning items.

Procedure

All questions were delivered via computer, presented as an item and a feature of that item; *muscle, feature is:* (e.g., *Anconeus, position is:*). The participant then selected from among two feature values displayed on the computer screen which one was correct for that feature (*posterior, anterior*) using the "f" and "j" keys, which had been relabeled "1" and "2," respectively. After the participant entered a response, the computer provided feedback, indicating whether the answer selected by the participant was correct or incorrect, and the correct answer was displayed. This feedback remained on the screen until the participant pressed the spacebar to initiate the next trial. The computer recorded participant accuracy and latency.

The experimenter began the experiment by explaining the keyboard layout and instructed participants to place their left index finger on the "1" key, and their right index finger on the "2" key. The experimenter then read the preliminary instructions to participants and answered questions.

The experiment consisted of five self-paced phases: pretraining, learning, test, transfer, and post-experimental questionnaire. The item order, feature order, and answer option order were all randomly determined within each phase.

Pretraining. For all conditions, the pretraining phase consisted of 24 individual items (2 features for each of 12 muscles). The number of pretraining items was selected based on the procedures employed by Clapper and Bower (1994, 2002) and by the results from Clapper and Bower's (1994) investigation of the impact of the amount of pretraining required for the pattern-sequence effect. Clapper and Bower (1994) used 16 pretraining items and Clapper and Bower (2002) used 12 pretraining items. The results of Clapper and Bower (1994; Experiment 3) indicated that there was a significant unsupervised learning advantage for a pretraining set of 12 items over sets of 8 or 4 pretraining items. The stimulus set size of 12 items used here contained fewer features to learn, but the results of Clapper and Bower (1994) indicated that more pretraining items lead to greater category learning. Feedback was delivered after each item indicating whether the answer given was correct or incorrect, and the computer presented the correct selection. In the contrast condition, twelve muscles from category A, each with two features, were presented sequentially in a random blocked design. The muscles were randomly sequenced, but presented with each of their two features in random order

before presenting the next muscle. The mixed condition was identical to the contrast condition except that the pretraining phase consisted of six category A muscles and six category B muscles. All pretraining items had two features, and were presented sequentially in a pseudo-random order as follows: the first item belonged to category A, and no more than two muscles from the same category were presented sequentially (still blocked based on muscle name). In the control condition, the stimuli were the same 12 muscle names as the mixed condition, but with uncorrelated feature values.

Learning. The learning phase consisted of 108 trials (three presentations of two features for each of 18 muscles; 9 from category A and 9 from category B). The computer provided feedback for all responses in this phase. In the contrast and mixed conditions, nine new muscles of each category were presented in a blocked pseudo-random order. The two properties for each muscle were presented sequentially in random order, with the order of muscles randomly determined except that no more than three muscles from a single category appeared in a row. The learning phase for the control condition was the same as the contrast and mixed conditions, but the feature values were distributed in an uncorrelated manner. Three blocks of the learning items was used because there were not enough features to adequately test participant's knowledge immediately after the presentation of each item as in Clapper and Bower (2002). Instead, category knowledge was tested during the testing phase, so participants had to learn not only which feature values went together, but also which muscle names went with which set of features.

Testing. All of the items from the pretraining and learning phases were presented once in blocked pseudo-random order. Each muscle was shown with all three of its features sequentially in random order; two features were those shown previously in the pretraining or learning phase (old items), the remaining feature was seen for the first time (new items). The order of muscles was determined randomly and feedback was not provided. On-screen instructions stressed speed and accuracy and informed participants that they would not receive feedback. Accuracy and RT were recorded.

Transfer. Participants were informed that they would see new items and they should try and learn the new items. They saw all three features for six new muscles (3 from each category) three times (54 trials). The computer provided feedback for all responses in this phase.

Post-Experimental Questionnaire. After completing the transfer task, participants were given a printed packet containing five questions that assessed participant awareness of the category structure used in the experiment. The first two questions were printed on the same page and were very broad ("How did you go about learning the muscles in the first part? Try to include all the methods that you used." "How did you go about learning the *new muscles* at the end of the experiment? If you used any of the same methods, please indicate which ones you used."). Each new question was presented on a new page and participants were required to show their answer to the experimenter before turning the page. The experimenter ensured that all questions were answered and requested clarification if the participant's handwriting was illegible or if the meaning was ambiguous.

Results

Hypotheses regarding above chance performance were assessed using single sample t-tests. Hypotheses regarding between-group differences were tested using one-way analyses of variances (ANOVAs). A Tukey correction was applied to all follow-up comparisons. Means and standard deviations for the behavioral data are presented in Table 22.

Behavioral Data

An analysis of the final block of learning trials and the test of old items was conducted to measure learning of the items seen during the learning phase (base-learning items). The results painted a mixed picture of learning. To ascertain whether any learning occurred over the three blocks, a 3 (condition: contrast, mixed, control) X 3 (learning block) mixed-design ANOVA with learning block as a within subjects variable revealed a significant main effect of learning block, F (2, 76) = 9.225, $p < .001$, indicating that accuracy improved from the first to the third block. The effect of condition and the interaction were not significant. A oneway ANOVA with condition as a between-subjects variable and accuracy on the third block of base-learning items approached significance, $F(2, 38) = 2.915$, $p = .066$. Single sample t-tests revealed that participants in the mixed and contrast conditions performed above chance on the final block of learning trials (chance = 18 out of 36 correct), t (13) = 2.50, $p < .05$ (mixed) and t (12) = 6.32, $p < .001$ (contrast), whereas participants in the control condition did not, t (13) = 1.59, *ns*. During the test phase the results were reversed. Participants in the

control condition scored significantly better than chance (chance = 18 out of 36 correct) on the base learning (old) items, t (13) = 2.40, $p < .05$, and the participants in the mixed and control conditions performed no different than chance, t (13) = 1.87, $p > .05$ (mixed) and t (12) = - 0.93, *ns* (contrast).

