AUTOMATICITY OF BASIC-LEVEL CATEGORIZATION ACCOUNTS FOR NAMING EFFECTS IN RECOGNITION MEMORY

By

Jennifer Joanna Richler

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Psychology

August, 2010

Nashville, Tennessee

Approved:

Dr. Thomas J. Palmeri

Dr. Isabel Gauthier

Dr. Tim Curran

Dr. Sean Polyn

Dr. Geoffrey Woodman

ACKNOWLEDGEMENT

This work was funded by an NSF grant to the Temporal Dynamics of Learning Center, and by a grant from the James S. McDonnell Foundation to the Perceptual Expertise Network. I would like to thank Justin Barisch, Kara Grubb, Steph Harrison, Sarah Muller and Magen Speegle for data collection, and Emily Merkel for helping to create stimuli.

I would also like to thank all the people with whom I have had the pleasure of discussing various aspects of this work, in particular all members of the Object Perception Lab, Category Laboratory, Perceptual Expertise Network, and my committee, Tim Curran, Sean Polyn and Geoff Woodman. Your comments and suggestions have been sincerely appreciated.

I am particularly indebted to Mike Mack for invaluable discussions and general assistance throughout graduate school. Having Mike as an officemate was an integral part of my ability to succeed.

I am especially grateful to Matt Crump for his unwavering encouragement and support. I owe many ideas presented here to his genuine interest in discussing my research with me.

Of course this work would not have been possible without the wonderful guidance and support provided by my advisors, Tom Palmeri and Isabel Gauthier. I could not have found more caring and thoughtful mentors, and I feel extremely fortunate to have had the opportunity to work with them.

TABLE OF CONTENTS

Page

Appendix

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

"In the fairy stories, naming is knowledge. When I know your name, I can call your name, and when I call your name, you'll come to me".
– Jeanette Winterson (from "Lighthousekeeping")

The ease with which we can name familiar objects belies the complexity of this task. The ability to name all the items on my desk means that I have learned dozens of associations between the visual properties of each object and a series of sounds that I am able to both recognize aurally and produce vocally. This feat is even more impressive considering that, aside from rare cases of onomatopoeic names (e.g., "zipper"), the relationship between objects and their names is purely arbitrary. Furthermore, multiple items on my desk share the same name (e.g., "pen") despite differences in visual properties, and I can call other objects by many different names that reflect a different level of specificity (e.g., "snack" vs. "wasabi peas"). Thus, naming is closely linked to the processes involved in categorization – I am typing on a "PC" and my office-mate is typing on a "Mac", but these both belong to the same category and thus share the name "computer".

So why do we give objects names? Aside from the obvious benefit of language in facilitating communication, having names may actually help us to know what things are. For example, recognition in non-visual modalities (e.g., taste, smell) is dependent on the ability to consistently and accurately name items (Lehrner, 1993; Lumeng, Zuckerman, Cardinal & Kaciroti, 2005), and differences in flavor and odor recognition between children and adults depends on adults' ability to *name* odors and flavors, rather than differences in sensory processing (Lehrner, Gluck & Laska, 1999).

Below I review work that addresses the role of basic-level category names in visual perception and cognition. This literature has traditionally focused on situations where we have names that are not being actively employed during a task, and how adding names to novel stimuli enhances performance. However, recent work has begun to ask whether overt naming influences performance, even when the act of naming is redundant. One study in particular makes an especially provocative claim, suggesting that naming shifts the representation of the object stored in memory toward the category prototype, resulting in reduced performance on a recognition memory task. The goal of my dissertation is to evaluate the representational shift hypothesis and test alternatives to this claim. Ultimately these experiments show that naming does not necessarily shift the memory representation toward the prototype, but rather that naming at the basic-level leads to a less salient and weaker memory representation because it is an instantiation of categorization that is performed relatively automatically[1].

---

[1] There are multiple views of automaticity in the literature. According to one prominent view, "automaticity" refers to the presence of specific properties of performance; tasks that lack one or more of these properties can no longer be considered "automatic" (e.g., Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). An alternative approach is a process view of automaticity that proposes a gradient of automatic processes (e.g., Logan, 1985). The goal of my dissertation is not to test whether naming is or is not an automatic process according to a strict list of criteria, but rather to show that naming is *more* automatic – *more* rapid, effortless and incidental - relative to another task, in this case preference judgements.

What is a "Name"?

It has been suggested that a "label" is whatever feature systematically co-occurs with a category (Colunga & Smith, 2002; Lupyan, McClelland & Rakison, 2007). Because labels are correlated with all the features of the object they label, labels may be treated as pointers to important correlated properties (Billman & Knutson, 1996; Colunga & Smith, 2005) and may help reduce the variability in category representations when exemplars are initially very different (Lupyan, 2005). Indeed, in infant category learning associating visually dissimilar objects with a single label promotes the formation of a single category (Plunkett, Hu & Cohen, 2008), while presenting two objects with two different labels promotes individuation (Xu, 2002; Xu, Cote & Baker, 2005). One neurocomputational model of infant visual categorization based on self-organizing maps suggests that these effects arise because the label is an additional feature that is common to all objects in the category, thus increasing the similarity of objects from the same category (Gliozzi, Mayor, Hu & Plunkett, 2009). In this way infants may treat labels as additional features that are processed in the same manner as other features; correlations between labels and objects impact performance in the same way as correlations between visual features of the objects.

Although some models of adult recognition memory are consistent with this idea and represent category labels in the same way as other category features (e.g., Anderson, 1990; Hintzman, 1986), many studies suggest that category labels have a special status. For example, Yamauchi & Markman (2000) showed that during inference learning where participants had to predict the value of a missing feature of an item given its category label along with its other visual features, participants made responses consistent with the category label despite contradictory similarity information from the other features of the stimulus. Furthermore, this effect was stronger when labels denoted the category name as opposed to the name of one of the features. Similar effects have also been observed in infants, where category labels guide inductive reasoning more than any other single object feature, overriding information about perceptual similarity (Dewar & Xu, 2009; Keates & Graham, 2008).

There is also evidence that *verbal labels* in particular have unique properties relative to other kinds of labels. Although infants initially accept a broad range of auditory signals in communicative contexts, this range narrows with development, such that by 20-months of age sound-object associations are no longer learned as well as word-object associations (Woodward & Hoyne, 1999). In fact, many developmental studies show that category formation is facilitated when a name is paired with objects from the same category compared with conditions where other auditory information (e.g., a tone) is paired with objects from the same category (e.g., Balaban & Waxman, 1997; Colunga & Smith, 2002; Fulkerson & Waxman, 2007; Plunkett, Hu & Cohen, 2008; Waxman & Markow, 1995; Woodward & Hoyne, 1999; Xu, 2002; Yoshida & Smith, 2005; but see Booth and Waxman, 2002, for evidence that object functions are as effective as verbal labels in infant category learning).

The distinctions between “label”, “verbal label” and “name” are often blurry. Here I use the term *label* to refer to any feature that systematically co-occurs with a category (Colunga & Smith, 2002; Lupyan et al., 2007), such as a sound, a colour or a location in space. A *verbal label* is a label that takes the form of a word that can be written and spoken. A *name* is a verbal label by which an object is designated, called, or known at the category-level. Thus, names and verbal labels are to some extent synonymous. However, a further distinction must be made between "labeling" and "naming". An object can be *labeled* in an infinite number of ways. For example, the same chair might be labeled as "the red one" or "mine". However, the *name* of the object always remains "chair". Although there can be names at different levels of abstraction (e.g., "fruit" vs. "apple" vs. "Granny Smith"), in general naming is more constrained than labeling, and people tend to use basic-level names when naming objects that are typical exemplars of the basic-level category (e.g., Jolicoeur, Gluck & Kosslyn, 1984).

Having Names

There is evidence to suggest that one consequence of having names for objects is that they are automatically accessed. In some situations this can be beneficial. For example, latencies to name a target picture are shorter when an accompanying distractor picture shares the same name or has a name that is phonologically related (e.g., *bell* vs. *belt*; Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005). Conversely, in visual search tasks search efficiency is reduced when distractors share the same name as targets but are not semantically related (e.g., baseball *bat* vs. the animal *bat*; Meyer et al., 2007). Thus, object names appear to become rapidly activated regardless of how this will ultimately impact performance.

This automatic access to names has consequences for perception and perceptual decision-making. As early as one year of age enhanced oscillatory gamma-band activity (measured with ERPs) is observed over visual cortex in response to objects for which the infant is familiar with the name compared with familiar objects for which the infant does not have a name (Gliga, Volein & Csibra, in press)[2]. In an fMRI study with adults, Tan and colleagues (2008) had participants discriminate between easy-to-name and difficult-to-name colour patches and found that perceptual discrimination for nameable colours evoked activation in regions involved in the colour naming localizer task. These results suggest that names and language can modulate perception and play a role in perceptual decisions when this information is available.

Such neural data is consistent with behavioral work on categorical perception, which refers to the finding that discrimination between two perceptually similar stimuli is easier when the stimuli cross a category boundary (see Davidoff, 2001; Masharov & Fischer, 2006; Ozgen, 2004; for reviews). Importantly, in studies using familiar stimuli the category boundary is determined by linguistic category, not perceptual similarity (Gilbert, Regier, Kay & Ivry, 2006; 2008; Goldstein & Davidoff, 2008; Roberson & Davidoff, 2000; Roberson, Davidoff & Braisby, 2000; Roberson, Pak & Hanley, 2008; Winawer, Witthoft, Frank, Wu, Wade & Boroditsky, 2007). For example, given colour patches arranged in a spectrum from blue to green, discrimination is fastest for a pair of stimuli when one stimulus

---

[2] In adults the amplitude of this activity is correlated with familiarity (e.g., Busch, Herrmann, Muller, Lenz & Gruber, 2006; Gruber, Muller & Keil, 2002). However, for adults, objects which are familiar generally also have names, so these two contributions to the enhancement of gamma-band activity are confounded.

is categorized as “blue” and the other stimulus is categorized as “green” despite equivalent distances in actual colour-space between the pairs of stimuli. One explanation for categorical perception is that items in the same category are represented more similarly due to extensive practice categorizing these stimuli together (e.g., Goldstone, 1994; Harnad, 1987).

However, a recent study showed that belonging to the same conceptual category does not necessarily lead to differences in perceived similarity, as search for a novel target was faster when non-targets were from the same linguistic category (e.g., non-targets were “B” and “b”) compared to different categories (e.g., “B” and “p”), but perceptual judgements were equally fast for B-b and B-p pairs (Lupyan, 2008b). These results suggest that effects that arise when objects are from the same linguistic category are not due to long-term changes in the perceptual representation, where objects from the same category are represented as perceptually more similar. Nor is it sufficient that objects belong to different semantic categories. What is key is that these distinct semantic categories have different names. Indeed, the advantage in discrimination for cross-category pairs is reduced when participants perform a concurrent verbal task (e.g., holding a word in working memory), but not a concurrent visual task (e.g., holding a visual pattern in working memory; Gilbert et al., 2006; 2008; Roberson & Davidoff, 2000; Winawer et al., 2007), implying that the ability to utilize category names is necessary for categorical perception in humans[3].

---

[3] Categorical perception has also been reported in animals (e.g., Maki et al., 2001; Wilson & DeBauche, 1981), in particular categorical perception of sound signals (e.g., Ehret, 1987; Morse & Snowdon, 1975; Nelson & Marler, 1989; Wyttenback et al., 1996). Although categorical perception in

Supporting this claim, studies using visual search arrays where the target and distracters come from either the same or different linguistic category find larger categorical perception effects when the target is presented in the right visual field (Gilbert et al., 2006; 2008) for rapid responses (Roberson et al., 2008). This right visual field advantage in categorical perception is presumed to arise because items presented in the right visual field have faster access to language processing in the left hemisphere. Interestingly, categorical perception of colours in toddlers is initially stronger for stimuli presented to the left visual field, but the right visual field advantage emerges once the toddlers have acquired the relevant colour terms (Franklin, Drivonikou, Bevis, Davies, Kay & Regier, 2008; Franklin, Drivonikou, Clifford, Kay, Regier & Davies, 2008).

In addition to simply having names, declarative knowledge of names may influence perception. Mitterer and colleagues (2009) had Dutch and German participants categorize yellow and orange hues. These hues were presented on pictures of four common objects: a banana (prototypically yellow), a carrot (prototypically orange), a sock (with no strong colour associations) and the center circle of a traffic light. Critically, the center light of a traffic light is called "yellow" in German and "orange" in Dutch, even though the objective colour of the center light of a traffic light is the same in both countries. If an ambiguous yellow-orange hue was presented, both groups were more likely to classify the hue as yellow if it was

---

this case cannot be attributed to *verbal* labels, categorical labeling is still believed to underlie the effect. For example, if a given dimension is most relevant for identifying a type of sound signal, selective attention to that dimension results in that dimension being treated as a "label" for that type of signal (Ehret, 1987). Alternatively, a "label" might take the form of the response: one sound might signal approach, while another sound signal might signal flee (Wyttenback et al., 1996). Names might be especially critical in human categorical perception because when language is available naming might become automatic, especially in experimental settings (Roberson & Davidoff, 2000).

presented on the prototypically yellow stimulus (and vice versa for orange). However, if an ambiguous hue was presented on the traffic light, Dutch participants were more likely to classify the stimulus as orange, and German participants were more likely to classify the stimulus as yellow. Therefore, it is not just having different category names, but also declarative knowledge of the category name within a particular context, that can influence perception.

Assigning Names to Novel or Unfamiliar Objects

The studies reviewed above demonstrate that simply having names can influence processes that were believed to be purely perceptual, possibly because names are automatically accessed. Consequently it is perhaps not surprising that similar advantages in perception for novel or unfamiliar objects can be gleaned by assigning names to them. Indeed, categorical perception arises after categorization training with novel stimuli (e.g., Goldstone, 1994). Because in these training studies response keys are uniquely associated with each category, participants are in effect learning verbal labels (i.e., the name of the response keys) during training in addition to learning the categories themselves.

A particularly fascinating demonstration that categorical perception for novel stimuli depends on the presence of names comes from a recent paper by Kikutani, Roberson & Hanley (2008). First they show that categorical perception for a morph line of face identity in a two-alternative forced-choice sequential matching task arises when the endpoints of the morph line are known famous faces, but not when the faces being used to create the morph line are unfamiliar. However, when participants passively view the unfamiliar faces *with names* for one minute prior to the matching task, categorical perception is observed. Critically, familiarizing participants with the unknown faces *without names* prior to the experimental task does not produce this effect. Therefore, briefly associating stimuli with names is sufficient to influence perceived similarity. Similarly, Lupyan and Spivey (2008) found that visual search for a novel target among very similar novel distractors was more efficient when participants were instructed to think of the target and distractors as rotated digital 2's and 5's, demonstrating that ascribing familiar names to novel stimuli improves visual processing by making perceptually similar stimuli appear more different.

Adding names to novel stimuli can also impact cognitive processes beyond the level of perception. For example, the presence of names can promote learning of novel categories even when the names are redundant and not necessary for the learning task. Lupyan, Rakison and McClelland (2007) had participants categorize novel "aliens" as those that should be approached or avoided. Critically, the responses being made in the training task were whether to move toward or away from the alien, so the response keys were not correlated with the categories[4]. After corrective feedback, a name for the category the alien belonged to was presented, or the alien moved to one side of the screen, representing the location where that category of alien lived. Participants who viewed a category name learned the categories faster and showed better transfer to new aliens not viewed during

---

[4] Direction of movement depended on the location of the alien and the symbol representing the participant on the screen, such that both left and right movement could be either moving toward or away from the alien. Consequently, "left" and "right" could be the correct response for either category.

training. A similar advantage for names over non-verbal associative information was found for subjects learning to classify event sequences (Cabrera & Billman, 1996). These results indicate that, similar to the developmental studies briefly described earlier, verbal category labels influence category formation more so than other non-verbal semantic information, even when this non-verbal information is perfectly correlated with category membership.

Assigning a familiar name to a novel stimulus may also improve recognition memory by adding an additional distinguishing feature to the stimulus. Supporting this hypothesis, Musen (1991) asked participants to either generate a meaningful name or count the number of lines during encoding of random line patterns, and found that recognition memory was better for objects for which names were generated during encoding. On the other hand, Koutstaal and colleagues (2003) found that presenting stimuli from categories of ambiguous shapes with disambiguating category names (i.e. all the stimuli from the same ambiguous category were given the same name) led to worse recognition memory performance due to increased rates of false recognition (incorrectly responding that a new item was previously studied). They suggested that this increase in false memory for named items occurred because semantic category information provided by the name attenuated or truncated encoding of item-specific perceptual information. Therefore, assigning familiar names to novel objects can have different effects on recognition memory depending on how many items share the same name, similar to the fan effect where retrieval is impaired due to interference when multiple items are associated with the same concept (Anderson, 1974; Anderson & Reder, 1999; Reder & Anderson, 1980; Reder, Donavos & Erickson, 2002).

