

1 General Introduction

Categories help us organize the world. They help us predict and hypothesize about category members, helping us quickly select the most appropriate response for each situation. Language plays a key role in categorization and category learning. It provides structure in the form of category labels and affects how we think about and even perceive the categories themselves. As Lupyan (2012) puts it, language augments our thought. Thus, any thorough investigation of how we learn categories must consider the role of language.

Indeed, many theoretical frameworks of category learning consider the role of language. For example, COVIS, a key theory in perceptual category learning, stands for "competition between verbal and implicit systems" (Ashby et al., 1998). Similarly, a theory put forth by Minda and colleagues is called "A theory of verbal and nonverbal category learning" (Minda & Miles, 2010). However, most theories of category learning that consider language primarily focus on whether or not language has influences category learning, rather than exploring what kind of influences language might have .

Thus, the current work seeks to both define a theory of category learning and explore the role language has in this theory. In this review I will synthesize multiple approaches to category learning. Following the synthesis, I will review relevant literature that provides suggestions as to how language might be involved in category learning. Through these efforts, I will provide a theoretical framework and hypotheses for this dissertation.

1.1 Dual-systems model for category learning

Multiple theories converge on the idea that there are two systems for category learning. In this section, I will first describe a generalized dual-systems model that pulls threads from all of these theories and then go on to describe how each theory fits into the overarching framework.

1.1.1 Proposed model

The proposed model involves two systems for category learning. The first, which I title the **associative system**, uses associative mechanisms in an iterative manner to learn distributions of features. This system is best suited for learning multidimensional *similarity-based* categories such as natural kinds, where it is difficult to describe necessary and sufficient rules for inclusion. Similarity-based categories have features that are correlated and probabilistic, such that a given category instance may not have all of the category-relevant features but does tend to have some distribution of them. For example, although Manx cats do not have tails, a typical feature of cats, they are still undeniably members of the category *cat*. Thus, the

associative system must be able to extract the most frequent pattern of features over many instances in order to learn a category.

In contrast, the **hypothesis-testing system** uses a more explicit learning method to test and adjust hypotheses about category boundaries. This method relies on selection of relevant features rather than representation of a distribution of feature probabilities. As such, it is most suited for learning rule-based categories, which typically have one or a few easily verbalizable rules for inclusion. For example, the *ad hoc* category *things to be sold at the garage sale* has a simple rule for inclusion that perfectly separates members from non-members.

Thus, we have two systems for category learning, each one ideal for learning a different type of category (similarity-based vs. rule-based). In the upcoming sections, I will describe theoretical and empirical evidence for a dual-systems model from five different approaches to category learning. I will show how each approach informs the current theoretical framework.

1.1.2 COVIS

COVIS stands for Competition between Verbal and Implicit Systems. First proposed by Ashby et al. in 1998, it is a prominent theoretical framework for perceptual category learning. This framework provides a dual-systems model that is grounded in neuropsychological data, allowing it to suggest neurobiological underpinnings for the two systems. It is important to note that this framework is mostly concerned with perceptual categories, which are defined as "a collection of similar objects belonging to the same group" (Ashby & Maddox, 2005, p. 151). This is in contrast to concepts, which Ashby and colleagues define as groups of related ideas. Thus, this approach focuses on categorizing objects that can be encountered and perceived in the real world.

As can be inferred from the title, the two category learning systems in COVIS are the **verbal** and **implicit** systems. The verbal system is COVIS' answer to our hypothesis-testing system. It is a declarative learning system that uses a hypothesis-testing method to learn category rules, typically for rule-based stimuli. Under COVIS, rule-based stimuli must have inclusion rules that are easy to describe verbally. Typically, rule-based stimuli used by Ashby and colleagues have a single rule for inclusion or two rules combined by "and" or "or." When a rule-based category involves multiple dimensions, decisions about each dimension are made separately, and these decisions are used to evaluate the logical operators. In other words, each dimension is considered on its own before their combination. These guidelines for rule-based categories ensure that an explicit hypothesis-testing method can be used to learn them. When learning a new category, the verbal system holds potential category inclusion rules in working memory that are then tested as stimuli

are encountered. Over time, hypotheses are tested and switched until they reflect the optimal strategy for categorization. Individual differences in rule-based category learning have been shown to be related to an individual's cognitive flexibility (Reetzke et al., 2016), suggesting that the verbal system relies at least partially on executive function.