The first hypothesis, that participants in the contrast condition would perform better than chance and better than participants in the mixed and control conditions on the base-transfer items, was not supported. Participants in the contrast condition achieved a mean accuracy of 9.15 out of 18 (SD = 1.99), which was not significantly greater than chance (chance = 9 out of 18 correct), single sample t (12) = 0.28, *ns*. A one-way ANOVA confirmed that there were no differences between the experimental conditions in accuracy on the base-transfer items, F (2, 38) = 0.92, *ns*.

The second hypothesis, that participants in the contrast condition would perform better than chance and better than participants in the mixed and control conditions on the final block of transfer items, was partially supported. Participants in the contrast condition achieved a mean accuracy of 11.15 out of 18 (SD = 3.02), which was significantly different than chance (chance = 9 out of 18 correct), single sample t (12) = 2.57, $p < .05$. Participants in the mixed condition also performed significantly above chance, with a mean accuracy of 11.29 correct (SD = 3.47), single sample t (13) = 2.46, $p < .05$. Mean performance for participants in the control condition did not exceed chance (M = 9.79, SD 3.04), t (13) = 0.97, *ns*. A one-way ANOVA failed to find any significant differences between the experimental conditions on accuracy on the final block of the transfer items, F (2, 38) = .94, *ns*.

The third hypothesis, that participants in the contrast condition would display faster RT than participants in the control condition on the base transfer items, was not supported. Individual RTs were normalized using a log transformation. The transformed values were then submitted to a one-way ANOVA which revealed no significant differences in RT between any of the three conditions, F (2, 38) = .33, *ns*. Log transformed RT for the base-transfer items was compared to the log transformed RT for the base-learning items (the items seen during the learning phase). The times did not differ by a significant amount for any of the conditions; contrast condition, paired sample t (12) = .39, *ns*, mixed condition, paired sample t (13) = -1.48, *ns*, control condition, paired sample t (13) = -1.46, *ns*.

Questionnaire Data

The questionnaire data revealed no difference in category knowledge between groups. Explicit knowledge of some or all of the category structure was evidenced by approximately 32% of all participants regardless of experimental condition. Table 23 reports the number of no awareness, some awareness, and full awareness cases in each condition for each of the explicit awareness questions (Questions 1, 3, and 4). The low number of cases necessitated the use of Fisher Exact tests to determine if there were any differences in awareness between categories for questions 1, 3, and 4. This test is more robust when cell sizes and/or expected values are small (Conover, 1999). None of the groups differed significantly on any of these questions. The mean score for Question 2 for the contrast condition was 0.69 (SD = 1.49) out of 4, 1.00 (SD = 1.18) for the mixed

condition, and 1.21 (SD = 1.42) for the control condition. This mean difference was not reliable, t(15) = , p > . The high standard deviations for these means suggests that there was substantial individual variation. Question 2 was also evaluated by comparing the number of participants in each group who demonstrated full knowledge of the pattern versus those that did not. In the contrast condition, 2 of 13 participants achieved the maximum score (full knowledge of the category structure), which was the same number as the control condition (2 of 14), and only 1 of the 14 in the mixed condition. These differences are not reliable (Fisher Exact test, all $p > .50$).

Discussion

The results provide weak support for unsupervised category learning. There was a significant effect of learning block, and participants in the contrast and mixed conditions had accuracy rates that were significantly above chance on the final block of learning trials, but they then performed at chance on those same items in the test phase. Participants in all conditions proceeded from the learning phase to the test phase, pausing only long enough to read the on-screen instructions for the test. The most likely explanation is that the learning effects reported here reflect the fact that participants were learning the category structure but not which muscles belonged to which category. In the learning phase, they would use the feedback from the first item-feature pair to determine which category the item belonged to, allowing them to get the second item-feature pair correct. There was no feedback in the test phase for participants to use to determine which category the item belonged to, reducing them to chance performance.

Participants in the contrast and mixed conditions also performed better than chance on the final transfer block, while participants in the control condition performed at chance. One possible interpretation is that the small size (18 total pairs: 6 items, each with 3 features) and isolation from the other learning materials allowed participants in the contrast and mixed conditions to memorize the set of transfer items without the benefit of category knowledge. This could explain the above chance performance observed without requiring any knowledge of the category structure. That participants in the control condition performed at chance speaks against this interpretation. If the small size and distinctiveness of the set of new transfer items made it easy for participants to memorize the entire set of 18 item-feature pairs without the benefit of category knowledge, then participants in the control condition should have enjoyed the same advantage as participants in the contrast and mixed conditions. Rather, it seems likely that this effect indicates that correlated feature values facilitates learning (Billman & Knutson, 1996). Additional support for this hypothesis comes from the difference in accuracy on the third block of learning trials approached significance, with the contrast and mixed condition participants scoring above chance, and the control condition participants scoring at chance. This effect is in the direction that supports the utility of correlated feature dimensions (Billman & Knutson, 1996).