Calling Objects by their Names

While it is clear from the above review that names exert a unique influence on how object categories are learned and perceived, what is less known is how overtly using names we already have influences cognitive processes. Certainly it may seem that actively naming should not influence performance beyond what happens when we simply *have* names, particularly given evidence that naming is automatic. For example, memory errors for ordered recall of briefly presented items tend to correspond to auditory confusions (e.g., recalling “B” instead of “D”) for letters (Conrad, 1964), words (Coltheart, 1993) and pictures (Coltheart, 1999; Sciano & Watkins, 1981), but this phonological similarity effect is eliminated if subjects are engaged in an irrelevant articulatory task (e.g., repeating a word or counting; Coltheart, 1993; Sciano & Watkins, 1981). Similarly, phonological similarity between objects presented in an array impairs performance at detecting whether objects are subsequently presented in new locations (Mondy & Coltheart, 2006), and concurrent verbal tasks (e.g., verbal shadowing) reduce accuracy for detecting an identity change but not a configuration change in change detection experiments (Simons, 1996), suggesting that participants automatically use names to encode objects in scenes.

A particularly elegant demonstration that naming is automatic was provided by Zelinsky & Murphy (2000) who showed that the amount of time participants look at objects is directly related to the length of the name of the objects. They presented participants with displays containing 4 line drawings that were equated in visual complexity and typicality. Critically, the names for two of the objects were one-syllable in length, while the names for the other two objects were three-syllables in length. After the participant terminated the study display, a single item was presented, and participants had to judge whether that object had been presented in the preceding display. Although participants were not asked to name the objects during study, subjects made more fixations and had longer gaze durations for objects with three-syllable vs. one-syllable names. Moreover, there was a significant correlation between spoken name duration and object gaze duration: objects that take longer to name aloud were looked at longer. Interestingly, a similar effect was obtained by Noizet & Pynte (1976) in a task where participants were asked to simply recognize each object silently – there was no explicit response requirement, and objects did not need to be held in memory. Together these results suggest that participants automatically code objects verbally regardless of task demands.

Surprisingly, despite the fact that participants are likely implicitly accessing and generating object names, there is evidence that *explicitly* providing object names influences performance. For example, visual search efficiency is improved when participants hear the name of a familiar target or distractor during the onset of the search display, even though participants already know the identity of the target and distractors for that trial (Lupyan, 2008b; Spivey, Tyler, Eberhard & Tanenhaus, 2001). This is a case where the experimenter is providing the object name to participants. What happens when participants themselves overtly and volitionally name objects? Although some studies have used naming as an incidental task to ensure that participants are paying attention to presented items (e.g., Seamon, Luo, Schlegal, Greene & Goldenberg, 2000), these studies do not make comparisons between naming and other encoding tasks. In fact, to date only two studies have explored this question. In these studies participants named familiar items with their basic-level names either prior to making a recognition memory judgement (Koutstaal & Cavendish, 2006), or immediately following encoding during the study phase (Lupyan, 2008). These two studies are described in detail below.

Koutstaal and Cavendish (2006) had participants view objects and make judgements about object size, then complete three blocks of recognition memory tests. Participants were required to make recognition memory judgements at either the conceptual (same category) or identical (same object) level, and the critical manipulation was whether the type of memory judgement in the third block was the same or different as the previous two blocks. They found that performing two blocks of conceptual old/new recognition led to an impairment at identical recognition (CC → I condition) compared to the case where participants performed three blocks of identical recognition (II → I condition). In other words, performing two blocks of identical recognition facilitated identical recognition in a third block. More interestingly, and critical to the current discussion, this difference was eliminated when participants named objects aloud prior to making their recognition memory response. The interpretation of this result was that naming enhanced recognition memory by introducing new information that could help isolate the relevant items in memory.

However, closer inspection of the data for the experiment with naming vs. the experiment without naming suggests that the difference between CC → I and II → I conditions was eliminated with naming due to both an improvement in performance in the CC → I condition and a decrease in performance in the II → I condition, which is inconsistent with the idea that naming enhanced recognition memory - recognition memory performance actually *decreased* in the II → I condition when objects were named. Moreover, Koutstaal and Cavendish do not provide a comprehensive account for why naming exerted this effect on performance - it is unclear how naming would improve discrimination between targets and lures when targets and lures were from the same basic-level category that shared the same name. Nevertheless, more convincing evidence that naming might facilitate access to category-level information and increase resistance to forgetting was demonstrated by a comparison of performance in the CC → C condition between experiments with and without naming at test. In this analysis, the authors showed that a decrease in performance across subsequent conceptual recognition blocks was significantly reduced when items were explicitly named at test. In other words, naming objects at test facilitated memory for whether an object sharing the same name was presented at study.

Naming objects at test may help isolate cues that facilitate memory for object representations that possess category-relevant features (Koutstaal & Cavendish, 2006). In contrast, naming at study may modulate memory by influencing how the object is encoded and thus the nature of the stored representation itself. Lupyan (2008) explored how overtly naming an object with its basic-level name at study influences recognition memory performance and obtained a particularly provocative result that will be the focus of my dissertation experiments. In Lupyan (2008) pictures of chairs and lamps were presented briefly (300ms) and then participants were asked to press a key denoting the name of the object ("chair" vs. "lamp"), or a preference judgement ("like" vs. "don't like"). Each study item was matched with a critical lure that differed from the study item in subtle details such as the presence or absence of a feature (e.g., arm rests), or colour. All targets and lures were presented in a subsequent surprise recognition memory test where participants decided whether each presented item was old or new. Recognition memory performance was worse for objects that were named at study. Critically, this difference in recognition memory performance was driven by a difference in hit rates, where hit rates were lower for objects that were named, without a difference in false alarm rates. Lupyan (2008) also collected typicality ratings for his stimuli and found that hit rates were lower for objects that were rated as being more typical, and that this relationship between hit rates and typicality was stronger for named objects compared with objects for which preference judgements were made.

Lupyan (2008) proposed a representational shift hypothesis to account for these results. According to this hypothesis, activating category names produces top-down feedback that activates the visual features stored with the category from previous occasions. These features become coactive with the features activated by bottom-up processing of the recently viewed item. In other words, the bottom-up representation of the object that was just seen is altered by top-down feedback invoked by the category label. The result is that when the same object is viewed during the recognition memory task, the object no longer matches the modified representation stored in memory. This study-test mismatch leads to a false sense that the old object is new, and thus a lower hit rate. This effect is modulated by typicality because typical objects have more features in common with the category prototype, and thus more features that can be modified when the prototype is activated by the category name.

Specific Aims

It is certainly interesting that overt naming appears to influence performance, even when these names are not adding extra information that can help perform the task. Moreover, the representational shift hypothesis is very provocative and intriguing, and may explain the results of studies described above where hearing the name of the target or distractor facilitates visual search: top-down feedback activated by the basic-level name may feed back onto lower-level visual representations, allowing more rapid perceptual processing of objects that are named (Lupyan, 2008b). The representational shift hypothesis also suggests that effects that depend on having access to names such as categorical perception may be modulated by overtly using category names during the task. More generally, the representational shift hypothesis, if supported, has implications for any cognitive processes that rely on object representations stored in long-term memory.

However, the representational shift hypothesis it is not unequivocally supported by the data, nor is it the most parsimonious explanation of the data.

Experiment 1 will provide a replication of Lupyan (2008) and Experiment 2 will test whether this effect generalizes to a situation where all objects have unique basic-level names that must be generated by the participant. In addition, both Experiment 1 and Experiment 2 test whether the effect reported by Lupyan (2008) is contingent on the type of lures being used. The remaining experiments (Experiments 3-6) test alternatives to the representational shift hypothesis that can account for reduced recognition memory performance (driven by a difference in hits) for objects that are overtly named. These alternatives are described in more detail below.

The representational shift hypothesis is centered on the effect of overt naming. Certainly it is difficult to dissociate the effects of naming and categorization because every act of naming is also an instance of categorization (Lupyan, 2005). This difficulty is evident in studies of categorical perception, where effects were presumed to arise because of categorization experience (e.g., Goldstone, 1994; Harnad, 1987) but access to category names has been shown to be a critical factor (e.g., Gilbert et al., 2006; Kikutani et al., 2008). When categorization responses and category name are dissociated, names have been shown to uniquely influence performance (Lupyan et al., 2007), but in Lupyan (2008) naming and categorization are confounded: participants are classifying objects as chairs or lamps by pressing response keys. Therefore, the results of this study cannot be unequivocally linked to overt naming - representational shift might arise due to basic-level categorization on its own. That is, it may not be the act of using the category name that is critical, but simply that objects are being categorized and thus considered at the category-level, which does not require attention to details that may facilitate later recognition when lures are highly similar to targets. This possibility is tested in Experiment 3 and Experiment 4 where tasks that involve basic-level categorization without naming are performed during the study phases.

A second criticism is that while representational shift is an intriguing and arguably more interesting hypothesis, simpler explanations for the naming effect have not been ruled out. One possibility is that the naming effect may be explained by a depth of processing account, where items presented during tasks that are more cognitively demanding are better recalled (Craik, 1979; Craik & Lockhart, 1972; Craik & Tulving, 1975). In the terminology of depth of processing, naming may lead to shallow processing because participants automatically name objects (e.g., Coltheart, 1999; Schiano & Watkins, 1981; Zelinsky & Murphy, 2000), so the overt naming response does not require additional effort. Thus, differences in performance between naming and preference may be due to an *enhancement* in performance when a more effortful preference judgement is made. Experiment 5 will support the depth of processing account by showing a gradient of recognition memory performance in this paradigm based on processing demands when a third study task is included. Experiment 6 will also support a depth of processing account by showing that memory performance is improved by the addition of the preference judgement, but is not reduced by the addition of a naming judgement, when multiple judgements are made at study.

It is important to note that the levels of processing framework has been heavily criticized (e.g., Baddeley, 1978; but see Lockhart & Craik, 1990) and transfer appropriate processing has been presented as a more general framework for understanding memory effects. According to transfer appropriate processing, memory performance depends not only on the nature of the encoding or study tasks, but also on the compatibility between study and test tasks (e.g., Morris, Bransford & Franks, 1977). In this view, naming may lead to worse recognition memory performance than preference judgements because the preference task may be better matched with the type of memory judgements that are required in this paradigm, where subtle differences between object features are relevant and category information is not ultimately informative. Although there are important distinctions between depth of processing and transfer appropriate processing accounts of memory effects, they cannot be distinguished in the present experiments where only encoding tasks are manipulated. Therefore, while I will focus on depth of processing, contributions to performance based on transfer appropriate processing are not ruled out. Importantly, the goal here is not to distinguish between depth of processing and transfer appropriate processing and determine which theory better accounts for the naming effect; the aim is to demonstrate that the effects attributed to representational shift following overt naming can be accounted for by more general principles of memory performance.

In sum, the following experiments will show that: the effect reported in Lupyan (2008) does not depend on the number of objects sharing the same name or the type of lures used (Chapter 2), reduced recognition memory performance is due to basic-level categorization as opposed to overt naming (Chapter 3), and recognition memory is impaired after basic-level naming because naming at the basic-level occurs relatively automatically, and thus leads to less processing than the more demanding preference task, consistent with a depth of processing account (Chapter 4).

CHAPTER II

REPLICATION & EXTENSION

Experiment 1

The purpose of Experiment 1 was to ensure that the main result reported in Lupyan (2008) could be replicated. Experiment 1 will also be used as a basis of comparison for some later experiments. To this end, Experiment 1 was nearly identical to Lupyan (2008), with the exception that lures that were not matched to study items were also included because there is some evidence that false alarm rates to similar lures increase with hits more so when dissimilar lures are also tested (Malmberg, 2008). Thus, Experiment 1 also tests whether the inclusion of non-critical, dissimilar lures differentially impacts false alarm rates for named objects vs. objects for which preference is rated.

Method

Participants

Twenty members of the Vanderbilt community (11 male; mean age 23.4 years) were given monetary compensation for participation. Data from two participants were discarded due to timing out on over 20% of the study trials, data from a third participant were discarded due to accuracy less than two standard deviations from the mean on the naming task, and data from a fourth participant were discarded due to below chance performance in the memory task.

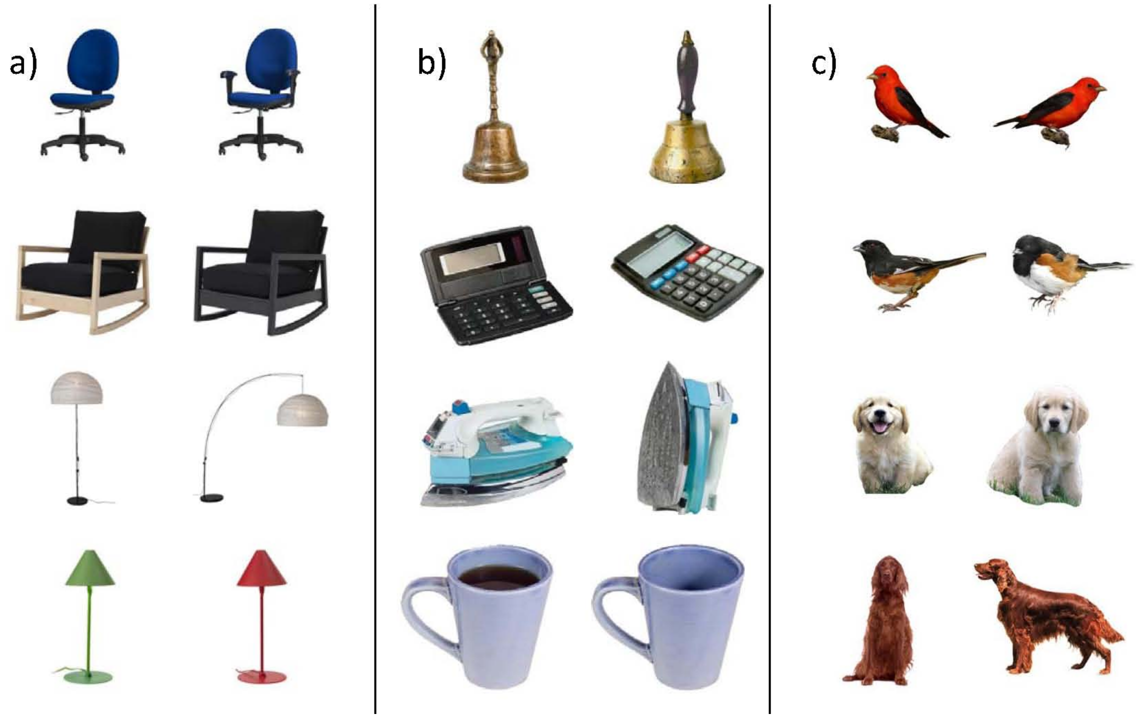

Figure 1. Examples of target-lure pairs used in a) Experiments 1, 4-6, b) Experiment 2, and c) Experiment 3. The top two examples show a) chairs, b) exemplar-change pairs, or c) birds. The bottom two examples show a) lamps, b) state-change pairs, or c) dogs.

Stimuli

Stimuli were 100 colour pictures of chairs and lamps (50 pictures for each category) that were downloaded from the IKEA online catalogue (www.ikea.com). Each picture was 250 x 250 pixels in size and showed a single chair or lamp on a white background. There were 40 pairs of chair and lamp pictures (20 for each category), such that each target chair or lamp was matched with a critical lure. Paired lures differed from targets in small but noticeable ways. For example, paired lures might be a different colour from the target, differ on a feature such as the presence or absence of arm rests, or have a different height-width ratio (see Figure 1a for examples). Pictures were randomly sorted into two sets, each containing ten target-lure chair pairs and ten target-lure lamp pairs. The 20 remaining pictures (ten from each category) were only presented at test as unrelated lures.

Procedure

There were two parts to the experiment. During the study phase participants were told that they would see pictures of chairs and lamps and be asked to make judgements about them as quickly as possible. Although they were not explicitly told that there would be a memory test, they were instructed to pay careful attention and try to remember as much as possible about each picture. Pictures were presented for 300ms followed by a question mark. Participants were told that responses would only be accepted when the question mark was on the screen, and that they should respond as quickly as possible. A tone was played if a response was not made within 700ms, at which point the trial timed out. These trials were not included in the study phase analyses.