The implicit system from COVIS is most similar to our associative system. Like the associative system, it uses incremental learning to find category boundaries. It is most ideal for learning information-integration categories, which are like similarity-based categories but have also some specific guidelines. Information-integration categories are defined by some combination of dimensions. However, while each dimension can be considered separately in rule-based categories, all dimensions must be considered simultaneously for information-integration categories. Information-integration category membership depends on both the values associated with each dimension as well as the relationship between these values. Information-integration category boundaries are difficult or impossible to describe verbally. The structure of information-integration categories require an iterative, associative learning method. COVIS suggests that the implicit system relies on an information stream that connects stimuli, motor responses, and feedback to learn category membership.

One of the most substantial contributions of COVIS is its strong grounding in neurobiology. In the original paper, Ashby and colleagues proposed specific brain regions involved the verbal (hypothesis-testing) and implicit (associative) systems, supported by neuroimaging and patient studies. The verbal (hypothesis-testing) system relies on the prefrontal cortex (PFC), anterior cingulate cortex (ACC), striatum, hippocampus, and the head of the caudate nucleus. Information about the stimuli are processed in fronto-striatal loops, where potential category rules are generated. The PFC keeps these rules in working memory while the ACC and the head of the caudate nucleus mediate switching between rules based on feedback. Finally, the hippocampus stores longer-term memory of which rules have already been tested. The hippocampus is only involved when the task is complex enough that previously tested rules cannot all be stored in working memory (Ashby & Maddox, 2005, 2011).

Patient data shows that individuals with frontal damage as well as individuals with Parkinson's disease, which affects the basal ganglia including the caudate nucleus, show difficulty in rule-based tasks such as the Wisconsin Card Sorting Test (Robinson et al., 1980) and an experimental rule-based category learning task (Ashby et al., 2003). This suggests that both frontal regions and the basal ganglia are involved in rule-based categorization. More recent neuroimaging work, however, is still mixed as to the involvement of different areas specifically for rule-based categorization. Soto et al. (2013) found that two separate rule-based tasks could be differentiated based on activation in ventro-lateral PFC, suggesting that specific rules are stored in that region. Nomura et al. (2007) found activation in the medial temporal lobe (MTL), which

contains the hippocampus, specifically for rule-based categorization. However, a later study failed to find any activation that was specifically greater for rule-based categorization (Carpenter et al., 2016). Thus, the neural underpinnings of the verbal (hypothesis-testing) system are still under debate.

The implicit (associative) system from COVIS has a different neurobiological pathway for category learning. It uses incremental learning rather than hypothesis testing to learn information-integration (similarity-based) categories. The main structure involved in this procedural learning system is the striatum, which is involved in reinforcement learning with dopamine as the reinforcement signal. From the striatum, information about the category is sent to the thalamus and the globus pallidus, which is within the basal ganglia. Information then runs to motor and premotor cortex. This system links stimuli, motor responses during categorization, and feedback to allow the participant to learn categories. Neuroimaging studies using the implicit system again are mixed, with some finding activation in the caudate body while others fail to find that activation, instead seeing activity in parahippocampal regions (Carpenter et al., 2016; Nomura et al., 2007). A separate study also found a role for the putamen in similarity-based category learning (Waldschmidt & Ashby, 2011). As with the verbal system, the neural basis of the implicit system requires more study.

COVIS provides us with a few key insights. First, it is one of the most studied dual-systems theories of categorization. While Ashby and colleagues generally use visual stimuli for their tasks, this paradigm has been extended to other perceptual domains such as hearing/speech (Chandrasekaran et al., 2014, 2016). As such, research on the current theoretical framework (associative/hypothesis-testing systems) has much COVIS literature which we can compare it to. It also makes clear claims about the neurobiological basis of the two systems of category learning. While the specifics of these claims are still under debate in the literature, they at least provide regions of interest for researchers who want to conduct neuroimaging research on a dual-systems model of category learning. Finally, this approach is one of the only ones to consider how the two systems interact.