The remaining behavioral results do not support the conclusion that unsupervised category learning occurred. Participants in the contrast and mixed conditions performed no better than chance on the base transfer items. Their RT for these items was no different than participants in the control condition, which also fails to support the

acquisition of category knowledge. Category knowledge should result in faster RT for the base transfer items, with greater gains in RT with stronger category knowledge. It is possible that participants who have acquired a strong category representation would display no difference in RT for the base learning and the base transfer items. This null effect is only meaningful if there is an interaction between experimental condition and item type (base-learning vs. base-transfer). Category learning can be inferred from similar RTs on the base-learning and the base-transfer items only if RTs for participants who have not acquired the category structure exhibit slower RT for the base-transfer items than the base-learning items. Participants in all three conditions displayed no RT differences between the base-learning and the base-transfer items. This cannot be interpreted as support for unsupervised category learning because participants did not achieve above chance accuracy on those same items and there was no interaction between experimental condition and item type.

The pattern of results is consistent with a failure to learn the learning items. Overall, participants in all three conditions failed to learn the items presented during the learning phase. Even though the performance of participants in the mixed and control conditions was above chance during the last block of learning trials, they performed no better than chance during the test on the same items. That participants in the control condition demonstrated the opposite pattern further supports the conclusion that all participants had not learned the associations of the two features with each item during the learning phase. Further support for a lack of participant learning comes from the irrelevant responses given for Question 2 in the post-experimental questionnaire. Six of

41 participants (14.6%) did not know which features the feature values went with, or responded with values not used in Pilot Study 1. Learning the item-feature pairs may be a prerequisite for category learning. Failure to learn the features of the items prevents learning the similarities between features that defines the categories used in this study. If participants did not learn the muscles and the muscle-feature pairs that they saw repeated during the learning phase, if they could not tell them apart, then it may have interfered with acquiring the category structure. The stimuli were balanced between categories on first letter, last letter and number of syllables. Knowing that "anterior" and "contraction" went with Pectineus and "posterior" and "extension" with Peroneus would not help participants learn the categories if the participant could not tell the two muscles apart.

Knowing the category structure would have allowed participants to perform above chance on new items (transfer phase) or new features (test phase) as long as they knew which category the item belonged to. Their accuracy on the base transfer items was at chance, and their RTs suggest that forming an answer to unseen items (base-transfer items) was just as effortful as forming an answer to the seen items used in learning (base-learning items). While there is some evidence that participants learned the new items in the transfer phase, these results are problematic. The lack of a difference between the RT for the base-transfer and the base-learning items, coupled with the lack of significant learning of the base-learning items, indicates that participants did not engage the task and learn about the features of the muscles.

Above chance performance on the final block of transfer items may reflect some category learning. The use of a small set of new items at the end of the experiment gave

them several advantages that may have contributed to potential category learning. A set of six items is less susceptible to confusion than a set of 18 items and also enjoyed a faster rehearsal time because of the reduced number of items in each block, which is associated with better immediate performance (Pavlik & Anderson, 2005). The additional repetitions of the learning materials in the test phase combined with several blocks of the new items could have resulted in category learning during the transfer phase, rather than before it. Methodological changes were made in Pilot Study 2 to address the possibility that failure to replicate the pattern-sequence effect was due to a failure to learn the items and their features.

Table 19 Pilot Study 1 Contrast Condition Stimuli.

Name	Set	Position	Function	Movement
		Category A		
Adductor	Pretraining	Posterior *	Abduction	Extension
Medialis	Pretraining	Posterior *	Abduction	Extension
Pollicis	Pretraining	Posterior *	Abduction	Extension
Rosorius	Pretraining	Posterior *	Abduction	Extension
Sternohyoid	Pretraining	Posterior	Abduction *	Extension
Ulnaris	Pretraining	Posterior	Abduction *	Extension
Buccinator	Pretraining	Posterior	Abduction *	Extension
Cubitalis	Pretraining	Posterior	Abduction *	Extension
Femoris	Pretraining	Posterior	Abduction	Extension *
Hallucis	Pretraining	Posterior	Abduction	Extension *
Mylohyoid	Pretraining	Posterior	Abduction	Extension *
Splenius	Pretraining	Posterior	Abduction	Extension *
Brachialis	Learning	Posterior *	Abduction	Extension
Fibularis	Learning	Posterior *	Abduction	Extension
Lumbrical	Learning	Posterior *	Abduction	Extension
Omohyoid	Learning	Posterior	Abduction *	Extension
Peroneus	Learning	Posterior	Abduction *	Extension
Popliteus	Learning	Posterior	Abduction *	Extension

Table 19 Continued.

Name	Set	Position	Function	Movement
Serratus	Learning	Posterior	Abduction	Extension *
Supinator	Learning	Posterior	Abduction	Extension *
Tibialis	Learning	Posterior	Abduction	Extension *
Cocygeus	Transfer	Posterior	Abduction	Extension
Dorsalis	Transfer	Posterior	Abduction	Extension
Gemellus	Transfer	Posterior	Abduction	Extension
		Category B		
Gracilis	Learning	Anterior *	Flexion	Contraction
Levator	Learning	Anterior *	Flexion	Contraction
Obturator	Learning	Anterior *	Flexion	Contraction
Pectineus	Learning	Anterior	Flexion *	Contraction
Piriformis	Learning	Anterior	Flexion *	Contraction
Plantaris	Learning	Anterior	Flexion *	Contraction
Scalenus	Learning	Anterior	Flexion	Contraction *
Scapularis	Learning	Anterior	Flexion	Contraction *
Thyrohyoid	Learning	Anterior	Flexion	Contraction *
Anconeus	Transfer	Anterior	Flexion	Contraction
Capitis	Transfer	Anterior	Flexion	Contraction
Radialis	Transfer	Anterior	Flexion	Contraction

Note. Items marked with an asterisk were only presented on the test (base-transfer items).