There were two study tasks. In the naming task, participants pressed '1' if the object in the picture was a chair, and '2' if the picture was a lamp. In the preference task, participants pressed '1' if they liked the object in the picture, and '2' if they disliked the object in the picture. The response-key mapping was kept constant for all participants. Tasks were blocked and the object set assigned to each task was counterbalanced across participants. Task blocks were repeated twice during the study phase. Task order and which object in a pair was the target or lure was also counterbalanced across participants. Prior to the experimental tasks participants were familiarized with the pace of the experiment with five practice trials. On the practice trials, participants saw pictures of different tables and pressed '1' if the table was round, and '2' if the table was square.

After the study phase participants performed a surprise recognition memory task. They were informed that some of the pictures would be exactly the same as those they saw during the study phase, some of the pictures would be new but very similar to the pictures they saw before, differing only subtly in details like shape or colour, and some of the pictures would be brand new. Pictures were presented on the screen one at a time and participants were instructed to press '1' if the object was 'old' and the exact same picture they saw before, and '2' if the object was 'new'. Pictures remained on the screen until participants made a response. All 100 objects were presented in the recognition memory phase in random order.

Typicality Ratings

Typicality ratings were collected from a second group of ten participants from the Vanderbilt Community (6 male; mean age 22.6 years). Each item was presented one at a time and participants were instructed to rate the typicality of each chair/lamp on a 5-point scale (1 = very typical, 5 = very atypical). Pictures remained on the screen until a response was made. Object category was blocked, with each object presented twice during the block, and category order was counterbalanced across participants.

Table 1.
Correct Response Times (and Standard Deviations) Recorded from the Onset of the Response Probe During the Study Phase for Each Study Task in Experiments 1-5

	Study Task	Response Times (ms)
Experiment 1	Naming	230.43 (102.42)
	Preference	338.24 (113.88)
Experiment 2	Naming	1002.91 (111.47)
	Preference	1043.62 (193.11)
Experiment 3	Naming	174.04 (76.76)
	Preference	267.90 (94.98)
	Category Induction	175.48 (64.09)
Experiment 4	Category-Matching	315.94 (143.00)
	Exemplar-Matching	323.95 (137.00)
Experiment 5	Naming	233.52 (74.16)
	Preference	347.82 (66.98)
	Colour Judgement	253.92 (75.05)

Results

Study Phase

Average accuracy on the naming task was 91.77% (*SD* = 5.31). Because there is no correct response for preference judgements, a reliability score was calculated for both the preference and naming tasks, which measures the percentage of times where the same response was given for both presentations of the same object. Reliability of preference responses was significantly greater than chance (t_{15} =

6.187, $p < .001$) indicating that responding in this task was not random. However, naming responses were significantly more reliable than preference responses (86.77% vs. 78.46%; $t_{15} = 2.545$, $p < .05$). Correct naming responses (recorded from the onset of the response probe) were also significantly faster than preference judgements ($t_{15} = 6.291$, $p = .001$; see Table 1).

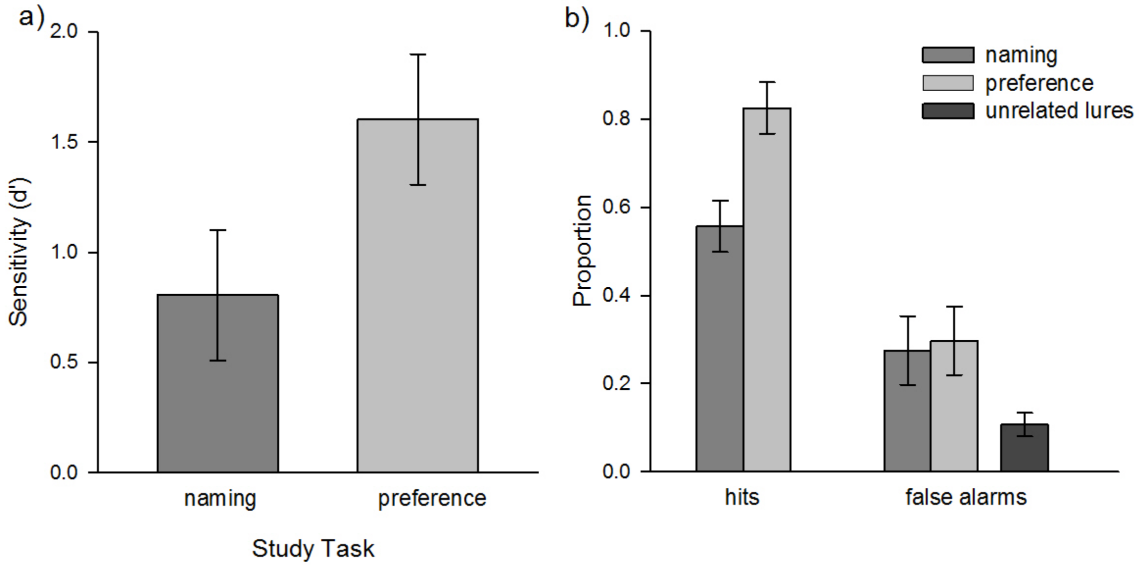

Figure 2. a) Overall memory performance (d') and b) hit and false alarm rates on the recognition memory test for objects that were named vs. objects for which preference was rated in Experiment 1. Error bars show 95% confidence intervals for the paired-sample *t*-test. Error bars for false alarms for unrelated lures are standard error of the mean.

Test Phase

Performance on the recognition memory test is shown in Figure 2. Paired-sample *t*-tests confirmed that the naming effect observed by Lupyan (2008) was replicated: overall performance was better for objects for which preference judgements were made vs. objects that were named ($t_{15} = 5.736$, $p < .001$), and this difference was driven by a difference in hit rates ($t_{15} = 9.830$, $p < .001$). A one-way repeated measures ANOVA on false alarm rates (naming vs. preference vs. unrelated lures) was significant ($F_{2,30} = 19.911$, $p < .001$), however post-hoc tests revealed that this effect arose because false alarms for named objects and preference objects were both significantly higher than false alarms to unrelated lures (ps < .001), but, critically, did not differ from each other. A similar repeated measures ANOVA on correct response times (RTs) revealed no differences in RTs between named objects, objects for which preference was rated, and unrelated lures (see Table 2).

Table 2.
Correct Response Times (and Standard Deviations) on the Recognition Memory Test for Objects Presented in Each Study Task in Experiments 1, 3-5.

	Study Task	Response Times (ms)
Experiment 1	Naming	1125.69 (237.24)
	Preference	1102.78 (243.17)
	Unrelated Lures	1133.82 (285.59)
Experiment 3	Naming	1343.50 (538.24)
	Preference	1303.05 (616.22)
	Category Induction	1248.82 (459.12)
Experiment 4	Category-Matching	1231.04 (435.55)
	Exemplar-Matching	1080.33 (314.24)
Experiment 5	Naming	1135.66 (279.59)
	Preference	1095.66 (223.09)
	Colour Judgement	1158.48 (247.62)

Typicality Effects at Test

Average typicality ratings for chairs and lamps was 3.00 and 2.68, respectively. The difference in typicality ratings for chairs and lamps was not significant.

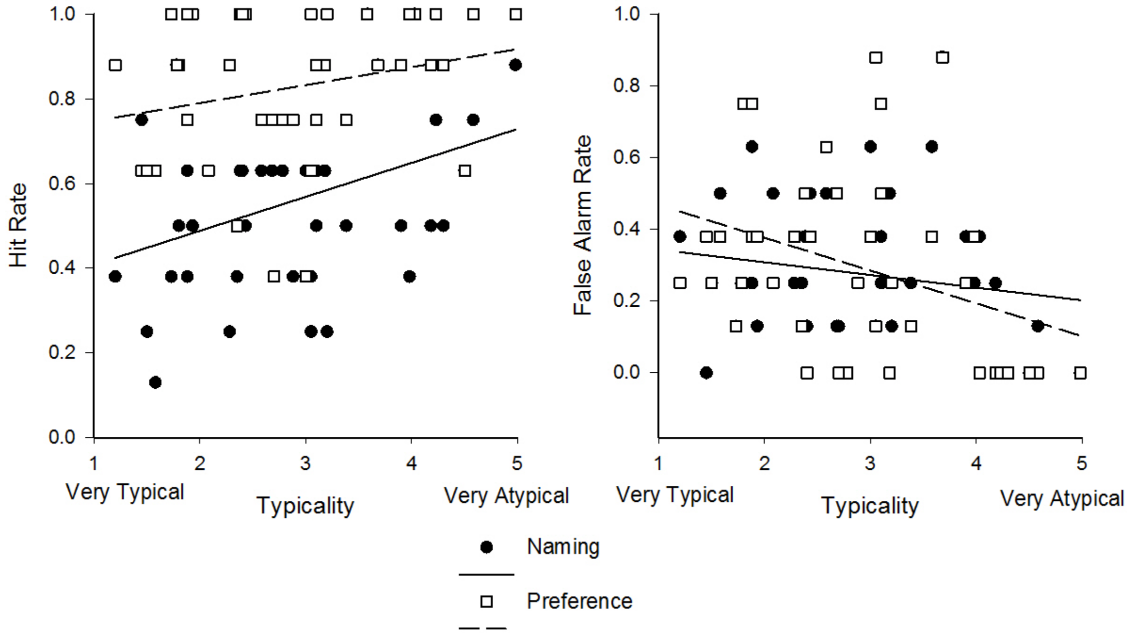

Figure 3. Correlations between Typicality and Hit Rate (left panel) and correlations between Typicality and False Alarm Rate (right panel) for objects that were named (black circles, solid line) and objects for which preference was rated (white square, dashed line) at study in Experiment 1.

Correlations between typicality and hit rate (see Figure 3) revealed that overall hit rates were higher for atypical vs. typical objects (r_{40} = .422, *p* < .01). The correlation between typicality and hit rate was significant for named objects (r_{40} = .379, *p* = .016) but not for objects for which preference was rated (r_{40} = .236). However, as can be seen in Figure 3, the lack of correlation for objects for which preference was rated may be due to a ceiling effect. In contrast to the results reported by Lupyan (2008), when typicality was entered as a covariate in a general linear model to predict hit rates the analysis revealed that the interaction between typicality and study task was not significant ($p > .3$). Thus, the relationship between typicality and hit rates was not significantly different between study task conditions.

The correlation between typicality and false alarm rate (see Figure 3) approached significance ($r_{40} = -.307$, $p = .054$), such that there was a trend toward higher false alarm rates for typical objects. Although the correlation between false alarm rate and typicality was significant for objects for which preference was rated ($r_{40} = -.343$, $p < .05$) but not named objects ($r_{40} = -.168$), when typicality was entered as a covariate in a general linear model to predict false alarm rates the interaction between typicality and study task was not significant ($p > .2$).

Discussion

Experiment 1 partially replicated the main results of Lupyan (2008): recognition memory was worse for objects that were named vs. objects for which preference was rated, and this difference was driven by a decrease in hits for named objects, with no difference in false alarms. Not surprisingly, rates of false memory were higher for matched lures compared with the baseline false alarm rate (i.e., false alarms to non-matched, unrelated lures). Importantly, the inclusion of these non-matched lures did not lead to differences between the results of Experiment 1 and the experiments in Lupyan (2008). Therefore, unrelated lures are not included in the remaining experiments. Naming and preference responses not only differed in how they influenced subsequent memory, but also differed in how fast and reliable they were during the study phase. Naming responses were significantly faster and more reliable than preference judgements. The implications of these differences in study phase performance are considered in Chapter 4.

Unlike Lupyan (2008), the relationship between typicality and hit rate did not differ between named objects and objects for which preference was rated, although there is a trend in this direction (the correlation is significant for named objects but not preference objects). Effects of typicality at test for all subsequent experiments are reported in Appendix A and are considered further in the General Discussion.

Experiment 1 established that the difference in recognition memory between objects that are named and objects for which preference is rated at study is replicable. Because Experiment 1 was nearly identical to the main experimental design of Lupyan (2008), these data will be used as a comparison for later experiments to determine whether the number of basic-level categories in the stimulus set influences recognition memory performance in this paradigm (Experiment 2) and whether different encoding tasks produce the same quantitative effects as naming (Experiment 4).

Experiment 2

In the experiments reported by Lupyan (2008), only two categories of objects were used. However, there is evidence that the number of items that share the same name can influence recognition memory performance. For example, in studies where familiar category names are assigned to novel stimuli, recognition memory is enhanced when a unique name is assigned to each novel object (Musen, 1991), but recognition memory is impaired when multiple novel objects share the same name (Koutstaal et al., 2003). In Experiment 2 all objects will come from different basic-level categories to test whether the naming effect generalizes to conditions where objects from multiple categories are presented at study.

In addition, because only two categories of objects were used in the experiments reported by Lupyan (2008), the alleged naming task may essentially be a categorization task that does not tap into overt naming at all - participants were given two response keys and were told to classify the items. In Experiment 2 because each study object will be from a unique category, participants will respond by typing the first letter of the name of the object. As such, Experiment 2 will also test whether the naming effect in recognition memory arises when participants are required to actually generate basic-level names.

Finally, Lupyan (2008) suggests that because naming study objects does not influence false alarm rates, this effect cannot be driven by basic-level categorization (see Chapter 3). Nonetheless, there are several procedural differences between the traditional false memory experiments (e.g., Koutstaal & Schacter, 1997; Sloutsky & Fisher, 2004) and the paradigm used by Lupyan (2008). Most notably, Lupyan (2008) used lures specifically paired with targets, rather than lures that were simply categorically related to the target. Such a difference may impact whether and how different encoding tasks influence recognition memory performance. Indeed, Marks (1991) found that conceptual encodings were more effective as lures became less similar to the studied pictures. Thus, Experiment 2 also tests whether the type of lure being used influences rates of false memory or general recognition memory performance in this paradigm by including both matched lures like those used in Lupyan (2008) and Experiment 1 and lures that are just categorically related to the targets.

Method

Participants

Seventeen Vanderbilt University undergraduates (2 male; mean age 19.1 years) participated in exchange for course credit. Data from one participant were discarded for naming accuracy that was more than two standard deviations below the mean.

Stimuli

Stimuli were 160 pictures of objects from the stimulus set created by Brady and colleagues (2008; http://cvcl.mit.edu/MM/download.html). Pictures were 256 x 256 pixels and showed a single object on a white background. There were two types of target-lure pairs (see Figure 1b for examples). For exemplar pairs, the target and lure were different exemplars of the same basic-level category. For state pairs, the target and lure were the exact same object in a different state or pose. Thus, the state pairs are analogous to the target-lure pairs used in Lupyan (2008) and Experiment 1, where targets and lures show the same object altered in a small but noticeable way. There were 40 exemplar pairs and 40 state pairs, and each pair consisted of an object from a unique basic-level category. Stimuli were randomly divided into two sets with an equal number of exemplar and state pairs in each set. The object set assigned to each study task was counterbalanced across participants.

Procedure

The procedure was the same as Experiment 1 with the following exceptions. For the naming task, participants typed the first letter of the name of the object. To equate response selection demands between the naming and preference tasks, instead of making "like" vs. "don't like" responses for the preference judgements, participants were asked to rate how much they liked the object in the picture compared with other objects from the same category on a scale of 1-5 (1 = dislike, 5 = like). Because of the additional response selection demands for these tasks compared with Experiment 1, there was no response deadline during the study phase. In addition, no unrelated lures were included in the test phase, and the instructions were modified to reflect this (i.e. participants were no longer told that any of the pictures would be brand new).

Recognition memory is better when objects are from many different categories (Mandler, Pearlstone & Koopmans, 1969; Koutstaal et al., 2003; Koutstaal & Schacter, 1997). Therefore several changes were made to the study phase of Experiment 2 to increase the overall task difficulty: twice as many pictures were presented at study, each picture was presented once for only 150ms (vs. twice for 300ms in Experiment 1). In addition, unlike Lupyan (2008) and Experiment 1 where the test phase immediately followed the study phase, the test phase in Experiment 2 was completed approximately 30 minutes after the study phase. During this time participants took part in another unrelated experiment.

Participants were familiarized with the pace of the experiment with five practice trials that showed pictures of chairs and lamps. Participants were asked to press ‘1’ if the picture was a chair, and ‘2’ if the picture was a lamp.

To code correct responses in the naming task, at the end of the experiment participants were given a sheet that showed images of all the target-lure pairs used in the experiment and were asked to write the name for each pair of objects. Correct responses were coded based on how participants named the objects at the end of the experiment.

Results

Study Phase

Average accuracy on the naming task was 88.5% (*SD* = 6.57). Because each object was presented only once during the study phase it was not possible to compute reliability scores. There was no difference in correct RTs between the two tasks in the study phase (see Table 1).