1.1.3 Dimensionality

The dimensionality approach, led by Lupyan and colleagues, considers categories in terms of the dimensions on which they cohere. Low-dimensional categories are the same on one or a small number of dimensions (e.g., color) and allow other dimensions to vary. Low-dimensional categories are similar to rule-based categories, as they can be described using relatively simple rules (e.g., *things that are red*). In fact, some of Lupyan's papers define low-dimensional categories as those that have a single dimension that can distinguish category members from non-members (Lupyan & Mirman, 2013). Examples of low-dimensional categories from this study include *things made of wood* and *things with handles*.

In contrast, high-dimensional categories are those that cohere on multiple dimensions, often so many that category rules are difficult to describe. Examples of high-dimensional categories from the previously-mentioned study include *birds*, *tools*, *things that fly*, and *objects that hold water*. Most natural kinds and artifacts are high-dimensional, as well as some *ad hoc* categories. Like similarity-based categories, high-dimensional categories require their members to be the same on most (but not all) relevant dimensions.

The core prediction tested using this approach is that low-dimensional categorization relies more heavily on language than high-dimensional categorization. The dimensionality approach postulates that language helps an individual select features, which is a process only helpful for low-dimensional categorization. High-dimensional categorization relies on creating associations across multiple features, which does not require language.

To explore this prediction, Lupyan and colleagues interfered with language ability in multiple ways across studies. In each study, they found that a reduction in language ability was associated with poorer performance on low- but not high-dimensional categorization. Lupyan & Mirman (2013) measured categorization ability in individuals with aphasia for both low- and high-dimensional categories. They found that the individuals with aphasia performed similarly to unimpaired controls on the high-dimensional categories, but showed significantly lower accuracy on the low-dimensional categories. Lupyan (2009) used a concurrent verbal load to reduce the verbal resources available during a categorization task. He found that individuals showed significantly poorer categorization with a verbal load as compared to a visuospatial load specifically for category judgments based on a single dimension (color or size) but not for those based on multiple dimensions (theme). Other studies manipulated language ability by using transcranial direct current stimulation (tDCS). One study found that reducing excitability in a language-critical region (left inferior frontal gyrus) led to poorer performance on low-dimensional but not high-dimensional categorization (Lupyan et al., 2012). Another study used stimuli that could either be categorized using a uni-dimensional or a bi-dimensional strategy. Reducing excitability over Wernicke's area made participants more likely to choose the bi-dimensional strategy, indicating that interfering with language functioning resulted in participants using higher-dimensional categorization (Perry & Lupyan, 2014).

The dimensionality approach to category learning and the studies done to test it provide multi-method evidence for the role of language in low-dimensional categorization. Unlike COVIS, where the verbal system largely uses language to describe and rehearse candidate category rules, the dimensionality approach suggests that language is used to select relevant features for a category. This idea has highly influenced this paper's dual-systems model, in which the hypothesis-testing system does select category-relevant features. However, the evidence for this approach is largely unable to speak for the system underlying high-dimensional categorization, as most of the effects for this system are null. Thus, it is not clear from

this approach whether the hypothesized broad inter-item association building is in fact how individuals learn high-dimensional categories. In addition, while the authors claim that poorer low-dimensional categorization performance reflects difficulty in selecting category-relevant dimensions, the studies mentioned above do not directly test how interfering with language ability affects selection or inhibition ability.

1.1.4 Statistical Density

The statistical density framework focuses on the structure of categories defined by the relationships among members and non-members. Pioneered by Sloutsky, it proposes two category learning systems that are each used to extract different types of regularities from a stream of information, allowing for flexibility in the data collected (Sloutsky, 2010). Sloutsky's main metric for describing categories is called *statistical density*. In this section, I will describe statistical density in a broad sense; for more detailed information on how to calculate it, see Appendix A (p. 30).

The statistical density of a category is related to the ratio between the amount of entropy within a target category and the entropy between the target category and other categories in the set. In this context, entropy refers to variation within features. Consider a set of shapes. These shapes can vary in shape, size, and color. The within-category entropy for squares is all of the different sizes and colors for a single shape: squares. The between-category entropy includes all of the variation in size, color, and shape for the items in the set. **Sparse** categories have lots of within-category entropy; the items in the category cohere on only one or a few dimensions. All other dimensions are allowed to vary freely. In our shape example, a sparse square category would have squares of all color and sizes, such that color and size was not related to shape. Thus, to find the category *square*, an individual would have to isolate the "shape" feature.