Table 20 Pilot Study 1 Mixed Condition Stimuli.

Name	Set	Position	Function	Movement
		Category A		
Adductor	Pretraining	Posterior *	Abduction	Extension
Medialis	Pretraining	Posterior *	Abduction	Extension
Pollicis	Pretraining	Posterior	Abduction *	Extension
Rosorius	Pretraining	Posterior	Abduction *	Extension
Sternohyoid	Pretraining	Posterior	Abduction	Extension *
Ulnaris	Pretraining	Posterior	Abduction	Extension *
Brachialis	Learning	Posterior *	Abduction	Extension
Fibularis	Learning	Posterior *	Abduction	Extension
Lumbrical	Learning	Posterior *	Abduction	Extension
Omohyoid	Learning	Posterior	Abduction *	Extension
Peroneus	Learning	Posterior	Abduction *	Extension
Popliteus	Learning	Posterior	Abduction *	Extension
Serratus	Learning	Posterior	Abduction	Extension *
Supinator	Learning	Posterior	Abduction	Extension *
Tibialis	Learning	Posterior	Abduction	Extension *
Cocygeus	Transfer	Posterior	Abduction	Extension
Dorsalis	Transfer	Posterior	Abduction	Extension
Gemellus	Transfer	Posterior	Abduction	Extension

Table 20 Continued.

Name	Set	Position	Function	Movement
		Category B		
Buccinator	Pretraining	Anterior *	Flexion	Contraction
Cubitalis	Pretraining	Anterior *	Flexion	Contraction
Femoris	Pretraining	Anterior	Flexion *	Contraction
Hallucis	Pretraining	Anterior	Flexion *	Contraction
Mylohyoid	Pretraining	Anterior	Flexion	Contraction *
Splenius	Pretraining	Anterior	Flexion	Contraction *
Gracilis	Learning	Anterior *	Flexion	Contraction
Levator	Learning	Anterior *	Flexion	Contraction
Obturator	Learning	Anterior *	Flexion	Contraction
Pectineus	Learning	Anterior	Flexion *	Contraction
Piriformis	Learning	Anterior	Flexion *	Contraction
Plantaris	Learning	Anterior	Flexion *	Contraction
Scalenus	Learning	Anterior	Flexion	Contraction *
Scapularis	Learning	Anterior	Flexion	Contraction *
Thyrohyoid	Learning	Anterior	Flexion	Contraction *
Anconeus	Transfer	Anterior	Flexion	Contraction
Capitis	Transfer	Anterior	Flexion	Contraction
Radialis	Transfer	Anterior	Flexion	Contraction

Note. Items marked with an asterisk were only presented on the test (base-transfer items).

Table 21 Pilot Study 1 Control Condition Stimuli.

Name	Set	Position	Function	Movement
		Category A		
Adductor	Pretraining	Anterior *	Flexion	Extension
Medialis	Pretraining	Posterior *	Abduction	Contraction
Pollicis	Pretraining	Anterior *	Abduction	Contraction
Rosorius	Pretraining	Posterior *	Flexion	Extension
Sternohyoid	Pretraining	Posterior	Flexion *	Contraction
Ulnaris	Pretraining	Anterior	Abduction *	Extension
Brachialis	Learning	Posterior *	Abduction	Extension
Fibularis	Learning	Anterior *	Flexion	Contraction
Lumbrical	Learning	Posterior *	Abduction	Contraction
Omohyoid	Learning	Anterior	Abduction *	Extension
Peroneus	Learning	Posterior	Flexion *	Extension
Popliteus	Learning	Anterior	Abduction *	Contraction
Serratus	Learning	Posterior	Flexion	Contraction *
Supinator	Learning	Anterior	Flexion	Extension *
Tibialis	Learning	Posterior	Abduction	Extension *
Cocygeus	Transfer	Posterior	Abduction	Extension
Dorsalis	Transfer	Posterior	Abduction	Extension
Gemellus	Transfer	Posterior	Abduction	Extension

Table 21 Continued.

Name	Set	Position	Function	Movement
		Category B		
Buccinator	Pretraining	Anterior	Flexion *	Contraction
Cubitalis	Pretraining	Posterior	Abduction *	Extension
Femoris	Pretraining	Anterior	Abduction	Contraction *
Hallucis	Pretraining	Posterior	Flexion	Extension *
Mylohyoid	Pretraining	Posterior	Abduction	Contraction *
Splenius	Pretraining	Anterior	Flexion	Extension *
Gracilis	Learning	Anterior *	Flexion	Contraction
Levator	Learning	Anterior *	Flexion	Extension
Obturator	Learning	Posterior *	Abduction	Extension
Pectineus	Learning	Posterior	Abduction *	Contraction
Piriformis	Learning	Posterior	Flexion *	Extension
Plantaris	Learning	Anterior	Flexion *	Contraction
Scalenus	Learning	Anterior	Abduction	Extension *
Scapularis	Learning	Posterior	Flexion	Contraction *
Thyrohyoid	Learning	Anterior	Abduction	Contraction *
Anconeus	Transfer	Anterior	Flexion	Contraction
Capitis	Transfer	Anterior	Flexion	Contraction
Radialis	Transfer	Anterior	Flexion	Contraction

Note. Items marked with an asterisk were only presented on the test (base transfer items).