Test Phase

Performance on the recognition memory test is plotted in Figure 4. 2 x 2 repeated-measures ANOVAs with factors Lure Type (exemplar vs. state) and Study

Task (naming vs. preference judgement) were performed on d', correct RTs, hit rate and false alarm rate.

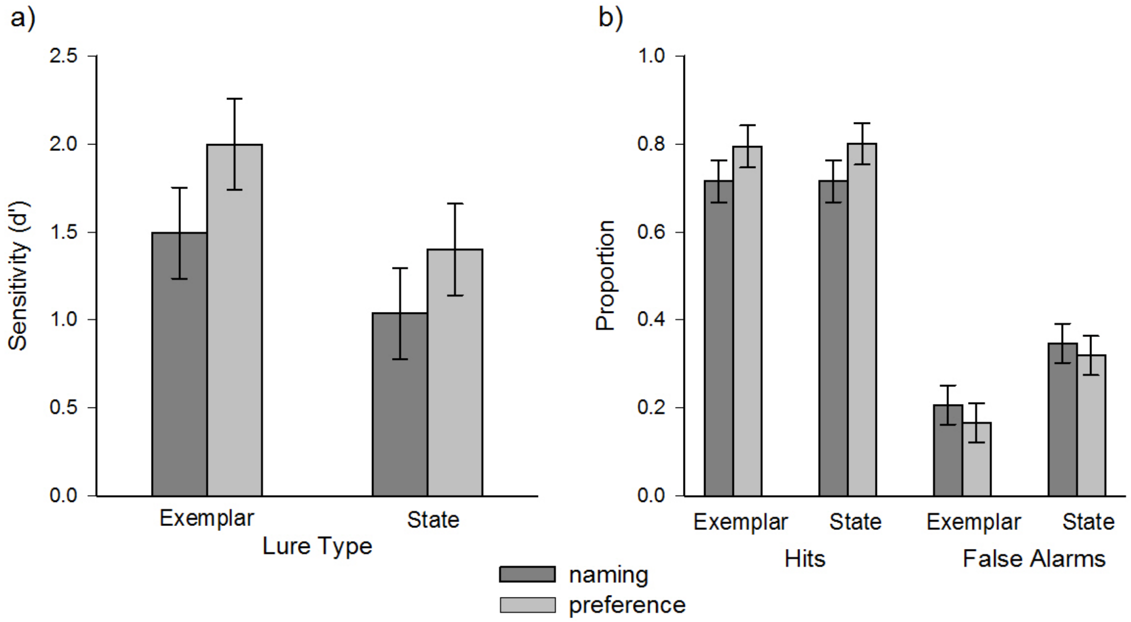

Figure 4. a) Overall performance (d') and b) hit and false alarm rates on the recognition memory test for exemplar and state target-lure pairs presented during the naming and preference study tasks in Experiment 2. Error bars show 95% confidence intervals of the paired-sample *t*-tests.

Overall performance was better for exemplar vs. state lure pairs ($F_{1, 15}$ = 12.797, *MSE* = .348, $p < .01$) and for objects that were named vs. objects for which preference was rated ($F_{1, 15}$ = 13.474, *MSE* = .224, $p < .01$). Critically, the interaction between Lure Type and Study Task was not significant. Paired-sample *t*-tests revealed that the naming effect (better performance for objects presented during preference vs. naming study blocks) was significant for both exemplar lure pairs (t_{15} = 2.960, $p = .01$) and state lure pairs (t_{15} = 2.159, $p < .05$).

Hit rates were lower for named objects vs. preference objects ($F_{1,15}$ = 14.006, *MSE* = .008, $p < .01$). Paired-sample *t*-tests revealed a significant difference in hits between objects presented in the naming vs. preference study tasks for objects paired with both exemplar lures (t_{15} = 2.815, p = .013) and state lures (t_{15} = 2.522, $p < .05$).

There were more false alarms for state lures compared with exemplar lures ($F_{1,15}$ = 21.447, *MSE* = .016, $p < .001$). Paired-sample *t*-tests revealed no significant differences in false alarm rates between study tasks for either lure type.

Table 3.
Correct Response Times (and Standard Deviations) on the Recognition Memory Test for State and Exemplar Target-Lure Pairs Presented During Naming and Preference Study Tasks in Experiment 2.

Study Task	Lure Type	Response Times (ms)
Naming	State	1256.73 (408.18)
	Exemplar	1189.62 (407.91)
Preference	State	1139.73 (367.59)
	Exemplar	1166.75 (375.82

Correct RTs for all combinations of study task and lure type are reported in Table 3. The interaction between Lure Type and Study Task approached significance in correct RTs ($F_{1,15}$ = 4.384, *MSE* = 8084.932, p = .054). Although the main effect of Study Task was significant ($F_{1,15}$ = 17.668, *MSE* = 4428.952, p = .001), paired-sample *t*-tests revealed that correct RTs to objects from the naming block were slower than correct RTs to objects from the preference block for state lure pairs only (t_{15} = 3.967, p = .001).

All Unique Categories vs. Two Categories

Only data from state lure pairs were used to compare with Experiment 1 because these lures are more similar to the kind of lures used in that experiment. A 2 x 2 mixed-factor ANOVA on overall performance (d') with within-subjects factor of Study Task (naming vs. preference) and Experiment (Experiment 1 vs. Experiment 2) as a between-subjects factor revealed a main effect of Study Task ($F_{1,30}$ = 28.216, MSE = .191, p < .001). The interaction between Study Task and Experiment approached significance ($F_{1,30}$ = 3.913, MSE = .191, p = .057), but independent sample t-tests revealed no significant differences in memory performance between experiments for objects that were named vs. objects for which preference was rated. As can be appreciated in Figure 5, this suggests that the trend toward a smaller naming effect in Experiment 2 is due to both an increase in naming performance and a decrease in preference performance compared with Experiment 1.

A similar ANOVA on hit rates revealed a main effect of Study Task ($F_{1,30}$ = 66.807, MSE = .007, p < .001) and, critically, a significant Study Task x Experiment interaction ($F_{1,30}$ = 18.212, MSE = .007, p < .001). Independent-sample t-tests revealed that hit rates were significantly lower for named objects in Experiment 1 vs. Experiment 2 (t_{30} = 3.609, p = .001). There were no significant effects in false alarm rates.

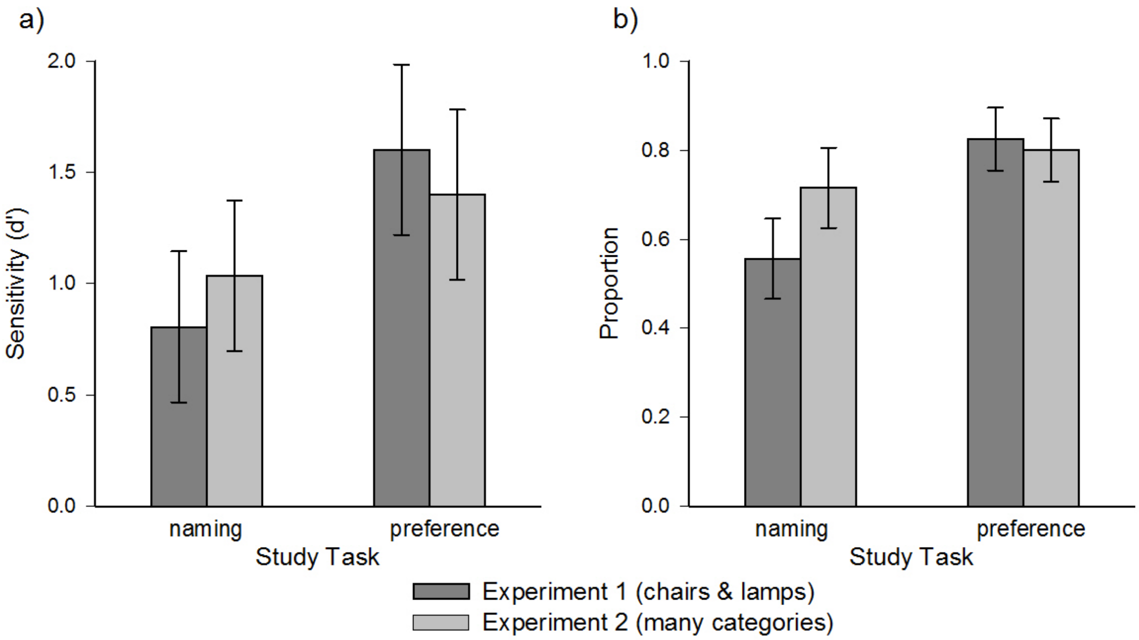

Figure 5. a) Overall performance (d') and b) hit rates on the recognition memory test for objects that were named and objects for which preference was rated in Experiment 1 (where stimuli were chairs and lamps) and Experiment 2 (where every object came from a unique basic-level category). Error bars show 95% confidence intervals for the independent-sample *t*-tests.

Discussion

Experiment 2 shows that naming leads to worse recognition memory performance driven by a decrease in hit rates relative to preference judgements when participants need to explicitly generate basic-level names. In addition, Experiment 2 shows that this pattern of results is obtained regardless of whether lures are matched to targets (state lures) or are just categorically related (exemplar lures), indicating that the naming effect is not influenced by the type of lures being used. Although overall memory was worse for state lure pairs, this was driven by more false alarms compared with exemplar lures (which is not surprising, since state lures look more similar to their matched targets). Critically, lure type did not influence the relationship between naming and preference.

Hit rates were higher in Experiment 2, where every object had a unique name, compared with Experiment 1 where objects were either chairs or lamps. There are several possible explanations for this result. According to the fan effect, retrieval is impaired when multiple items are associated with the same concept (Anderson, 1974; Anderson & Reder, 1999; Reder & Anderson, 1980; Reder, Donavos & Erickson, 2002). When all study objects are from only two categories, objects from the same category that already share many visual features are linked to the same conceptual feature – the basic-level name – during the naming task, possibly resulting in interference at test and reduced recognition memory performance. In Experiment 2, where every object had a unique name, memory may have been better because of less overlap in conceptual fans for named objects. However, although the difference between naming and preference in hit rates was smaller when objects that were named were no longer associated with the same names, this difference was not completely eliminated, indicating that the fan effect on its own cannot account for the naming effect. Performance differences between Experiment 1 and Experiment 2 may also be due to the fact that actually generating the basic-level name was a more cognitively demanding task that improved subsequent memory (Craik & Lockhart, 1972; see Chapter 4). Alternatively, this difference in hit rates may just reflect that recognition memory in general is facilitated when objects are from many different basic-level categories (Mandler et al., 1969; Koutstaal et al., 2003; Koutstaal & Schacter, 1997), possibly due to a memory benefit when encoding episodes are more distinct (Craik & Tulving, 1975; Klein & Saltz, 1976; Moscovitch & Craik, 1976).

Importantly, although there was an increase in hit rates in Experiment 2 compared with Experiment 1, the difference in memory performance between objects that were named and objects for which preference was rated was still significant. This is consistent with Craik & Tulving (1975) who showed that although increasing the uniqueness of encoding operations for a particular study task improved memory performance for objects encoded during that particular task, the ordering of different study tasks in terms of performance remained unchanged.

In summary, Chapter 2 provided a replication of the main result presented by Lupyan (2008), and showed that this effect is not dependent on the number of basic-level categories or the type of lures being used. The following experiments test whether overt naming itself is actually necessary for representational shift (Chapter 3), and whether there is an alternative, more parsimonious explanation for the pattern of results used to support the representational shift hypothesis (Chapter 4).

CHAPTER III

OVERT NAMING VS. BASIC-LEVEL CATEGORIZATION

The naming task in Lupyan (2008) is also a basic-level categorization task: participants are pressing keys to classify stimuli as chairs or lamps, and undeniably every act of naming is also an act of categorization (Lupyan, 2005). Thus, while it is possible that naming drives the decrease in recognition memory, it is also possible that the effect is due to basic-level categorization. The false memory literature finds that categorical processing disrupts item-specific recognition performance because of higher false alarms for lure items that are from the same category as studied items (Koutstaal & Schacter, 1997; Seamon et al., 2000). Moreover, processing study items at the category level has been shown to lead to an increase in false alarms (Koutstaal et al., 2003; Sloutsky & Fisher, 2004). The explanation for this false memory effect is that considering objects at the category level leads to coarse encoding of only the features that are relevant to the category. As a result, objects from the same category that share the same category features are falsely recognized.

Lupyan (2008) posits that because the effect he observes is not in false alarms, the effect cannot be attributed to coarse encoding due to categorization and must instead be caused by representational shift following overt naming. But in the previous studies of false memory it is the categorical relationship between target and lures or the size of a category at study that are manipulated, not encoding tasks: false memory refers to the tendency to falsely recognize lures that are categorically related to the target *relative to categorically unrelated lures* (Koutstaal & Cavendish, 2006; Koutstaal et al., 2003; Koutstaal & Schacter, 1997; Seamon et al., 2000; Sloutsky & Fisher, 2004). These within-category false alarms increase as the number of studied items from that category increase (Koutstaal et al., 2003; Koutstaal & Schacter, 1997) and are modulated by the typicality of lures as exemplars of the studied category (Seamon et al., 2000).

In fact, only two studies of false memory for pictures report differences in false alarm rates *between study tasks*: Marks (1991) found higher false alarm rates following a conceptual judgement that required participants to access semantic information (e.g., "is it a mode of transportation?") vs. a judgement about surface perceptual features (e.g., "is it red?"); Sloutsky and Fisher (2004) found higher false alarm rates for pictures of animals following a category induction task where an unobservable feature correlated with category membership compared to a task where participants judged the age of the animal or passive viewing. Critically, any conclusions about increases or decreases in false alarms between tasks necessarily depend on the two tasks being compared. Preference judgements are certainly not the same as surface-level perceptual judgements, naming does not necessarily probe semantic information, and neither of these tasks map onto the tasks used by Sloutsky & Fisher (2004). In short, there is no basis for predictions regarding differences in false alarms between naming and preference judgements, and not finding differences in false alarm rates between naming and preference judgements is insufficient evidence that the effect attributed to overt naming in Lupyan (2008) is not in fact due to basic-level categorization.

The following two experiments will test whether the reduction in recognition memory performance following naming observed in Lupyan (2008) can be attributed to basic-level categorization as opposed to overt naming by using tasks where participants must consider category membership (Experiment 3) or attend to category-level information (Experiment 4) without requiring an overt naming response.

Experiment 3

In category induction tasks, participants are told about an unobserved feature of a member of a category. They are then shown pictures of objects from the same category or a different category and asked if that object shares the unobserved property. This task is believed to be performed by first identifying the object as a member of a category, and then making inductive inferences on the basis of this categorization (Gelman & Markman, 1986; Rehder & Hastie, 2001; Ross & Murphy, 1996; Yamauchi & Markman, 2000). Categorizing stimuli is believed to lead to the formation of a gist or category-level memory trace because item-specific perceptual information is irrelevant to the categorization task (Koutstaal et al., 2003), and this ultimately leads to an increase in false alarms: any test object that shares the category-level information will be falsely recognized. Because category induction requires consideration of category membership, category induction should also lead to increases in false alarm rates. This hypothesis was supported by Sloutsky and Fisher (2004) who found that performing a category induction task led to an increase in false alarms in a subsequent recognition memory task relative to a task where the judgement being made (e.g., “is the animal young?”) encouraged focusing on item-specific perceptual features, or compared with passive viewing.

In Experiment 3 participants performed three tasks at study: naming, preference and category induction. Birds and dogs were used as stimuli (as opposed to chair and lamps) because there is evidence that category induction is stronger for natural categories compared to artifacts (e.g., Gelman & Markman, 1986, 1987). If the category induction task leads to an increase in false alarms, while the naming task leads to a decrease in hits, relative to preference judgements, this would suggest that the effect of naming is distinct from memory effects related to basic-level categorization.

Method

Participants

Eighteen Vanderbilt University undergraduates (5male; mean age 20 years), participated in exchange for course credit. Data from one participant were discarded for below chance performance on the memory test. Data from two more participants were discarded for accuracy on the category induction task that was more than two standard deviations below the mean.