In contrast, **dense** categories have little within-category entropy; their members have multiple intercorrelated features that together are predictive of category membership. There are few irrelevant features in dense categories. Within our set of shapes, the square category would be considered dense if all squares shared the same color and size. The distribution of these other features are what determine the statistical density of a category. If irrelevant features (here, color and size) are correlated with the relevant feature(s), the category is dense. If they vary independently of the relevant features, the category is sparse. Thus, statistical density expresses the relationships between features within a category as well as within an entire set of items. A particularly interesting feature of this metric is that statistical density is a continuous spectrum: categories can be very dense, very sparse, or anywhere in between.

This framework also outlines two systems used to learn categories with different densities. Dense categories are best learned by the compression-based system, which takes input and reduces it by representing

some but not all features. With more instances, relevant features for a given category will be represented more frequently and survive the compression. In contrast, features that appear infrequently will be mostly filtered out. The compression-based system does not use conscious selection to determine which features are represented. Instead, redundant and probable features are more likely to continue on. The many correlated features of a dense category are easily extracted using this system, which is quite similar to our associative system.

The second learning system is called the selection-based system. This system directs attention towards relevant features, sampling those features for later representation. This system learns by aiming to reduce error. As feedback is encountered, the system shifts attention from those dimensions that create categorization errors to those that do not. The selection-based system relies heavily on multiple aspects of executive function, including inhibition and selection. It is best for learning sparse categories. While over time the compression-based system could be able to learn sparse categories, as the freely varying irrelevant features would eventually be less frequent than relevant features, this process would be much more inefficient than selecting and testing individual features. The selection-based system is Sloutsky's version of our hypothesis-testing system. Some research shows that sparse categorization is correlated with performance on a flanker task, which is often used to measure selection and inhibition (Perry & Lupyan, 2016). This suggests that at least some executive functions are related to sparse category learning.

The statistical density framework also discusses the development of these two systems. It suggests that children have access to the compression-based (associative) system early in development, as its mechanisms involve brain structures that develop relatively early, such as inferotemporal cortex (Rodman, 1994). In contrast, the selection-based (hypothesis-testing) system involves more frontal regions that develop later, such as dorsolateral prefrontal cortex and anterior cingulate cortex (Eshel et al., 2007; Lewis, 1997; Segalowitz & Davies, 2004). Thus, this framework posits that the compression-based system develops before the selection-based system. Sloutsky and others have done some studies on different age groups testing the two systems with categories of different densities to verify this claim.

Kloos & Sloutsky (2008) tested both of these systems in children and adults. They engaged the two systems separately by modifying task demands. Some participants learned novel categories by being taught the rules for inclusion (e.g., "Ziblets have a short tail."). This activated the selection-based (hypothesis-testing) system. Other participants learned these categories by viewing multiple members, engaging the compression-based (associative) system. Thus, the authors could test how well individuals could learn novel categories of different densities depending on whether the category density matched the system being engaged. For both children and adults, learning performance was high when the category density and task instructions matched. However, while the adults were able to adapt and learn the categories in

mismatch conditions, children were specifically unable to learn sparse categories just by viewing multiple instances. This suggests that children are not able to use the selection-based system without direct guidance from task instructions.

Other evidence for the developmental course of the two systems comes from a study of infants and adults. This study used a switching paradigm to investigate whether individuals were selecting specific features (using the selection-based/hypothesis-testing system) or processing the entire stimulus holistically (using the compression-based/associative system). They found that when viewing sparse categories, adults showed a significant switch cost as the category-relevant feature was changed, while infants did not. This indicates that adults viewing sparse categories were focusing on a specific feature, while infants were processing the entire stimulus. Eye movement data also suggested that even with holistic processing, infants were able to learn the categories (Best et al., 2013). Thus, this study found that adults and infants used different systems for learning the same categories.

A similar finding comes from a study of change detection. One study found that while both children and adults showed high accuracy when detecting change in a cued stimulus, children were better than adults at detecting change in task-irrelevant stimuli. Similar results were found when children and adults were asked to perform familiarity judgments on items seen during a visual search task: high performance for both groups when probing relevant features, and higher performance for children than adults when probing irrelevant features (Plebanek & Sloutsky, 2017). The results from both of these experiments suggest that children attend to a stimulus in a diffuse manner, even when task demands suggest a selective strategy. This is consistent with a later-developing selection-based system, as children may be processing these features using the compression-based system, which preserves even category-irrelevant features.