Table 22 Pilot Study 1 Means and Standard Deviations for Learning, Test, & Transfer Accuracy, and Test Latencies.

	Contrast		Mixed		Control	
	Mean	SD	Mean	SD	Mean	SD
Learning Trials Accuracy						
Learning Block 1	48.93%	9.11%	50.00%	12.00%	46.83%	6.96%
Learning Block 2	54.27%	5.51%	53.00%	8.00%	55.16%	8.00%
Learning Block 3	57.91%	4.51%	56.00%	9.00%	54.56%	10.76%
Test Trials Accuracy						
Base Learning	47.86%	8.27%	55.00%	10.00%	54.56%	7.11%
Base Transfer	50.85%	11.08%	46.00%	10.00%	51.19%	12.17%
Test Trials Latency						
Base Learning	2262.61	979.11	2328.08	869.9	2074.22	737.36
Base Transfer	2062.35	1002.42	2530.11	1077.22	2143.85	826.85
Transfer Trials Accuracy						
Transfer Block 1	52.56%	17.94%	54.00%	14.00%	51.98%	16.54%
Transfer Block 2	53.85%	15.28%	63.00%	15.00%	53.17%	17.26%
Transfer Block 3	61.97%	16.80%	63.00%	19.00%	54.37%	16.90%

Table 23 Pilot Study 1 Post Questionnaire Analysis: Questions 1, 3, 4.

	Question 1 Awareness			Question 3 Awareness			Question 4 Awareness		
Condition	None	Some	Full	None	Some	Full	None	Some	Full
Contrast	9	3	2	9	3	2	11	0	3
Mixed	11	2	1	8	4	2	11	2	1
Control	8	5	0	10	2	1	12	0	1
Total	28	10	3	27	9	5	34	2	5

Note. Question 1 asked participants to describe the strategies they used to learn the items. Question 3 asked participants to describe how, if at all, the plants or their features were similar. Question 4 informed participants that the items belonged to categories and asked them to describe how those categories were formed. See Appendix C for all of the post-experimental questions.

APPENDIX F

PILOT STUDY 2

Four changes were made to the method used in Pilot Study 1 in an effort to replicate the pattern-sequence effect of Clapper and Bower (1994, 2002). The changes were all intended to increase learning of the items and their features and to expand the range of the dependent measures to allow a greater difference between chance performance and knowledge of category structure.

First, examination of the behavioral data and the post-experimental questionnaires suggested that the materials were too difficult for participants. Specifically, several participants in Pilot Study 1 noted that the Latin names used were too difficult and that they focused on parts of the words, usually the final one or two letters, rather than learning the entire word and its associated characteristics (a useless strategy as the stimuli were balanced across groups on first letter, last letter, last syllable, number of syllables, and number of letters). Shorter Latin tree names were used in Pilot Study 2 in an effort to make the items easier to differentiate and learn yet remaining unfamiliar. Several participants in Pilot Study 1 reported trying to associate muscles into groups based on where they were in the body (anterior/posterior or upper/lower body). In order to facilitate the learning of the related features, more familiar and imaginable characteristics were used for the plants (e.g., leaf shape: pointed or rounded) than for the muscles (e.g., function: abduction or flexion). It was thought that concrete features like stem type, root

type, leaf shape, and flower type would be easier for participants to visualize than abstract muscle features like position, function, and movement.

Second, the learning task for Pilot Study 2 required participants to type in their response. Spending more time typing in responses should require participants to pay more attention to the feature values and the items they go with. It will also slow participants down, forcing them to spend more time on each item-feature pair. Previous research on category induction required participants to spend more time on the learning task by requiring them to type in the response rather than simply selecting from two options (Taraban & Hayes, 2000, 2001).

Third, a fourth feature was added for two reasons. First, the stimuli used by Clapper and Bower (2002) may have been learned so readily in part because nine correlated features produced high systematicity (Billman & Knutson, 1996). Second, the use of three features makes it statistically difficult to differentiate someone who is guessing on each feature for each item from someone who got the first feature wrong for an item (especially since the muscle names used in Pilot Study 1 were so difficult to remember), but then corrected him/herself and got the remaining two items correct. It seems unlikely that this occurred during the test phase of Pilot Study 1, but it is possible that the pattern-sequence effect was masked by the slight difference between chance performance and performance based on category knowledge during the final block of the transfer phase combined with the small sample size used in Pilot Study 1. The pattern-sequence effect depends on a learning difference between the contrast and mixed conditions. Both of these groups performed above chance on the final block of the

transfer phase, and participants in the control condition did not, but the ANOVA failed to find any difference between the three groups.

Finally, the number of pretraining and learning items was adjusted. Four items were added to the pretraining sequence and two items were removed from the learning sequence. This was done to ensure counterbalancing of the learning and base transfer items now that there were four instead of three features. Overall, the addition of a fourth feature and a net gain of two items meant that participants in Pilot Study 2 experienced 60 more trials than participants in Pilot Study 1.