Stimuli

Stimuli were 42 images of dogs and birds (21 images from each category), 250 x 250 pixels in size. Images were obtained from the internet and were edited in

PhotoShop such that each image showed a single dog or bird on a white background. There were 21 pairs of dog and bird pictures, where each target dog or bird was matched with a critical lure. Lures differed from targets in small but noticeable ways. Target-lure pairs consisted of dogs of the same breed or birds of the same species (see Figure 1c for examples). A given breed or species was only used to create one target-lure pair. One additional bird and dog image were created to use as an example during the instructions that preceded the category induction block. This bird and dog were not the same breed/species as any of the target-lure pairs, and this image was not presented at test. Images were sorted into three sets of 14 pairs (seven dog pairs, seven bird pairs). For each participant an object set was assigned to each of the three study tasks. The object set assigned to each task, task order, which object in the pair was the target and which was the lure were all counterbalanced across participants. The target category in the category induction task was randomized, such that the target was “dog” for half of the subjects and “bird” for the other half.

Procedure

There were two parts to the experiment. The study phase was identical to Experiment 1, with the exception that a third study task, category induction, was included in addition to the naming and preference tasks. The category induction task was based on the procedure used by Sloutsky & Fisher (2004). At the beginning of the block, participants were shown a picture of a dog (or bird) and told that the animal in the picture “has beta cells in its body”. For the subsequent pictures in the block, participants were instructed to determine which other animals also have beta cells. They were instructed to press ‘1’ if they thought the animal in the picture had beta cells, and ‘2’ if they thought the animal did not have beta cells. Correct responses were based on category membership, such that if a dog was presented during the instructions, participants should have responded that all dogs, but no birds, have beta cells. Corrective feedback (the word “correct” or “incorrect” presented for 500ms) was provided after each trial to ensure that participants learned to make the induction on this basis.

Task blocks were repeated twice during the study phase. Prior to the experimental tasks participants were familiarized with the pace of the experiment with five practice trials that were identical to the practice trials in Experiment 1.

After the study phase participants performed a surprise recognition memory task that was identical to Experiment 2. All 84 objects were presented in the recognition memory phase in random order.

Results

Study Phase

Average accuracy on the naming task (93.95%; *SD* = 4.97) and the category induction task (94.68%; *SD* = 4.09) did not differ from each other. A repeated-measures ANOVA on reliability scores revealed no significant differences in reliability between the study tasks (naming: 90.16%; preference: 83.50%; category induction: 90.22%). However, a similar ANOVA on correct study RTs revealed significant differences in RTs between tasks ($F_{2,28} = 15.628$, $MSE = 2775.825$, $p < .001$; see Table 1). Paired-sample *t*-tests indicated that RTs to rate preference were significantly slower than RTs on the naming task ($t_{14} = 4.263$, $p = .001$) and the category induction task ($t_{14} = 4.473$, $p = .001$). RTs did not differ between the naming and category induction tasks.

Test Phase

Performance on the recognition memory test is shown in Figure 6. A repeated-measures ANOVA on d' (naming vs. preference vs. category induction) revealed a significant effect of study task on recognition memory performance ($F_{2,28} = 5.297$, MSE = .343, p = .011). Paired-sample *t*-tests indicated that performance was significantly better for objects studied in a preference block compared with objects that were studied in a naming block ($t_{14} = 3.260$, $p < .01$) or category induction block ($t_{14} = 2.269$, $p < .05$). Recognition memory performance did not differ between objects studied in naming vs. category induction blocks.

Similar ANOVAs on hit and false alarm rates revealed a significant effect of study task on hit rate ($F_{2,28} = 31.092$, $MSE = .014$, $p < .001$). Paired-sample *t*-tests indicated that hit rate was significantly higher for objects studied in a preference block compared with objects that were studied in a naming block ($t_{14} = 8.189$, $p < .001$) or category induction block ($t_{14} = 5.107$, $p < .001$). Hit rates did not differ between objects that were studied in naming vs. category induction blocks.

Study task also had a marginally significant effect on false alarm rates ($F_{2,28} = 3.253$, $MSE = .016$, $p = .054$). Surprisingly, paired-sample *t*-tests revealed that there were significantly more false alarms for objects studied in the preference block compared with objects that were named ($t_{14} = 2.167$, $p < .05$).

Study task did not have a significant effect on correct RTs (see Table 2).

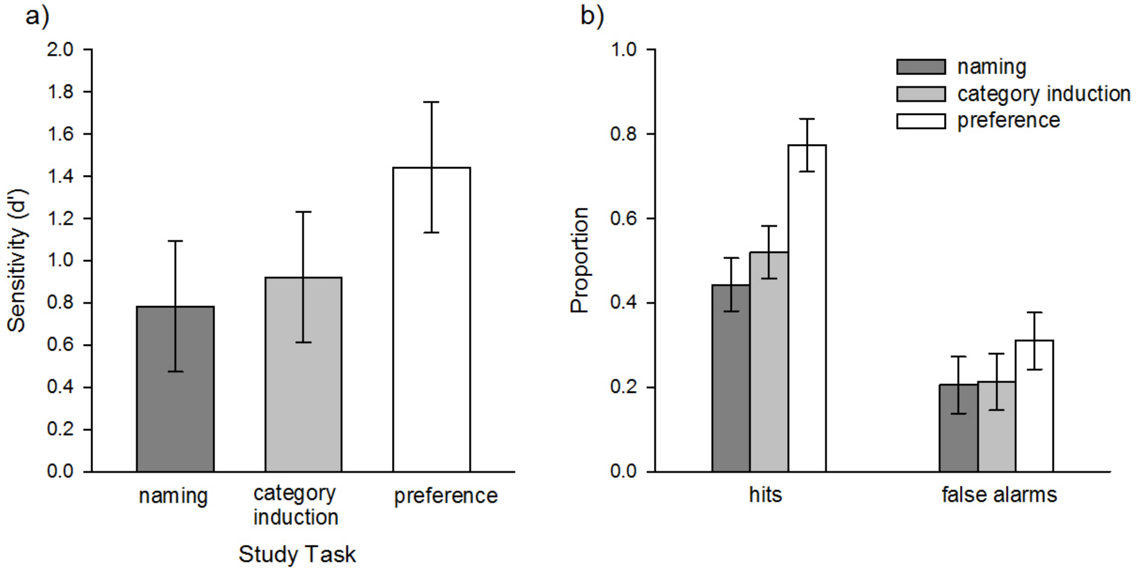

Figure 6. a) Overall memory performance (d') and b) hit and false alarm rates in the recognition memory test for objects presented during each study block (naming, preference and category induction) in Experiment 3. Error bars show 95% confidence intervals of the within-subjects effect.

Discussion

Experiment 3 replicated the difference in memory for objects that are named vs. objects for which preference is rated reported in Lupyan (2008) and Experiment 1 using a different set of stimuli, birds and dogs. As in Lupyan (2008) and Experiment 1, this difference in recognition memory performance between naming and preference study tasks was driven by a difference in hit rates. In contrast to these previous experiments, in Experiment 3 a significant difference in false alarm rates was also observed between objects that were named and objects for which preference was rated. However, this difference was in the opposite direction than what might be predicted based on the false memory literature: there were more false alarms for lures that matched preference objects vs. named objects. Thus, worse recognition memory performance for named objects was still exclusively driven by a difference in hit rates. More interestingly, Experiment 3 showed the same pattern of performance – worse recognition memory driven by a difference in hit rates – when objects from the preference study task were compared with objects from the category induction study task. This suggests that the pattern of results that Lupyan (2008) argues must be due to overt naming also arises due to consideration of category membership on its own, without an overt naming response.

However, an alternative possibility is that because participants were required to make an induction based on category membership, participants may have simply re-coded the response keys for "yes" and "no" to "bird" and "dog" (or "dog" and "bird", depending on the target category). This could be especially likely in the second block of category induction because participants had presumably already learned the induction rule in the first block. Indeed, performance on this task was very high, indicating that participants learned the induction rule very quickly. Thus, participants may have been making naming responses despite the response labels given in the instructions.

Experiment 4

Experiment 4 provides a second test of the hypothesis that impaired recognition memory following naming can also arise due to basic-level categorization on its own. In this experiment participants performed a sequential matching task with chairs and lamps. In one block participants were instructed to judge whether two sequentially presented items were from the same category (category-matching); attending to category-level information may be similar to naming since this is the level at which the objects were being named in the previous experiments. In the second block participants were instructed to judge whether two sequentially presented items were the exact same exemplar (exemplar-matching), requiring attention to the details of the exemplar itself; this may be similar to preference judgements because of the requirement to consider the specifics of the presented object – this is *a* chair vs. I like *this* chair. Importantly, two items were presented on every trial and whether the second item was a chair or lamp did not correlate with whether the correct response was "same" or "different", making it impossible for participants to re-map the response keys as "chair" and "lamp".

Methods

Participants

Twenty-four Vanderbilt undergraduates (10 male; mean age 21.1 years) received course credit in exchange for participation in this experiment. Data from four participants were discarded for below chance performance on the memory test.

Stimuli

Stimuli were 144 colour pictures of chairs and lamps (72 from each category; 36 target-lure pairs) that were obtained in the same manner as Experiment 1. Pictures were sorted into four sets. Two sets contained 20 target-lure pairs and were designated as “target sets”. Two other sets contained 16 target-lure pairs and were designated as “non-critical sets”. One target set and one non-critical set were assigned to the category-matching block and exemplar-matching block for each participant. Set assignment was counterbalanced.

During the study phase both target and lure objects from the non-critical set were presented once. Only the target items from the target sets were presented (assignment as a target or lure was counterbalanced), and these objects were displayed twice, with both presentations in the same trial.

Procedure

On each trial participants saw a fixation cross (500ms) followed by the first image (300ms), then a random pattern mask was presented (500ms), followed by the second image (300ms). A question mark was then displayed, cueing participants to respond. Participants were instructed that responses would only be accepted when the question mark was on the screen. Participants had 700ms to respond, and they heard a tone if they responded too slowly at which point the trial timed-out. Time-out trials were not included in study phase analyses. In the category-matching block participants were instructed to press ‘1’ if the two objects were from the same basic-level category, and ‘2’ if the two objects were from different basic-level categories; in the exemplar-matching block participants were instructed to press ‘1’ if the two objects were the exact same object, and ‘2’ if the two objects differed in any way. Participants completed one block of exemplar-matching and one block of category-matching, with task order counterbalanced across participants. A five-trial practice block where participants had to decide if two sequentially presented tables were the same or different shape preceded the experimental blocks to familiarize participants with the pace of the experiment.

There were 52 trials in each task block. The different trial types and their frequency for the category-matching and exemplar-matching blocks are illustrated in Figure 7. In both blocks there were 20 critical target trials (created with targets from the target object set). On these trials the exact same object was presented twice and the correct response was “same” (the exact same object is both the same exemplar and the same category). The remaining 32 trials in each block were created from objects in the non-critical object set and were designed to create either a category-matching or exemplar-matching context. For category-matching blocks the remaining 32 trials consisted of 16 non-critical “same” trials, where the two objects were different exemplars from the same category, and 16 “different” trials, where the two objects were from different categories. For the exemplar-matching block the non-target trials consisted of target-lure pairs from the non-critical object set, and required a “different” response. In this way very subtle differences needed to be detected on different trials in the exemplar-matching block, and participants could not rely on global similarity. Prior to the experiment participants were shown examples of same and different trials for each type of matching judgement so they were aware of the kind of subtle differences that would need to be detected during the exemplar-matching block.

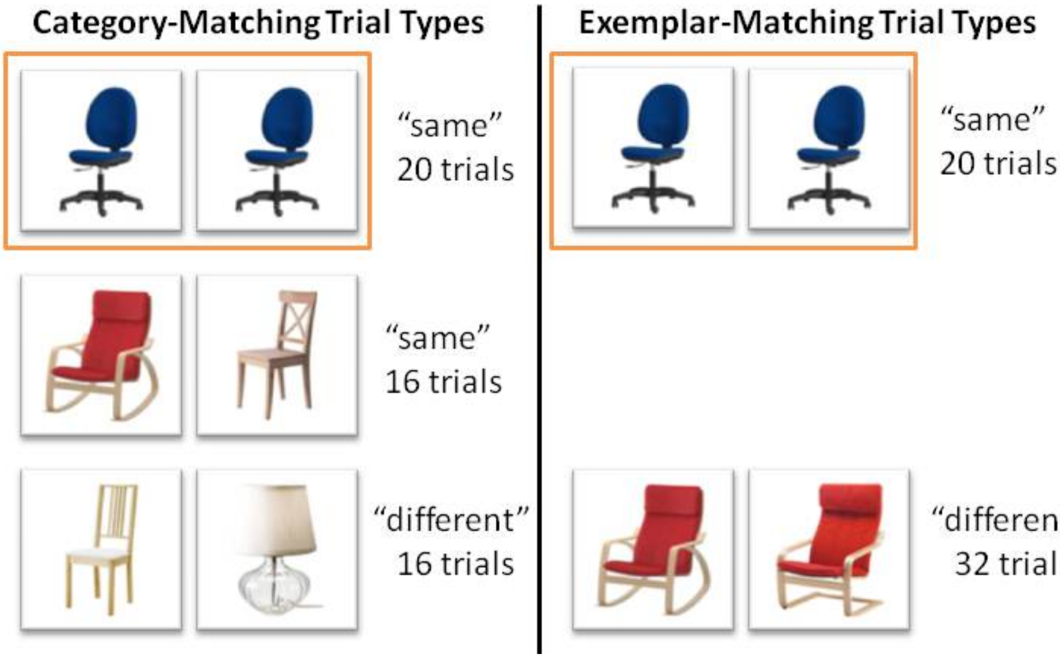

Figure 7. Trial types and their frequencies in category-matching (left) and exemplar-matching (right) study blocks in Experiment 4. Both blocks contained an equal number of critical target trials (bordered in orange).

Although the overall proportion of same and different trials differed between the category-matching and exemplar-matching blocks, the critical target trials were identical: both blocks contained the same number of trials where the same object was presented twice within the same trial and where the correct response was "same", and these are the critical trials that were later tested for recognition. The difference between blocks was necessary to create the different contexts in which the critical target trials were presented.

Matching blocks were followed by a surprise recognition memory test that was identical to the previous experiments. Recognition memory was only tested for objects presented on critical target trials, resulting in a total of 80 test trials (the 20 target set items from each study block and their matched lures).

Results

Study Data

Accuracy was significantly higher for category-matching (93%; *SD* = 3.61) vs. exemplar-matching (87%; *SD* = 4.61; $t_{19} = 4.46$, $p < .001$). Correct RTs did not differ between tasks (see Table 1).

Test Data

Performance on the recognition memory test is plotted in Figure 8. Paired-samples *t*-tests revealed a trend toward better overall recognition memory performance (d') for objects presented in the exemplar-matching block vs. objects presented in the category-matching block ($t_{19} = 2.03$, $p = .056$). In addition, there was a significant difference in correct RTs in the test block (see Table 2), with slower performance for objects presented in the category-matching block ($t_{19} = 2.869$, $p = .01$). This suggests that some of the effect occurred in RTs, as opposed to just d' as in Experiments 1 and 2.

There was a significant difference in both hit rates and false alarm rates, with more hits for objects in the exemplar-matching block vs. category-matching block

($t_{19} = 4.66$, $p < .001$). Unexpectedly, the difference in false alarm rates was also driven by more false alarms for items in the exemplar-matching vs. category matching block ($t_{19} = 2.90$, $p < .01$).

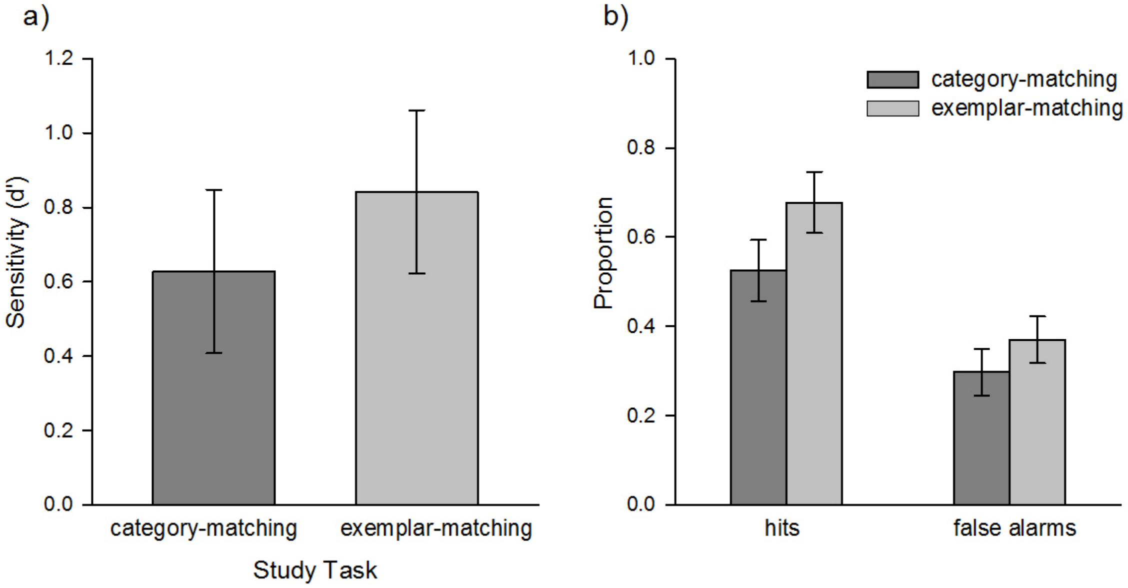

Figure 8. a) Overall performance (d') and b) hit and false alarm rates on the recognition memory test for objects presented in the category-matching and exemplar-matching study blocks in Experiment 4. Error bars show 95% confidence intervals for the paired-sample *t*-tests.