Thus, the statistical density approach to category learning provides two major points for consideration. First, the statistical density metric itself emphasizes the idea that there aren't two distinct types of categories (e.g., rule-based and similarity-based). Instead, categories exist on a spectrum ranging between these extremes. It is still unknown how a dual-system model would deal with stimuli that lie directly in the middle of this spectrum, however. Second, this framework is one of only a few that describes a developmental trajectory for a dual-systems framework of category learning.

1.1.5 Verbal/nonverbal

Like some of approaches discussed above, the verbal/nonverbal approach is a dual-systems model of category learning. While other approaches discuss the role of language in category learning, none make it as central as this approach by Minda & Miles (2010). The two systems in this approach are called the

verbal and **nonverbal** systems. These systems align well both with the framework outlined in this paper as well as with other approaches. The verbal system uses hypothesis testing to determine the verbal rules best suited to characterize a category. In contrast, the nonverbal system uses associative mechanisms to learn categories, iteratively learning which features go together in predicting category membership.

A unique feature of this approach to category learning is its emphasis on traditional models of working memory and their role in the category learning process. Minda & Miles (2010) state that the verbal system relies heavily on working memory, especially the phonological loop and central executive, to rehearse and select potential rules (A. D. Baddeley & Hitch, 1974). The nonverbal system, meanwhile, uses the visuospatial sketchpad to store and rehearse visual information, but overall uses working memory to a lesser extent than the verbal system. Evidence for these hypotheses comes from a study showing that children exhibited adult-like performance when learning categories using the nonverbal system and reduced performance for categories that required use of the verbal system. This study also showed that adults showed more child-like performance when learning categories suited to the verbal system while under concurrent verbal load, suggesting that the verbal system indeed needs verbal working memory resources (Minda et al., 2008).

While the two systems described in Minda & Miles (2010) are quite similar to the systems hypothesized in this paper, there remains a core difference: the nonverbal system does not posit a role for language. This is likely due to the way Minda & Miles (2010) ground their dual-systems model in working memory. As will be discussed shortly, language can be very useful even for iterative, association-based learning, although perhaps not in the form of a verbal working memory resource. Thus, the verbal/nonverbal dual-systems model of category learning provides us with evidence that verbal working memory and executive resources support rule-based category learning but does not fully consider the ways in which language may influence category learning.

1.1.6 Taxonomic/thematic

As these previous frameworks have shown, when considering categories we must think carefully about how the items in a category relate to each other. The taxonomic/thematic framework is yet another way to consider relations within categories. **Taxonomically** related items are those we might think of as belonging to the same everyday category (e.g., animals, plants, tools, etc.). **Thematically** related items are those that go together in everyday life but are not necessarily part of the same category (e.g., needle and thread, apple and worm).

Similar to and perhaps even more so than the statistical density approach, the taxonomic/thematic

framework has been able to provide many valuable insights about the developmental trajectory of categorization. The typical task in this line of research is a grouping task, where individuals are given a set of items and asked to group the ones that are "alike" or "the same." Early research on this topic suggests that children primarily categorize items using thematic relations in kindergarten and switching to taxonomic relations later in childhood, although even this early work indicated that young children are able to learn taxonomic relations if necessary (Piaget et al., 1964; Vygotsky, 1962). Smiley & Brown (1979) found that the preference for taxonomic versus thematic relations switches between first and fifth grade as well as between college and old age, such that the very young and the elderly both show a preference for thematic relations. However, in another study, college-aged adult participants chose a thematically-related item more frequently in a triad task across ten different experiments, including one with the same stimuli used in Smiley and Brown's paper (Lin & Murphy, 2001).

Rather than being tied directly to age or ability, the preference for thematic or taxonomic classification may depend on an individual's goals. Markman & Hutchinson (1984) had children between the ages of 2 and 4 complete a triad task. The children were shown a target picture (e.g. a tennis shoe) as well as two options: one that was taxonomically related (e.g., a high-heeled shoe) and one that was thematically related (e.g., a foot). The children were then asked to "find the one that is the same." With these directions, the children chose the thematically-related object about half of the time. However, when a novel label was applied to the task (e.g., This is a *dax*. Can you find another *dax*?), the children were more likely to choose the taxonomically related item. Thus, having a category label focused the task and directed attention towards taxonomic category structure rather than thematic relations. Further research in children between the ages of 2 and 4 manipulated many parts of the typical triad task, including experimenter instructions and medium of presentation (pictures vs. physical objects). They found that the thematic preference seen in Smiley & Brown (1979) seemed to be strongly affected by task instructions and age (Waxman & Namy, 1997). Some research suggests that what is developing in young childhood is not a sensitivity to different types of relations but instead the ability to flexibly switch between thematic and taxonomic relations according to task demands (Blaye & Bonthoux, 2001).