Methods

Participants

Participants were 34 students (16 in the contrast condition, 18 in the mixed condition) from a General Psychology course and upper level psychology courses who participated for course credit. One extreme outlier was identified. Examination of the responses by a participant in the contrast condition who achieved 6.25% accuracy on the final block of learning trials revealed that this participant regularly responded (72.92% of responses) with feature values belonging to other features (e.g., when prompted to enter the root type, the participant responded with a feature value associated with the leaf or flower). This participant was excluded, resulting in a total of 33 participants. None of the participants in Pilot Study 2 had taken part in Pilot Study 1.

Materials

Stimuli. The stimuli for Pilot Study 2 were selected to be simpler and easier to pronounce and differentiate than those used in Pilot Study 1. Latin plant names were used. They were shorter (5.61 letters, 2.41 syllables) than the stimuli in Pilot Study 1 (8.34 letters, 3.55 syllables). The features and the feature values are shown in Table 24 and Table 25, and were selected to be easier to learn and more familiar than those used in Pilot Study 1. As in Pilot Study 1, one of the features was reserved as a test of category knowledge and was not shown during the pretraining and learning phases (base-transfer).

Procedure

The procedure for Pilot Study 2 was the same as that for Pilot Study 1 except as follows. Participants typed in their responses for every part of Pilot Study 2 except for the testing phase. The testing phase displayed the two answer options just as in Pilot Study 1 so that RT could be accurately measured. The number of pretraining and learning items differed from Pilot Study 1. There were 48 pretraining trials (3 features for each of 16 items) and 144 learning trials (3 blocks of 16 learning items, each with 3 features). The procedure for the test phase and transfer phase were identical to Pilot Study 1, except for the number of items. There were 64 trials in the test phase (48 old items, 16 new items) and 72 trials in the transfer phase (3 blocks of 6 items, each with 4 features).

Results

As in Pilot Study 1, single sample t-tests were used to test above chance performance and group comparisons were assessed by one-way ANOVAs. Means and standard deviations for the behavioral data are presented in Table 26.

Behavioral Data

The main goal of Pilot Study 2 was to improve learning of the experimental items. Learning was assessed separately over the course of the three learning blocks and over the course of the three transfer blocks. A 3 (Learning block) X 2 (condition: contrast, mixed) mixed ANOVA, with learning block as a within subjects variable, revealed a significant main effect of block, F (1.942, 60.217) = 16.897, $p < .001$. Participants in the contrast condition performed above chance on the final block of learning items, $t(14) = 2.667$, $p = .018$, whereas participants in the mixed condition did not, $t(17) < 1$). The main effect of condition and the interaction of block and condition were not significant.

A second 3 (transfer block) X 2 (condition: contrast, mixed) mixed ANOVA, with transfer block as a within subjects variable, failed to reveal any significant effects, all Fs < 1, even though participants in both conditions performed above chance (50%) in the second transfer block (contrast: $t(14) = 3.318$, $p = .005$; mixed: $t(17) = 2.279$, $p = .036$), and participants in the contrast condition performed above chance in the first (contrast: $t(14) = 2.881$, $p = .012$; mixed: $t(17) = 1.725$, $p = .103$) and the third transfer block, (contrast: $t(14) = 2.990$, $p = .010$; mixed: $t(17) = 2.018$, $p = .060$).

The first hypothesis, that participants in the contrast condition would perform better than chance on the base transfer items, was not supported. Participants in the contrast condition achieved a mean accuracy of 52.08% (SD = 15.79%), which was not significantly greater than chance (chance = 50% correct), single sample t (14) = 0.511, p = 0.617, *ns*. A one-way ANOVA confirmed that there were no differences between the control and mixed conditions in accuracy on the base transfer items, F (1, 31) = 0.755, *ns*.

The second hypothesis, that participants in the contrast condition would perform better than chance on the final block of transfer items, was supported. Participants in the contrast condition achieved a mean accuracy of 71.11% (SD = 27.34%), which was significantly different than chance (chance = 50% correct), single sample t (14) = 2.990, p = .010. Participants in the mixed condition did not perform significantly above chance (M = 60.88%; SD = 22.87%), single sample t (17) = 2.018, p = .060. A one-way ANOVA failed to find any significant differences between the experimental conditions on accuracy on the final block of the transfer items, F (1, 31) = 1.371, p = .251.

Comparisons of the response latencies were again performed on log transformed RTs on the test items. There were no differences between the base-learning (old) items and the base-transfer (new) items for either group (contrast: t(14) = 1.367, p = .193; mixed: t(17) < 1). The groups did not differ significantly on either the base learning, F(1, 31) = 3.468, p = .072, or the base-transfer items, F(1, 31) < 1.

Questionnaire Data

The pattern of results for the questionnaire data for Pilot Study 2 was very similar to that of Pilot Study 1. Explicit knowledge of some or all of the category structure following the experiment (Question 1) was evident for 24 of the 33 participants (73.52%). The number of participants demonstrating no, some, and full awareness for Questions 1, 3, and 4 is presented in Table 27. Fisher Exact tests were again used to compare no awareness versus some or full awareness across groups. None of the tests were significant, all $p > .50$. The mean score for Question 2 for the contrast condition was 6.60 (SD = 2.47) out of 8, and 4.83 (SD = 2.43) for the mixed condition. This mean difference was reliable, $t(31) = 2.062$, $p = .048$. As in Pilot Study 1, the high standard deviations for these means suggests that there was substantial individual variation and the number of participants demonstrating full knowledge of the pattern was again compared for those in the contrast versus the mixed conditions. In the contrast condition, 11 of 15 participants achieved the maximum score (full knowledge of the category structure), whereas only 5 of the 18 participants in the mixed condition reached the same criteria. This difference was reliable, Fisher Exact test, $p < .05$.