Sequential Matching vs. Naming/Preference

It was hypothesized that requiring attention at the category-level vs. exemplar-level would produce the same effect in recognition memory performance as making naming vs. preference judgements. Performance in Experiment 4 was compared with performance in Experiment 1 (see Figure 9) to determine whether and how making sequential matching judgements led to quantitatively different results than making naming and preference responses. If overall performance is worse for objects that were named compared with objects that were matched at the category level, this would suggest that although category-matching and exemplar-matching led to the same qualitative results as naming and preference, naming influences recognition memory more than basic-level categorization.

A 2 x 2 mixed-factor ANOVA with within-subjects factor Task (naming/category-matching vs. preference/exemplar-matching) and Experiment (Experiment 1 vs. Experiment 4) as a between-subjects factor revealed a significant main effect of Task ($F_{1, 34}$ = 34.937, MSE = .130, p < .001), a main effect of Experiment ($F_{1, 34}$ = 19.677, MSE = .199, p < .001) and a significant interaction between Task and Experiment ($F_{1, 34}$ = 11.618, MSE = .130, p < .01), such that the difference in recognition memory performance between study tasks was smaller in Experiment 4 than Experiment 1.

Independent samples *t*-tests comparing performance on naming vs. category-matching and preference vs. exemplar matching revealed that performance was significantly higher for objects in the preference condition in Experiment 1 compared with objects that were matched at the exemplar level in Experiment 4 ($t_{1, 34}$ = 4.929, $p < .001$). Critically, there was no difference in performance between objects that were named in Experiment 1 and objects that were matched at the category level in Experiment 4. These results suggest that the smaller difference in overall memory performance between category-matching and exemplar-matching in Experiment 4 compared with the difference between naming and preference in Experiment 1 is due to lower performance for objects in the exemplar-matching task vs. the preference judgement task, rather than a difference in performance for objects in the naming vs. category-matching tasks.

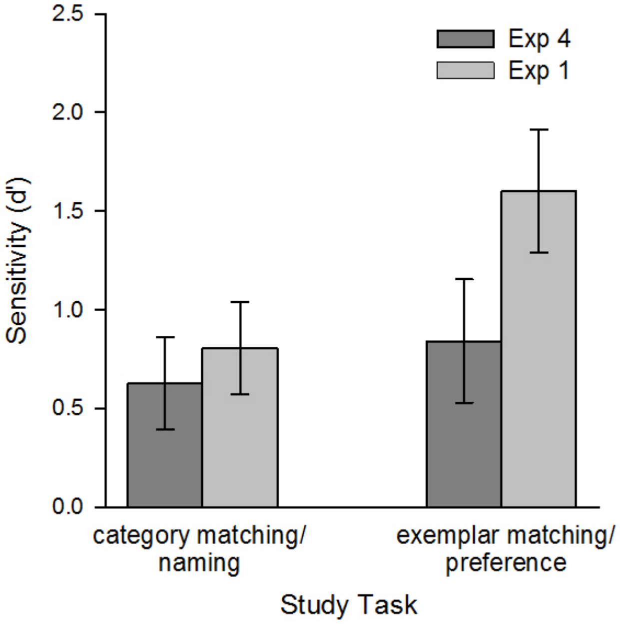

Figure 9. Comparison of overall memory performance (d') between Experiments 4 (category-matching and exemplar-matching study tasks)and Experiment 1 (naming and preference study tasks). Error bars show 95% confidence intervals of the independent-sample *t*-tests.

Discussion

In Experiment 4 recognition memory performance was worse for objects presented in the context of category-matching vs. objects presented in the context of exemplar-matching, and this difference was driven by a difference in hit rate. Like Experiment 3, although a difference in false alarm rates was observed, this was in the opposite direction than what might be predicted based on the false memory literature: false alarms were higher for objects presented during exemplar-matching. Thus, lower performance for objects matched at the category level was exclusively driven by a lower hit rate.

Together, the results of Experiments 3 and 4 demonstrate that, in contrast to the claim made by Lupyan (2008), overt naming is not required to obtain the pattern of results used to support the representational shift hypothesis: differences in recognition memory performance driven by differences in hit rates can be obtained in tasks that require basic-level processing but where the response labels are not category names (Experiment 3 and Experiment 4), and where category does not correlate with the correct response keys (Experiment 4). Additionally, in Experiment 3 performance for objects presented during the category induction task was not significantly different from performance for objects that were named, and memory performance did not differ between objects that were named in Experiment 1 and objects that were matched at the category-level in Experiment 4. Thus, overt naming does not disrupt recognition memory more than other basic-level processing tasks.

The results of Experiment 3 and Experiment 4 also suggest that tasks that encourage category-level processing do not necessarily lead to an increase in false alarms relative to other tasks. Neither the category induction task nor the category-matching task led to increases in false alarms relative to the other study tasks in those experiments. Therefore increases in false memory for categorically related lures (e.g., Koutstaal & Schacter, 1997; Koutstaal et al., 2003) is not necessarily an outcome of explicitly focusing on the category-level information during encoding. Furthermore, whether a given task leads to an increase in false alarms relative to another task depends on the specific tasks being compared: category induction did not lead to more false alarms than the other study tasks in Experiment 3, despite the fact that this task has been linked to increases in false alarms in previous work (Sloutsky & Fisher, 2004).

In sum, Chapter 3 establishes that if differences in recognition memory performance driven by differences in hit rates are indicative of representational shift, then representational shift does not depend on overt naming but occurs whenever a memory representation is processed at the basic-level. Chapter 4 addresses a possible alternative to the representational shift hypothesis that may account for this pattern of results.

CHAPTER IV

DEPTH OF PROCESSING ACCOUNT OF THE NAMING EFFECT

It is difficult to draw conclusions about processing based on differences between study tasks because such a comparison always depends on the specific tasks being compared: naming leads to worse memory performance compared with preference judgements, but may lead to better performance compared with another encoding task; this difference in performance between objects presented during naming and preference study tasks might arise because naming *impairs* recognition memory, but could also arise if preference judgements *enhance* recognition memory. Traditionally the effects of different encoding tasks on subsequent memory have been investigated in experiments where participants study word lists. The general finding is that memory is worse for words where participants are asked to make a surface level judgement (e.g., is the word written in upper-case or lower-case; shallow encoding) compared to words where participants make judgements that tap into conceptual processing, such as judgements about the meaning of the word (deep encoding; see Craik & Lockhart, 1972 and Craik, 1979, for reviews), when memory is tapped by a recognition memory task (Lockhart & Craik, 1990).

Chapter 3 demonstrated that naming leads to the same degree of memory impairment as basic-level processing. Although conceptual processing is a deep encoding strategy (Craik, 1979; Craik & Lockhart, 1972; Craik & Tulving, 1975), semantic processing may not facilitate recognition if it is not useful in differentiating targets and lures (Moscovitch & Craik, 1976; Morris et al., 1977). Because targets and lures are from the same category, such conceptual information may not be beneficial in the paradigm being used here. Furthermore, in classic depth of processing experiments the conceptual tasks go beyond a basic-level name, and instead ask specific questions about meaning, often aimed at levels of abstraction other than the basic-level (e.g., "Is it a kind of furniture?" requires knowledge about the superordinate category). Since for most objects the basic-level is the entry level of processing (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), processing objects at a level of abstraction other than the basic-level requires more time and effort (Jolicoeur et al., 1984). Indeed, basic-level categorization may occur relatively automatically (Grill-Spector & Kanwisher, 2005; Mack, Gauthier, Sadr & Palmeri, 2008).

Therefore, naming may be a relatively shallow task because it may be relatively automatic (e.g., Coltheart, 1999; Schiano & Watkins, 1981; Zelinsky & Murphy, 2000). If pictures of objects automatically activate their names (e.g., Meyer et al., 2007; Meyer & Damian, 2007), then better recognition memory performance following preference judgements vs. naming could be explained by a depth of processing account: performance might be better for objects where preference judgements are made because the preference task is more effortful than naming, and thus leads to a stronger and more persistent memory trace (Craik, 1979; Craik & Lockhart, 1972; Craik & Tulving, 1975).

Lupyan (2008) argues that the naming effect cannot be explained by a depth of processing account because the same effect was obtained in an experiment where preference and naming trials were randomized, and participants were not aware of which type of response they would need to make until after the object was removed from the display. Crucially, however, this only equates encoding processes, not differences in post-encoding processing that may be driven by differences in task demands. Indeed, depth of processing is often associated with longer response times (Craik & Tulving, 1975), and response times were longer for preference vs. naming responses during the study phase in both Experiments 1 and 3 (see Table 1), and the majority of experiments reported by Lupyan (2008) despite the fact that responses were made after the image was removed from the screen. Additionally, preference responses were less reliable than naming responses during the study phase in Experiment 1. One possible explanation for this difference in reliability is that participants were still considering the object during the second presentation and reversing their original decision about whether they liked it or not. According to depth of processing, such additional contemplation of the object would result in better encoding. Moreover, comparisons between Experiment 1 and Experiment 4 revealed that preference judgements led to significantly better performance than exemplar matching. This could also be consistent with greater depth of processing for preference judgements: the exemplar-matching task only required attending to physical properties of the objects, whereas a preference judgement requires considering not only the physical properties themselves, but also how one feels about them (i.e., self-reference effect; Symons & Johnson, 1997).

The following two experiments will test a depth of processing account of the naming effect that is contingent on the automaticity of naming. Experiment 5 includes a classic depth of processing manipulation (surface-level judgement) in addition to naming and preference study tasks to explore whether recognition memory performance adheres to predictions based on depth of processing. Experiment 6 investigates whether differences in performance between naming and preference are driven by impairment following naming or improvement following preference judgements.

Experiment 5

According to depth of processing, retention depends on both the qualitative nature of encoding, where any degree of conceptual analysis is more beneficial than structural analysis, and the amount of encoding, where multiple encoding tasks lead to more processing and thus better retention (Craik, 1979; Craik & Lockhart, 1972; Craik & Tulving, 1975). Therefore, depth of processing predicts a gradient of performance based on the type and number of encoding tasks. In Experiment 5 participants named objects, made preference judgements about objects, or judged whether objects were in any colour or were neutral (black/white/beige). If naming is automatic (e.g., Coltheart, 1999; Schiano & Watkins, 1981; Zelinsky & Murphy, 2000) then the colour judgement task will in essence consist of both naming (automatic) and the additional colour response. Because the colour judgement task will therefore be two tasks, performance should be better following colour judgements vs. following naming, even though colour judgement is a surface-level, shallow processing task (Banks & Barber, 1977; Treisman & Gormican, 1988). At the same time, depth of processing predicts that performance following colour judgements will be lower than performance following preference judgements because the colour judgement is a shallow processing task while preference involves deeper encoding – participants must consider why they like the object, producing a self-reference effect (Symons & Johnson, 1997). Thus, a depth of processing account predicts a gradient of recognition memory performance with the lowest performance for objects that are named (one shallow processing task), followed by objects where a colour judgement is made (two shallow processing tasks), followed by objects for which preference is rated (one shallow processing task and one deep processing task).

Method

Participants

Eighteen members of the Vanderbilt community (9 male; mean age 22.5 years) participated in exchange for monetary compensation. Data from one participant were discarded for over 20% timeouts during the study phase. Data from a second and third participant were discarded for accuracy less than two standard deviations below the mean on the naming and colour judgement tasks, respectively.

Stimuli

Stimuli were 84 colour pictures of chairs and lamps (21 target-lure pairs for each category) obtained in the same manner as Experiment 1. Pictures were randomly sorted into three sets, each containing seven target-lure pairs from each category.

Procedure

The procedure was identical to Experiment 3, except a colour judgement task was used instead of the category induction task. In the colour judgement task, participants pressed '1' if the object in the picture was a colour, and '2' the object was neutral (gray/beige/black/white). Note that this task does not require a response based on the specific colour of the study item, and so does not necessitate elaborative processing of the specific item or explicitly require a response about a dimension that might be relevant (although a target and lure could differ in terms of colour, both items in the pair were either two different colours, or two different neutral shades).

Results

Study Phase

Accuracy for naming and colour judgements were 91.53% (*SD* = 6.60) and 85.20% (*SD* = 11.24) respectively, and this difference was not significant. A repeated-measures ANOVA revealed no difference in reliability between the study tasks (naming: 86.73%; preference: 83.80%; colour: 83.80%).

There was, however, a significant difference between the study tasks in correct RTs ($F_{2,28} = 70.721$, $p < .0001$). Paired *t*-tests between all pairs of study tasks revealed that preference judgements were slower than both naming and colour judgements (preference vs. naming: $t_{14} = 10.074$, $p < .0001$; preference vs. colour: $t_{14} = 10.007$, $p < .0001$), and there was a trend for naming judgements to be faster than colour judgements ($t_{14} = 2.056$, $p = .059$). Correct RTs for all study tasks are reported in Table 1.

Test Phase

Performance on the recognition memory test is plotted in Figure 10. A repeated-measures ANOVA on performance (d') in the recognition memory task revealed a significant effect of task ($F_{2, 28} = 8.653$, $p = .001$). Paired-sample *t*-tests comparing each pair of study tasks showed that participants had lower recognition memory for objects that were named compared to objects for which preference judgments were made ($t_{14} = 4.050$, $p = .001$) and there was a trend for worse memory performance for objects that were named compared to objects for which colour judgements were made ($t_{14} = 2.034$, $p = .061$). Recognition memory was also poorer for objects for which colour judgements were made compared to preference judgements ($t_{14} = 2.175$, $p < .05$).

Subsequent analyses of hit and false alarm rates revealed that these differences in recognition memory performance arose because of differences in hits between study task conditions ($F_{2, 28} = 17.860$, p < .0001), such that there was a significant difference in hits between all pairs of tasks (*ps* < .01). There were no significant differences in false alarms[5].

There were no significant differences in correct RTs at test between study task conditions (see Table 2).

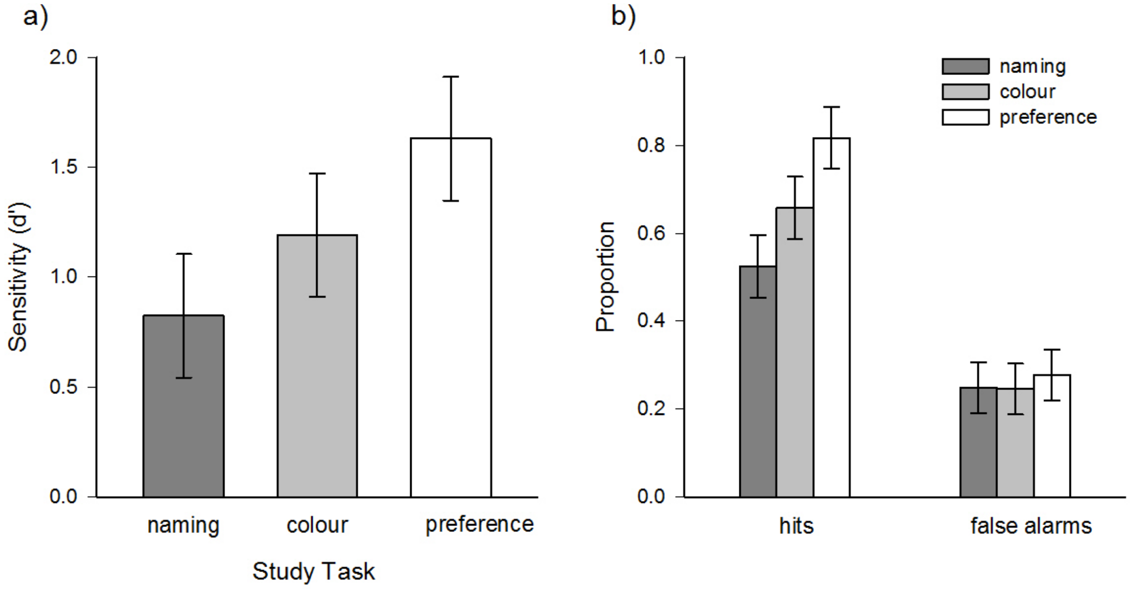

Figure 10. a) Overall performance and b) hit and false alarm rates on the recognition memory task in Experiment 5 for objects presented during each study task (naming, preference judgements, colour judgements). Error bars show 95% confidence intervals of the within-subjects effects.