Taxonomic and thematic categories and processing share many similarities with the approaches discussed above. Taxonomic categories are like similarity-based categories. Both are what a typical individual would consider to be a "category;" they include natural kinds and artifacts. In contrast, thematic categories are more similar to rule-based categories. Both can be defined using a rule like "usually found in a kitchen" or "used for sewing." Thinking about rule-based categories in terms of thematic relations brings a new aspect to these categories: situational similarity. Often, rule-based categories are *ad hoc*, or created for and bound to a certain situation (e.g., "things to be sold at the garage sale"). Thus, when we think about how

we learn and process rule-based categories using the hypothesis-testing system, we should keep in mind how we use our knowledge of situations or episodes in categorization.

1.2 Vocabulary/labels and category learning

Much theory and research has considered how having a single word for a category or concept affects how an individual learns and processes that category. In this document, we will consider the word form associated with a given category (either spoken or written) to be the **category label**. Thus, a category has two potential pieces. First, there is the category's meaning, or the way in which members belong to a category. As discussed previously, this can be a set of defined rules (e.g., anything you plan to sell is a part of the category *things to sell at the garage sale*) or an implicit set of fuzzy category boundaries (e.g., the ways in which you judge whether an item is a *chair*). The second piece of the category is its label. Individuals learning new categories often learn both the meaning and the label.

There have been multiple viewpoints on just how labels interact with the category or concept they describe and refer to. One line of thought postulates that labels are attached to concepts that can be formed in their absence (Gillette et al., 1999; Snedeker & Gleitman, 2004). This framework tends to focus on early-acquired object concepts, which are thought to be built nonverbally in the infant before language is acquired. Experiments done under this framework reveal interesting and important findings about the information that best supports a mapping between a category meaning and its label (e.g., having a syntactic frame for a category label leads to much quicker learning than just observing the use of the label in multiple situations). However, this viewpoint places little importance on the interplay between the label and the meaning; at best, the label is an additional way to access the meaning but does not seem to differ from any other feature.

Other researchers suggest that labels dynamically interact with meanings, and that having a single word for a meaning fundamentally changes how individuals think about and even perceive a category. In the words of Waxman & Markow (1995), words (labels) are "invitations to form categories". When a child encounters a novel word form applied to an object, they are initially biased to interpret that word form as a label for a category rather than the name of that singular object. Indeed, receiving a label for a category helps 12-month-old infants focus on common features more than just directive speech (Althaus & Mareschal, 2014). In adults, labels promote category learning even when they are redundant, and they do so even more than additional nonverbal features (Lupyan et al., 2007). Even more interestingly, having a label can change perceptual processing across development. Infants shown a certain set of objects without an accompanying label will sort these objects into multiple categories using visual features. However, if a single label is applied to the same set of objects, the infants will create only one category (Plunkett et al.,

2008). In adults, hearing category labels affects visual perception. Participants asked to find 2s or 5s in a visual display showed better accuracy and shorter reaction time when hearing two or five immediately before the display appeared (Lupyan & Spivey, 2010).

The evidence cited above suggests that labels are special in some way—they are not simply additional features of fully-formed concepts. This may be because labels encourage individuals to focus on features that are more diagnostic (i.e., more often associated with members of a category) rather than features that are specific to a given instance. A number of studies from Lupyan and colleagues support this idea. For example, Edmiston & Lupyan (2015) found that adults tended to look at more typical instances of a category when hearing a label. Thus, when hearing the word "bird," participants were more likely to look at a robin (a more typical bird) than a penguin (a less typical bird). They also found that when listening to sounds associated with a category (e.g., bird chirp), participants tended to look at more likely sources of the sound (e.g., images of birds with their mouths open). This suggests that labels activate a typical, abstracted representation of a category while other sounds activate a more specific instance of that category that is congruent with the sound itself.