Discussion

There was mixed support for improved learning in Pilot Study 2 over Pilot Study 1. First, there was a significant effect of learning block in the repeated measures ANOVA, which indicates that some learning may have occurred. Second, participants in the contrast condition performed significantly above chance on the final block of learning

trials. That participants in the contrast condition demonstrated above chance performance in the final block of the learning phase when feedback is present, but did not perform any better in chance during the test phase when feedback is absent suggests that participants have acquired the category structure without learning which items belong to which category. The present methodology successfully produces the pattern-sequence effect; participants are learning the category structure in the contrast condition but not in the mixed condition. Furthermore, knowledge of the category structure is abstracted from individual exemplars, indicated by drop to chance performance in the absence of feedback.

Experiment 1 will incorporate several changes in an effort to improve learning and provide a test of category knowledge consistent with what was learned. First, in order to accurately assess participant's knowledge of the unseen category features, those items must be placed in context. One way to accomplish this is to change the testing procedure to more closely resemble the learning procedure by providing feedback to participant responses. Experiment 1 will therefore use the same presentation format for the generalization items as for the learning items, requiring participants to enter their responses via the keyboard and receiving feedback on each item, and will query three features for each items throughout the experiment. In the learning phase, they will be the learning items. In the test phase, two will be learning items, the remaining one will be the generalization item. Second, a blocked design will be used instead of a contrast design in the pretraining phase such that the items from Category A will be presented twice before the items from Category B are presented twice. The use of the same items

in the pretraining and learning phases as well as the repetition of the items in the pretraining phase should improve learning by reducing the number of items to learn. Third, a new cover story will be used to discourage participants from looking for a "trick" in the experiment. Participants in both pilot studies mentioned using irrelevant information to complete the task, including the starting and ending letters of the muscle or plant names as well as the answer sequence. Participants in Experiment 1 will be told that they are participating in a norming study to establish the relative difficulty of stimuli to be used in future experiments. A difficulty-rating task will be added to support this cover story. The consistent use of three feature values throughout the experiment will also support this cover story.

Table 24 Contrast Condition Stimuli Used in Pilot Study 2.

Name	Set	Root	Stem	Leaf	Flower
		Category A			
Acorus	Pretraining	Taproot*	Smooth	Pointed	Headed
Briza	Pretraining	Taproot*	Smooth	Pointed	Headed
Cereus	Pretraining	Taproot	Smooth*	Pointed	Headed
Cynara	Pretraining	Taproot	Smooth*	Pointed	Headed
Mentha	Pretraining	Taproot	Smooth	Pointed*	Headed
Ricinus	Pretraining	Taproot	Smooth	Pointed*	Headed
Sedum	Pretraining	Taproot	Smooth	Pointed	Headed*
Trapa	Pretraining	Taproot	Smooth	Pointed	Headed*
Apium	Pretraining	Taproot*	Smooth	Pointed	Headed
Betula	Pretraining	Taproot*	Smooth	Pointed	Headed
Calluna	Pretraining	Taproot	Smooth*	Pointed	Headed
Caltha	Pretraining	Taproot	Smooth*	Pointed	Headed
Morus	Pretraining	Taproot	Smooth	Pointed*	Headed
Ruscus	Pretraining	Taproot	Smooth	Pointed*	Headed
Salvia	Pretraining	Taproot	Smooth	Pointed	Headed*
Taxus	Pretraining	Taproot	Smooth	Pointed	Headed*
Acer	Learning	Taproot*	Smooth	Pointed	Headed
Buxus	Learning	Taproot*	Smooth	Pointed	Headed

Table 24 Continued.

Name	Set	Root	Stem	Leaf	Flower
Ferula	Learning	Taproot	Smooth*	Pointed	Headed
Hordeum	Learning	Taproot	Smooth*	Pointed	Headed
Juncus	Learning	Taproot	Smooth	Pointed*	Headed
Nolana	Learning	Taproot	Smooth	Pointed*	Headed
Rumex	Learning	Taproot	Smooth	Pointed	Headed*
Vitis	Learning	Taproot	Smooth	Pointed	Headed*
Larix	Transfer	Taproot	Smooth	Pointed	Headed
Punica	Transfer	Taproot	Smooth	Pointed	Headed
Sorbus	Transfer	Taproot	Smooth	Pointed	Headed
		Category B			
Alnus	Learning	Fibrous*	Woody	Rounded	Spiked
Borago	Learning	Fibrous*	Woody	Rounded	Spiked
Geum	Learning	Fibrous	Woody*	Rounded	Spiked
Hedera	Learning	Fibrous	Woody*	Rounded	Spiked
Juglans	Learning	Fibrous	Woody	Rounded*	Spiked
Nigella	Learning	Fibrous	Woody	Rounded*	Spiked
Rubus	Learning	Fibrous	Woody	Rounded	Spiked*
Vitex	Learning	Fibrous	Woody	Rounded	Spiked*
Linum	Transfer	Fibrous	Woody	Rounded	Spiked

Table 24 Continued.