Discussion

The results of Experiment 5 reveal a gradient of recognition memory performance depending on the nature of the encoding task: memory performance was greatest for objects for which preference was rated, followed by objects where a colour judgement was made, and memory was lowest for objects that were named.

---

[5] The same pattern of results was obtained when colour-change target-lure pairs were analyzed separately.

These results are consistent with a depth of processing account, where retention depends on both the qualitative nature of the encoding task as well as the number of encoding tasks (Craik, 1979; Craik & Lockhart, 1972; Craik & Tulving, 1975).

Comparisons between study tasks must be interpreted with caution, as such conclusions are limited to the tasks being compared and the specific details of the tasks themselves. For example, colour judgements are assumed to be surface-level judgements that can be made after relatively rapid sensory-level analysis (Banks & Barber, 1977; Treisman & Gormican, 1988), and this may be particularly true in Experiment 5 where participants could squint their eyes and be unable to make out the shape of the object but would still be able to detect if colour is present or absent. However, asking participants to name the specific colour of an object may lead to more elaborative processing of the visual object features. Critically, if naming is automatic then the colour judgement task combines two study tasks: automatic naming and an explicit colour judgement. Thus according to depth of processing performance is better following colour judgements vs. naming regardless of the precise nature of the colour judgement because naming is only one automatic task, whereas a colour judgement involves naming plus a colour response.

Lupyan (2008) argues that depth of processing cannot account for the naming effect because it was obtained under conditions where study tasks were randomized and participants did not know which response was required until the image was no longer displayed. However, differences in depth of processing can still arise post-stimulus presentation - just because the image is no longer present does not mean encoding has terminated. In fact, differences in response times between naming and preference were still observed under randomized study conditions in Lupyan (2008). In Experiment 5 although study tasks were blocked, there were significant differences in response times (recorded from the onset of the response probe) that match the pattern of subsequent memory performance and the predictions based on depth of processing: naming responses were fastest, followed by colour judgements, and preference judgements were slowest. If depth of processing manipulations only exert an influence on encoding when the stimulus is present, how do we account for these response time differences? Even if encoding has terminated when participants are making their response, these response time differences must still correspond with differences in processing at some level.

According to the representational shift hypothesis, performance should be worse for named objects compared to all other tasks, but there are no predictions about how performance should differ between other tasks. If anything, the representational shift hypothesis predicts no differences between other study tasks because depth of processing is assumed not to be involved. Moreover, the representational shift hypothesis cannot account for differences in study response times. Thus the results of Experiment 5 are more consistent with a depth of processing account than with the representational shift hypothesis.

Experiment 6

The representational shift hypothesis suggests that naming impairs recognition memory. However, if naming is automatic, then a depth of processing account suggests that memory differences between named objects and objects for which preference judgements are made arise because preference judgements *enhance* memory - the strength of the memory trace is enhanced for objects for which preference is rated because this is a more effortful task compared with a task that is automatic and therefore less cognitively demanding. These competing predictions are tested in Experiment 6, where participants made two judgements for each object. In this experiment, there were two groups of subjects. One group named all objects in the study phase (primary naming group). On some trials, after the naming judgement was made participants were probed to respond about the location of the image (i.e. if the image was presented above or below fixation). Importantly, this location judgement does not require additional processing of the stimulus itself and is therefore a shallow processing task. On other trials, after the naming judgement participants were probed to make a preference judgement. A second group of subjects made preference judgements for all objects (primary preference group) and either made location judgements or named the object for the second judgement.

Predictions are illustrated in Figure 11. If naming impairs performance, as suggested by the representational shift hypothesis (Lupyan, 2008), then performance will be lower in all conditions where the object is named at study, regardless of whether naming is the first or second task, compared with the single condition where no naming response is made (preference judgement followed by location judgement; Figure 11, left panel). On the other hand, if preference enhances performance, as suggested by a depth of processing account, performance should be better when the second task is preference vs. location for the primary naming group, because the preference judgement requires additional processing of the stimulus that is not required for the location judgement. In addition, if naming is automatic, then no additional processing of the stimulus should be required to generate a naming response, so naming and location judgements should have equivalent effects as secondary tasks, leading to no difference between conditions for the primary preference group (Figure 11, right panel). Therefore, both the representational shift account and depth of processing account predict a between-group interaction, but differ in the precise nature of this interaction.

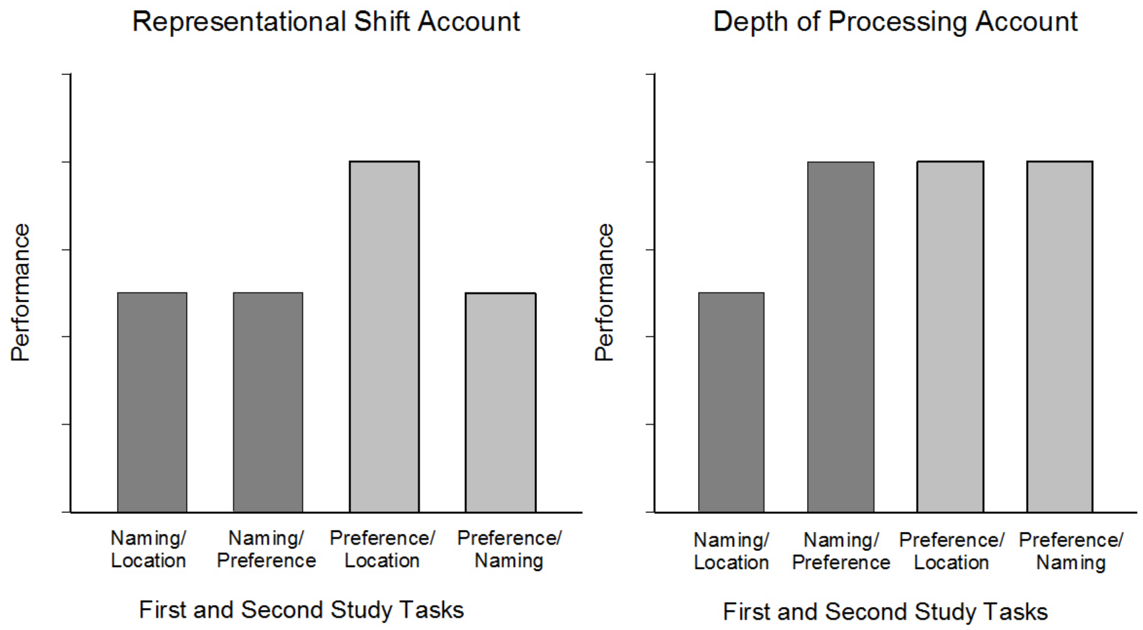

Figure 11. Predicted recognition memory performance in Experiment 6 based on the Representational Shift Account (left) and Depth of Processing Account (right).

Methods

Participants

Twenty-four members of the Vanderbilt community (9 male; mean age 22.1 years) were given monetary compensation in exchange for participation. Participants were randomly assigned to either the primary naming (n = 12) or the primary preference group.

Stimuli

Stimuli were the same chairs and lamps used as target-lure pairs in Experiment 1. Images were sorted into four sets of 20 items (five target-lure pairs for each category). For each participant, one object set (counterbalanced) was designated as the critical second task object set.

Procedure

On each trial participants saw a picture of a chair or lamp presented above or below fixation for 300ms. Participants in the primary naming group were then probed to name the object, pressing one key for “chair” and another key for “lamp”. On 75% of the trials, they were then probed to indicate the location where the object was presented relative to fixation, pressing one key for “above” and another key for “below”. On 25% of the trials (critical second task trials), participants were probed to make a preference judgement following the naming response, pressing one key for “like” and another key for “dislike”. The procedure was identical for participants in the primary preference group, except their first response was always to rate their preference for the object, and the naming response was only probed on 25% of the trials. Participants in each group knew which judgement they would always be making first (primary judgement), and although they were not informed of the exact proportion of location vs. critical second task trials, they were told that the location judgement would be probed more frequently.

For each response type, one response was made with the left hand, and the other response was made with the right hand. The same two keys were used for all response types. The response probes were the words “NAME?”, “RATE?” or “PLACE?” printed in the center of the screen for the naming, preference, and location tasks, respectively. The response probes included the two response options on the bottom left and right of the probe image to remind participants which key to press for which response. Response keys assigned to each response were kept constant across participants. Because of the constant re-mapping of response keys, there was no response deadline.

Each object was presented twice during the study phase (once above fixation and once below fixation) for a total of 80 trials. The primary judgement was followed by a location judgement on 60 trials, and the primary judgement was followed by the critical second judgement on 20 trials.

The test phase was identical to the test phase in the previous experiments.

Results

Study Data

Accuracy on the naming task (whether performed first or second) and the location task were all above 96%. Preference judgements were more reliable when the preference judgement was the primary task (86.5%) compared with when the preference judgement was the second task (72.5%; $t_{22} = 2.510$, $p < .05$).

Table 4.
Correct Response Times (and Standard Deviations) Recorded from the Onset of the Response Probe for the First and Second tasks in Experiment 6.

		Response Times (ms)
First (Primary) Task	Naming	748.18 (145.49)
	Preference	844.54 (234.08)
Second Task	Naming	891.09 (140.90)
	Preference	1189.28 (300.34)
	Location (following naming)	720.32 (216.68)
	Location (following preference)	665.43 (99.09)

Correct RTs recorded from the onset of the response probe for all primary and secondary tasks are reported in Table 4. Independent sample *t*-tests revealed that RTs for the primary task did not differ between groups. However, for the second response, preference judgements were significantly slower than naming responses (t_{22} = 3.114, p < .01). Paired-sample *t*-tests showed that location judgements were significantly faster than either critical second task (preference second: t_{11} = 13.706, p < .001; naming second: t_{11} = 21.883, p < .001). The difference in RTs for the location task did not differ between primary task groups.

Test Data

Performance on the recognition memory test is plotted in Figure 12. A 2 x 2 mixed factors ANOVA on overall performance (d') with within-subjects factor of Second Task and between-subjects factor of Primary Task revealed a significant main effect of Second Task, such that overall performance was worse when the second task was a location judgement compared with either naming or preference ($F_{1,22}$ = 29.336, p < .001). Both the representational shift and depth of processing accounts predict an interaction between primary task and secondary task (see Figure 11). Although there was no significant interaction in d', the representational account predicts that performance should be worse when the second task is naming vs. when the second task is location for the primary preference group. As can be seen from Figure 12, the opposite pattern of results was obtained, and performance for the primary preference group was actually *better* when the second task was naming vs. location (t_{11} = 2.829, p = .016). In contrast, the d' results largely support the depth of processing account, because performance was better when the second task was preference vs. location for the primary naming group (t_{11} = 4.942, p < .001).

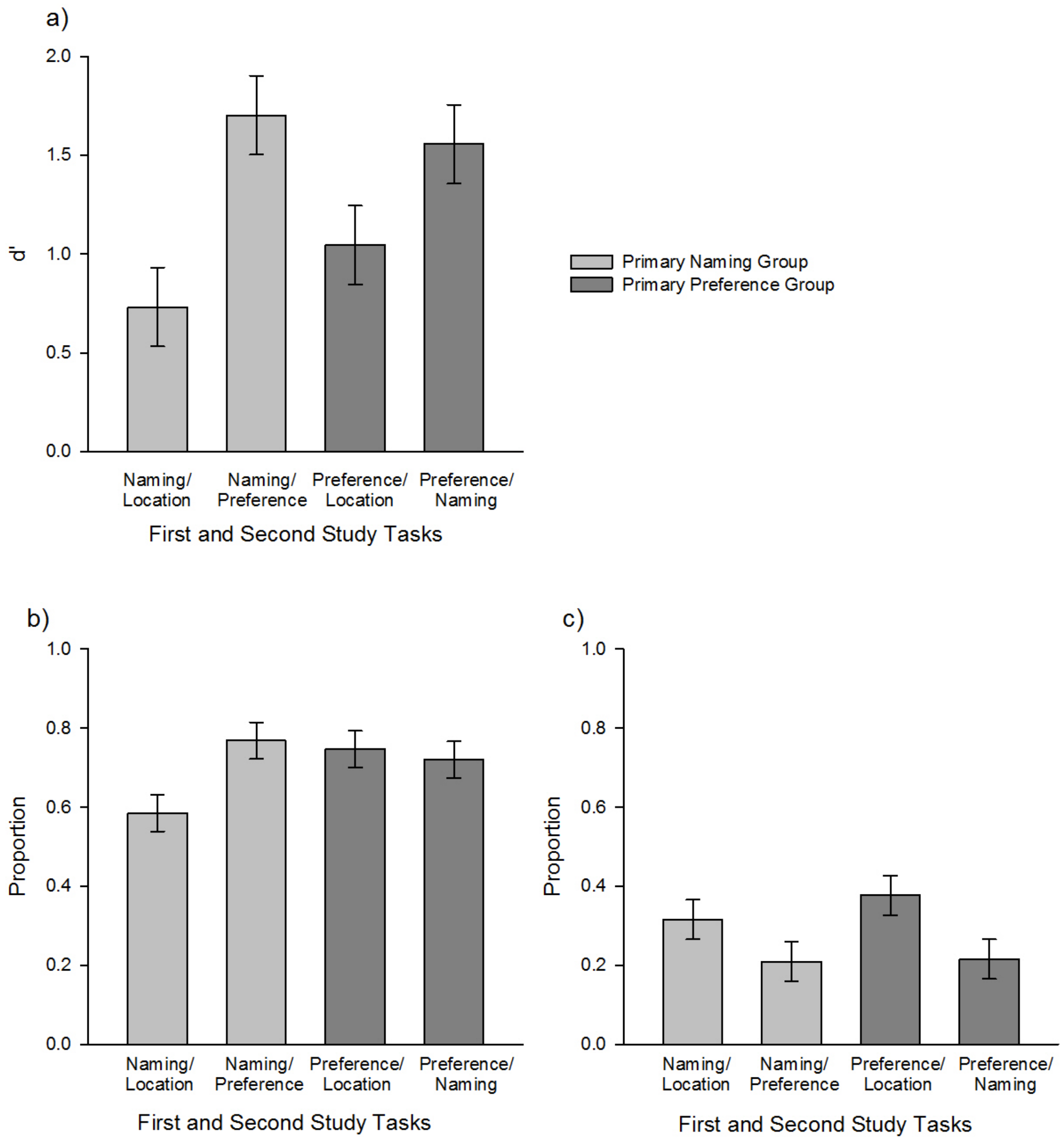

Figure 12. a) Overall performance (d'), b) hit rates and c) false alarm rates on the recognition memory test in Experiment 6 for all combinations of first (primary) and second study tasks. Error bars show 95% confidence intervals of the within-subjects effects.

The hit rate data also support a depth of processing account. A similar 2 x 2 mixed factor ANOVA conducted on hit rates revealed a main effect of Second Task ($F_{1,22}$ = 7.394, p = .01). But, critically, there was also a significant Second Task x

Primary Task interaction ($F_{1,22} = 9.737$, $p < .01$). Paired-sample t-tests revealed that, in line with the predictions based on depth of processing, there was no difference in hit rates between the two second tasks (location and naming) when the primary task was preference, but hit rates were significantly lower for the location task vs. secondary preference task when the primary task was naming ($t_{11} = 4.280$, $p = .001$). Moreover, independent sample t-tests indicated that hit rates were lower for the location task when the primary task was naming vs. when the primary task was preference ($t_{22} = 3.537$, $p < .01$), while the hit rates on the critical second task trials (naming and preference) did not differ between groups. As can be seen in Figure 12, these results indicate that hit rates were the same for any condition where a preference judgement was made, but were lower for the single condition where no preference judgement was made (naming followed by location judgement).

Table 5.
Correct Response Times (and Standard Deviations) on the Recognition Memory Task in Experiment 6 for all Combinations of First and Second Study Task.

First (Primary) Study Task	Second Study Task	Response Time (ms)
Naming	Preference	1299.44 (413.25)
	Location	1326.05 (343.86)
Preference	Naming	1412.79 (738.28)
	Location	1199.90 (282.73)

A similar ANOVA conducted on false alarm rates revealed more false alarms for the location task compared with either critical second task ($F_{1,22}$ = 13.081, $p <$ .01). Accordingly, there may not have been a between-subjects interaction in the d’ data because both groups had more false alarms for the location task, and thus lower d’ for the location task relative to the other second tasks.