Similar findings come from a study looking at the formal category triangles. Triangles are by definition figures with three sides any figure with three sides can be labeled a triangle. However, Lupyan (2017) found that typicality effects for triangles in multiple tasks were introduced when the word triangle was used. When asked to draw a triangle, participants most often drew isosceles or equilateral triangles with their base parallel to the horizontal (i.e., more canonical triangles). However, when instructed to draw a three-sided figure, participants drew a variety of triangles. The same typicality-related pattern of results was found for multiple other tasks, including typicality judgments, speeded recognition, and shape judgment. Another study found that pairing category instances with labels increased fixations on category-relevant features, as compared to pairing them with random words or silence, even for sparse categories (Barnhart et al., 2018). This study used an associative learning environment, where participants viewed many instances, were not asked to make category judgments, and were not provided any feedback on categorization. Thus, when the associative system is engaged, labels draw attention towards the most category-relevant features available.

This phenomenon is related to other research showing that other seemingly rule-based categories (e.g., grandmothers, odd numbers) show typicality effects (Armstrong et al., 1983; Lupyan, 2013). Armstrong and colleagues suggest that typicality effects are seen in what might be considered rule-based categories because these categories are defined both by rules for inclusion (e.g., having a grandchild) as well features that are used in identification (e.g., gray hair, tendency to bake cookies). This line of reasoning implies a continuum between rule-based and similarity-based categories, where categories with definite and verbalizable rules for inclusion are subject to processing most often associated with similarity-based categories.

Thus, having a label for a category changes how individuals process that category, even when it has clearly-defined rules for inclusion.

Insight into why this might be the case comes from the Attentional Learning Account (ALA; Smith et al., 2002; Yoshida & Smith, 2005). The ALA posits that infants and young children extract statistical regularities from their environment and then use that knowledge to direct their attention towards future learning. For example, early-acquired words in English often refer to objects that are grouped based on their shape (e.g., ball). This regularity teaches the child to direct their attention towards shape when they learn a novel word. Children who are taught this regularity specifically in the laboratory also show greater vocabulary growth than untrained peers (Smith et al., 2002).


When thinking about the ALA, it is important to discuss the use of the word "attention." Attention can be driven either by the individual (endogenous) or by the environment (exogenous). In the endogenous case, the individual expends effort to focus on specific aspects of the stimulus (Engle & Kane, 2004). Alternatively, the environment can direct an individual's attention to these different aspects. This exogenous case is more similar to the way attention is described in the ALA. As the individual learns that certain features tend to co-occur in a given stimulus (e.g., the name and shape of an object), an instance of one of those features draws attention towards the other. Since the label of a category is perhaps its most frequent feature, it co-occurs most often with other frequent (i.e., typical) features of that category. Thus, the typicality effects seen specifically for category labels may be the result of individuals learning statistical regularities between labels and features.

This type of iterative learning where feature distributions are learned over time closely matches the associative system. In contrast, the hypothesis-testing system is much more focused on selecting one or a few relevant features and discarding those that do not characterize category membership. In fact, many of the categories best learned by the hypothesis-testing system (e.g., *ad hoc* categories) do not have a single-word category label. Thus, a core hypothesis of this dissertation is that category labels affect learning in the associative system but not in the hypothesis-testing system. In the next section, I will discuss how language might play a role in the hypothesis-testing system.

1.3 Executive function and category learning

The hypothesis-testing system involves many executive functions (e.g., selecting and maintaining relevant category rules, inhibiting irrelevant rules). Both inhibitory control and working memory have been shown to be related to rule-based category learning (Rabi & Minda, 2014). In addition, interfering with language resources specifically affects the low-dimensional, rule-based categorization that is best processed by the

hypothesis-testing system (Lupyan, 2009; Minda et al., 2008). This suggests that language is important for this system. Thus, it is possible that language and executive functions work together in the hypothesis testing system.