Name	Set	Root	Stem	Leaf	Flower
Pilea	Transfer	Fibrous	Woody	Rounded	Spiked
Salix	Transfer	Fibrous	Woody	Rounded	Spiked

Note. Items marked with an asterisk were not show during learning (base-generalization items).

Table 25 Mixed Condition Stimuli Used in Pilot Study 2.

Name	Set	Root	Stem	Leaf	Flower
		Category A			
Acorus	Pretraining	Taproot*	Smooth	Pointed	Headed
Briza	Pretraining	Taproot*	Smooth	Pointed	Headed
Cereus	Pretraining	Taproot	Smooth*	Pointed	Headed
Cynara	Pretraining	Taproot	Smooth*	Pointed	Headed
Mentha	Pretraining	Taproot	Smooth	Pointed*	Headed
Ricinus	Pretraining	Taproot	Smooth	Pointed*	Headed
Sedum	Pretraining	Taproot	Smooth	Pointed	Headed*
Trapa	Pretraining	Taproot	Smooth	Pointed	Headed*
Acer	Learning	Taproot*	Smooth	Pointed	Headed
Buxus	Learning	Taproot*	Smooth	Pointed	Headed
Ferula	Learning	Taproot	Smooth*	Pointed	Headed
Hordeum	Learning	Taproot	Smooth*	Pointed	Headed
Juncus	Learning	Taproot	Smooth	Pointed*	Headed
Nolana	Learning	Taproot	Smooth	Pointed*	Headed
Rumex	Learning	Taproot	Smooth	Pointed	Headed*
Vitis	Learning	Taproot	Smooth	Pointed	Headed*
Larix	Transfer	Taproot	Smooth	Pointed	Headed
Punica	Transfer	Taproot	Smooth	Pointed	Headed

Table 25 Continued.

Name	Set	Root	Stem	Leaf	Flower
Sorbus	Transfer	Taproot	Smooth	Pointed	Headed
		Category B			
Alnus	Learning	Fibrous*	Woody	Rounded	Spiked
Borago	Learning	Fibrous*	Woody	Rounded	Spiked
Geum	Learning	Fibrous	Woody*	Rounded	Spiked
Hedera	Learning	Fibrous	Woody*	Rounded	Spiked
Juglans	Learning	Fibrous	Woody	Rounded*	Spiked
Nigella	Learning	Fibrous	Woody	Rounded*	Spiked
Rubus	Learning	Fibrous	Woody	Rounded	Spiked*
Vitex	Learning	Fibrous	Woody	Rounded	Spiked*
Apium	Pretraining	Fibrous*	Woody	Rounded	Spiked
Betula	Pretraining	Fibrous*	Woody	Rounded	Spiked
Calluna	Pretraining	Fibrous	Woody*	Rounded	Spiked
Caltha	Pretraining	Fibrous	Woody*	Rounded	Spiked
Morus	Pretraining	Fibrous	Woody	Rounded*	Spiked
Ruscus	Pretraining	Fibrous	Woody	Rounded*	Spiked
Salvia	Pretraining	Fibrous	Woody	Rounded	Spiked*
Taxus	Pretraining	Fibrous	Woody	Rounded	Spiked*
Linum	Transfer	Fibrous	Woody	Rounded	Spiked

Table 25 Continued.

Name	Set	Root	Stem	Leaf	Flower
Pilea	Transfer	Fibrous	Woody	Rounded	Spiked
Salix	Transfer	Fibrous	Woody	Rounded	Spiked

Note. Items marked with an asterisk were not show during learning (base-generalization items).

Table 26 Pilot Study 2 Means and Standard Deviations for Learning, Test, & Transfer Accuracy, and Test Latencies.

	Contrast		Mixed	
	Mean	SD	Mean	SD
Learning Trials Accuracy				
Learning Block 1	52.92%	18.66%	44.44%	11.63%
Learning Block 2	62.22%	21.30%	51.39%	17.00%
Learning Block 3	64.31%	20.77%	51.04%	15.45%
Test Trials Accuracy				
Base Learning Accuracy	54.17%	13.57%	50.69%	8.66%
Base Transfer Accuracy	52.08%	15.79%	47.92%	11.74%
Test Trials Latency				
Base Learning Latency	1732.07	627.96	2235.42	874.77
Base Transfer Latency	1886.49	914.60	2181.06	1002.42
Transfer Trials Accuracy				
Transfer Block 1	67.50%	23.53%	59.03%	22.20%
Transfer Block 2	70.28%	23.67%	59.03%	16.80%
Transfer Block 3	71.11%	27.34%	60.88%	22.87%

Table 27 Pilot Study 2 Post Questionnaire Analysis: Questions 1, 3, 4.

	Question 1 Awareness			Question 3 Awareness			Question 4 Awareness		
	None	Some	Full	None	Some	Full	None	Some	Full
Contrast	4	8	3	2	5	8	2	4	9
Mixed	5	11	2	4	9	5	1	11	6
Total	9	19	5	6	14	13	3	15	15

Note. Question 1 asked participants to describe the strategies they used to learn the items. Question 3 asked participants to describe how, if at all, the plants or their features were similar. Question 4 informed participants that the items belonged to categories and asked them to describe how those categories were formed.