 Indeed, language ability and executive function have been shown to be related in multiple studies to varying degrees. For example, Figueras et al. (2008) found significant positive correlations between language measures such as vocabulary and receptive grammar and a wide variety of executive function tasks for school-age children. Berninger et al. (2017) found that performance on inhibition and verbal fluency sub-tests of the D-KEFS, a standardized measure of executive function, was correlated with language outcomes in children between the ages of 9 and 15. Children with specific language impairment have also been shown to have some executive function deficits, specifically in updating and inhibition (Im-Bolter et al., 2006). Findings have been more mixed for the nature of the causal relationship between these skills. One study found a strong concurrent relationship between language and executive function longitudinally for children between ages 4 and 9, but no cross-lagged effects, suggesting that language and executive function are not directly influencing each other (Gooch et al., 2016). However, another study found that language ability at 2-3 years predicts executive function at 4 years (Kuhn et al., 2014). Thus, it is possible that the relationship between executive function and language ability changes over development. Regardless, language and executive function at least develop concurrently.

More evidence for the relationship between executive function and language comes from research on adults showing that interfering with verbal resources, usually through articulatory suppression, can negatively impact task switching (A. Baddeley et al., 2001; Emerson & Miyake, 2003). In a task-switching paradigm, performance typically decreases when an individual has to switch between tasks as compared to when they can perform the same task repeatedly. This decrease in performance is known as the switch cost. Articulatory suppression provides verbal interference by having the participant use language-related resources to repeat a nonsense string (e.g. the the the). In 6- and 9-year-old children, articulatory suppression has been shown to impair performance during task-switching but not during a flanker (inhibition) task (Fatzer & Roebbers, 2012).

Interestingly, the negative effect of articulatory suppression on task switching is specific to instances where the individual must represent the task rules internally. For example, if participants must switch between different arithmetic functions such as addition and subtraction, verbal interference does not have an effect when the plus, minus, and equal signs are printed on the page (A. Baddeley et al. (2001). A similar effect is found in a task-switching paradigm where participants must pay attention to different features of a stimulus. When the cue is the whole word (e.g., shape, color, etc.), articulatory suppression has no effect on switch cost. However, when the cue is just one letter (e.g., S, C, etc.), articulatory suppression increases


the switch cost (Miyake et al., 2004). This effect suggests that task switching in these instances require a participant to use language to represent and formulate task rules (Cragg & Nation, 2010). These results indicate that language is important for representing and selecting rules, which may be similar to how the hypothesis testing system learns rule-based categories.

1.4 Interaction between category learning systems

While much research has focused on outlining dual-systems models of categorization, very little research has investigated how these two systems might interact. Both COVIS and the verbal/nonverbal approach suggest that the two systems in a dual-systems model operate in parallel. Stimuli are processed by both systems, but category decisions are made using the faster system or the system with the strongest evidence. However, some research suggests that the hypothesis-testing system may be the default. Behavioral studies encouraging participants to switch between hypothesis-testing and associative strategies in a perceptual category learning task show that unless participants are cued towards which type of strategy to use on a given trial, they tend to use hypothesis-testing strategies for all trials (Ashby & Crossley, 2010; Erickson, 2008). This suggests that the hypothesis-testing system can overpower the associative system when both are equally activated. Still, this line of research requires much more empirical evidence before definitive claims can be made.

1.5 Overview of the current study

This document consists of two experiments, aiming to answer three questions. First, **Experiment 1** investigates the relationship between the associative and hypothesis-testing systems. Almost all category learning studies using a dual-systems model utilize a between-subjects design to avoid transfer effects in learning. Thus, it is still unclear how an individual switches between the systems in response to task demands and stimulus characteristics. To investigate this switching, we test three order effects in Experiment 1 using a task adapted from Kloos & Sloutsky (2008). The first order effect compares switching between systems in matching blocks, where both the stimulus and the task demands push participants towards either the hypothesis-testing or associative system. The second order effect compares switching between the two systems for just dense stimuli, and the third for just sparse stimuli. We use task demands to encourage use of either the hypothesis-testing or associative system. We also test whether these order effects vary according to an individual's language ability. Thus, this experiment can both contribute to how we understand an individual's ability to switch between category-learning systems and assist in within-subjects experimental design with a dual-systems paradigm.

We use **Experiment 2** to answer the two remaining questions. The  experiment is largely designed to test the core hypothesis that the associative system relies on labels while the hypothesis-testing system relies on executive function ability. We test individuals' vocabulary, executive function, and category learning in both systems. In addition, we are the first to directly compare category learning approaches within subjects. We use three different category learning tasks (from the COVIS, statistical density, and taxonomic-thematic approaches) to measure category learning ability. In this experiment, we directly compare performance on these three tasks for each participant to see if these category learning tasks, which have a similar theoretical basis, are indeed comparable.