Correct RTs did not differ between conditions (see Table 5).

Discussion

The results of Experiment 6 challenge the representational shift hypothesis: overall performance was actually *higher* for objects where the second task was naming vs. location judgement for the primary preference group, and hit rates were equivalent between these conditions, demonstrating that naming does not necessarily impair recognition memory. Instead, the results imply that performance differences between objects that are named and objects for which preference is rated arise because preference judgements *improve* recognition memory: hit rates were equivalent for any combination of study tasks that included a preference judgement, regardless of whether this was the first or second response; hit rates were lower for the single condition where no preference judgement was made (naming followed by location judgement). Taken together these results indicate that depth of processing provides a better account of the so-called naming effect than the representational shift hypothesis. Importantly, participants were not aware which second task they would be performing until after making their first response, well after the stimulus was no longer displayed, so these results cannot be due to differences in strategic encoding.

In addition, the hit rate data in Experiment 6 support the hypothesis that naming is a relatively automatic task – for the primary preference group adding a naming response did not produce quantitatively different results from adding a location judgment, a task which did not require additional processing of the stimulus itself. These results are consistent with a number of studies showing that participants tend to spontaneously use a verbal encoding strategy for pictures when faced with simple short term memory tasks (e.g., Coltheart, 1999; Simons, 1996; Zelinsky & Murphy, 2000), and studies suggesting that pictures of objects automatically activate their names (e.g., Meyer & Damian, 2007; Meyer et al., 2007); if participants are implicitly naming the study object anyways, the actual naming task, like the location task, does not require any additional effort or processing of the stimulus.

In summary, Chapter 4 provides evidence against the hypothesis that naming reduces recognition memory by shifting the representation of the object stored in memory towards the category prototype because it was demonstrated that naming does not necessarily impair recognition memory performance at all. Instead, the experiments in Chapter 4 support a simpler explanation for the observed difference in recognition memory between objects that are named and objects for which preference is rated: preference judgements enhance recognition memory because this is a more effortful task compared to naming which is relatively automatic, consistent with a depth of processing account.

CHAPTER V

GENERAL DISCUSSION

Lupyan (2008) proposed that overtly naming objects impairs subsequent recognition memory because the representation of the object is altered by top-down feedback invoked by the category name. This leads to a reduced hit rate for named objects because the same object shown at test no longer matches the representation stored in memory. Although the notion that conceptual information feeds back and influences lower-level processes is intriguing and has certainly been convincingly demonstrated in other studies (e.g., Curby, Hayward & Gauthier, 2004; Gauthier, James, Curby & Tarr, 2003; Goldstone, 1994; Mitterer et al., 2009), the experiments in my dissertation challenge this interpretation of the data presented by Lupyan (2008).

Chapter 3 demonstrated that overt naming is not necessary to obtain the pattern of results used to support the representational shift hypothesis: processing objects at the basic-level in the absence of an explicit naming response also led to a decrease in hits relative to objects for which preference judgements were made. In fact, processing objects at the basic-level led to performance that was quantitatively the same as naming.

Chapter 4 provided evidence that depth of processing better accounts for the difference in recognition memory for objects that are named vs. objects for which preference is rated: patterns of performance were consistent with predictions based on depth of processing and the notion that preference judgements enhance performance. More importantly, Experiment 6 revealed that naming does not universally disrupt memory performance, providing a clear falsification of the representational shift hypothesis.

Finally, although a difference in memory performance following naming vs. preference judgements was obtained in all the experiments reported here, the typicality effects reported by Lupyan (2008) were never replicated (see Experiment 1 and Appendix A). That is, there was no significant difference in the correlations between typicality and hit rates based on study task in any experiment. In fact, in Experiment 5 the trend was in the opposite direction (the correlation between typicality and hit rate was significant for objects presented during preference blocks but not named objects). Importantly, despite the fact that Lupyan (2008) uses this typicality effect as evidence for the representational shift hypothesis, the same result could be used to support a depth of processing account that assumes that naming is automatic: atypical objects may be more strongly associated with their subordinate-level name (Jolicoeur et al., 1984), so calling an atypical object by its basic-level name may be more effortful, leading to better performance for atypical objects. Thus the typicality results are not exclusively indicative of representational shift. Furthermore, there may be several reasons why I failed to replicate the typicality effects reported in Lupyan (2008), such as differences in the range of typicality in our stimulus sets (particularly in Experiment 3) and lack of power when more study tasks were used (Experiment 3 and Experiment 5), that are not theoretically important. However, one avenue for future work is to specifically test the shift-to-prototype aspect of the representational shift hypothesis that the correlation between typicality and hit rates is used to support.

Why No Differences in False Alarms?

The depth of processing account proposed in Chapter 4 is contingent on naming being automatic (e.g., Coltheart, 1999; Schiano & Watkins, 1981; Zelinsky & Murphy, 2000). This was also supported by the results of Experiment 6, where adding a naming response produced quantitatively the same performance as making a location judgement, even though determining location does not require additional processing of the study object itself. Interestingly, the automaticity of naming may explain why the memory effects in these experiments are in hit rates but not false alarm rates, a critical aspect of the representational shift hypothesis according to Lupyan (2008). The false memory literature typically finds that categorical processing influences false alarm rates when lures come from the same category as targets because of coarse or gist encoding (Koutstaal et al., 2003; Koutstaal & Schacter, 1997; Sloutsky & Fisher, 2004). However, if naming is automatic then all objects presented during study in recognition memory tasks are implicitly named. Thus, categorical effects may be equivalent for both objects that are named and objects for which preference is rated because both are named (and thus categorized) to the same extent, resulting in no difference in false alarms between these conditions. Indeed, in the first study of false memory for pictures Koutstaal & Schacter (1997) made this assumption. In this study, participants rated their preference for pictures on a five-point scale. The general finding was that false memory for studied items increased as category size at study increased. The authors suggested that this may occur because "…the studied items may have been named more readily…" (p. 570). In other words, effects of category size were presumed to occur due to implicit naming/categorization during a preference judgement task. Whether equivalent false alarm rates between study conditions is due to implicit naming could be investigated in future work using verbal interference manipulations. If naming is prevented or reduced during the preference judgement blocks, more false alarms may be observed for named objects. Future work can also investigate the extent to which the automaticity of naming influences recognition memory by comparing performance following naming to performance following passive viewing. On the one hand the act of overtly generating a response, even one where the response itself is relatively automatic, may increase processing beyond passive viewing, but it is equally possible - and would certainly be a challenge for the representational shift hypothesis – if naming and passive viewing did not differ.

Implications for Names & Naming

Rather than asking what happens when we call objects by their names, the experiments presented in my dissertation suggest that we should instead be asking what is the consequence of *always* calling objects by their names. There are several studies that demonstrate that objects are automatically named (e.g., Coltheart, 1999; Sciano & Watkins, 1981; Zelinsky & Murphy, 2000), and many studies assume this to be the case (e.g., Koutstaal & Schacter, 1997; Koutstaal et al., 2003; Roberson & Davidoff, 2000). Yet the potential influence of nameable stimuli and phonological variables are often ignored. For example, a recent fMRI study sought to explore whether there are separate processing regions for perception of form and texture in the ventral stream (Cant, Arnott & Goodale, 2009). However, the object shapes were novel (and therefore had no known names), but participants were exposed to the textures *with names* prior to the experiment. Thus the two features of interest differed in whether or not participants could name them and this could have influenced their results.

In addition, studies that make claims about conceptual categories influencing perception need to test whether these effects actually arise due to having names that are automatically accessed. For example, Lupyan, Thompson-Schill & Swingley (in press) found that participants took longer to say that the letters “B” and “b” were not physically identical (compared with “B” and “p”) when the letters were presented sequentially. They suggested that this effect is due to the fact that “B” and “b” are perceptually more similar because they are from the same conceptual category. However, an alternative explanation is that both “B” and “b” automatically activate the same name, and this match at the phonetic level conflicts with the mismatch at the perceptual level, leading to interference in producing the required “different” response (cf. Posner, 1978).

The present results focus on naming at the basic-level, and certainly the basic-level is the level at which we parse the world because it is the most informative and efficient (Rosch et al., 1976). However, objects can be named at different levels of abstraction. For example, while I consider all of my stimuli chairs and lamps, someone who works at Ikea might call each chair by its unique name:

Poang, Ektorp, and so on. One question for future work is whether overtly naming objects at a level of abstraction beyond the basic-level influences performance, and whether subordinate or superordinate level names can become automatically activated in the same way as basic-level names. Indeed, atypical exemplars are more readily named at the subordinate level (Jolicoeur et al., 1984), and a behavioral hallmark of perceptual expertise is the entry-level shift, where subordinate-level names are accessed as rapidly as basic-level names (Tanaka & Taylor, 1991). With expertise are both the basic-level and subordinate-level names automatically activated in response to objects? Do both types of names influence performance equally? Although a recent study suggests that individuation is sufficient to produce perceptual expertise without learning individual names (Bukach, Vickery, Kinka & Gauthier, under revision), this only suggests that names are not necessary for the acquisition of perceptual expertise, and does not rule out the possibility that subordinate-level names do contribute to performance in real-world experts when they are available.

The idea that pictures of objects automatically evoke their names and that the level of naming (basic vs. subordinate) can depend on typicality and expertise is consistent with the idea that automatic processing biases are the result of experience associating different tasks with different categories of objects (Wong, Palmeri & Gauthier, 2009). But what about automatic responses to objects that are not names? On the one hand automatic naming may be no different than another type of automatic response that arises due to experience associating that response with an object or category of objects. For example, there is both behavioral and neural evidence that action affordances are automatically activated when pictures of objects are viewed (e.g., Ellis & Tucker, 2000; Grezes, Tucker, Armony, Ellis & Passingham, 2003; Tucker & Ellis, 2000); experience reaching for and grasping objects automatically prepares the appropriate action response. However, the developmental studies reviewed in Chapter 1 suggest that verbal labels have a special status in infant category learning, and this advantage for category labels appears to persist into adulthood (e.g., Lupyan et al., 2007; Yamauchi & Markman, 2000). Thus, it may be that verbal responses to objects – names – maintain a special status throughout the lifespan, and naming objects is an automatic response that differs from other learned automatic responses.

General Implications for The Study of Memory

The experiments in Chapter 4 demonstrating that an effect originally attributed to a naming-specific mechanism can be accounted for by depth of processing speaks to a larger issue in the study of memory. Although it is tempting to propose hypotheses that are task-specific to account for memory differences, long-established general principles of memory should not be overlooked. For example, the research on the verbal overshadowing effect bears many similarities to the work presented here in this respect. The verbal overshadowing effect refers to the finding that verbally describing a face interferes with later identification of that face. The initial explanation for this effect was that verbalization led to the formation of a verbally recoded memory representation (Schooler & Engstler-Schooler, 1990). Similar to the representational shift hypothesis, decreased identification accuracy was thought to arise because this recoded memory representation would no longer match the target item when presented at test. However, follow-up work suggested that this effect – initially described in terms of modified representations – occurs because verbalization produces a general processing shift away from the non-verbal processing operations that are critical to face recognition (see Schooler, 2002, for a review), an explanation that is consistent with the transfer appropriate processing framework. Like the verbal overshadowing effect, decreased memory for named objects can be accounted for by a general principle of memory rather than a task-specific explanation.

Conclusion

Names have been shown to impact various aspects of performance in both perceptual and cognitive tasks: names influence perception (e.g., Gilbert et al., 2006; 2008; Goldstein & Davidoff, 2008; Roberson & Davidoff, 2000; Roberson et al., 2000; 2008), visual search (e.g., Lupyan, 2008b; Lupyan & Spivey, 2008; Spivey et al., 2001), category learning (e.g., Balaban & Waxman, 1997; Colunga & Smith, 2002; Fulkerson & Waxman, 2007; Lupyan et al., 2007; Plunkett et al., 2008; Waxman & Markow, 1995; Woodward & Hoyne, 1999; Xu, 2002; Yoshida & Smith, 2005) and recognition memory (e.g., Koutstaal et al., 2003; Koutstaal & Cavendish, 2006; Lupyan, 2008; Musen, 1991). In many cases names may exert an influence on performance because they are automatically activated in response to objects (e.g., Meyer & Damian, 2007; Meyer et al., 2007), and people automatically name objects (e.g., Noizet & Pynte, 1976; Zelinsky & Murphy, 2000). I have shown that explicitly calling objects by their names at study does not have a unique influence on memory performance in and of itself. Instead, naming leads to worse memory performance relative to other tasks because naming is an automatic instantiation of basic-level categorization and is thus a shallow processing task.

APPENDIX A

TYPICALITY EFFECTS AT TEST FOR EXPERIMENTS 2-6

Experiment 2

Typicality ratings were collected from a separate group of eight Vanderbilt Undergraduates (1 Male; mean age 19.1 years). The procedure for collecting typicality ratings was the same as Experiment 1 with the exception that each item was only presented once.

Average typicality was 2.66 for exemplar lure pairs and 2.60 for state lure pairs and this difference was not significant.

The correlations between typicality and hit rate and typicality and false alarm rate are plotted in Figure 13. Typicality correlated with both average hit rate ($r_{80} = -.315$, $p < .01$) and average false alarm rate ($r_{80} = .269$, $p = .016$). Curiously, the direction of these correlations is the opposite of the previous experiments and what would be expected based on previous work: hit rates were higher for typical objects, and false alarm rates were higher for atypical objects.

The correlation between typicality and hit rate was significant for objects for which preference was rated ($r_{80} = -.342$, $p < .01$) but not objects that were named ($r_{80} = -.173$). However, when typicality was entered as a covariate in a general linear model to predict hit rates, the interaction between typicality and study task was not significant ($p > .3$).

The correlation between typicality and false alarm rate was significant for objects that were named (r_{80} = .238, p < .05) and the correlation approached significance for objects for which preference was rated (r_{80} = .216 p = .054). However, when typicality was entered as a covariate in a general linear model to predict false alarm rates, the interaction between typicality and study task was not significant (*p* > .6).

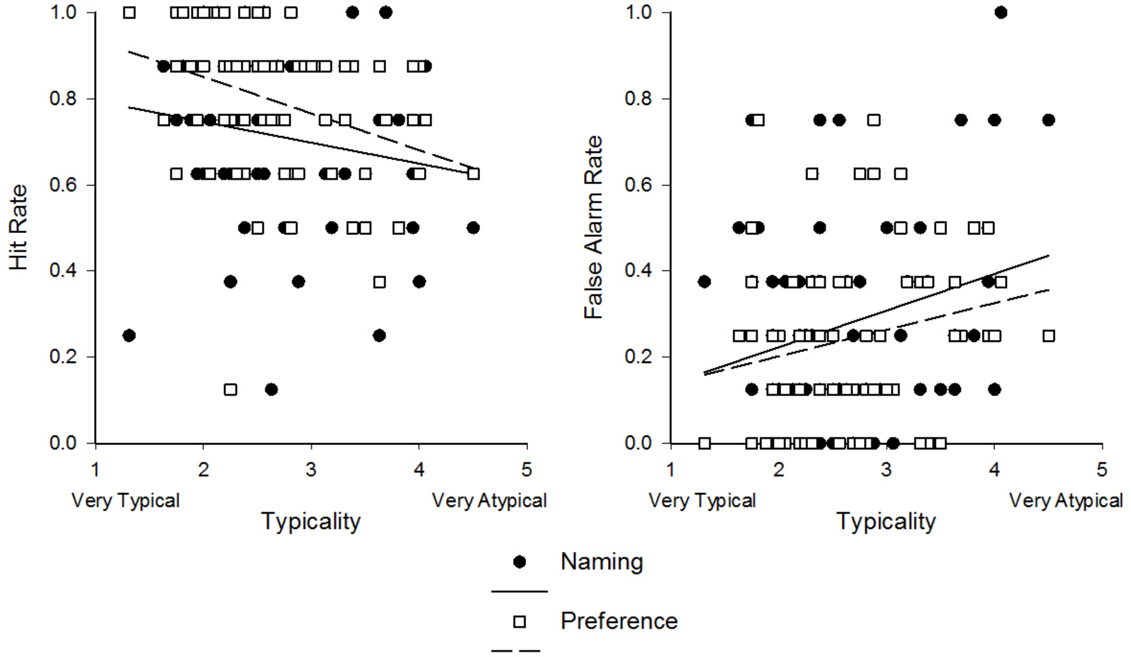

Figure 13. Correlation between Typicality and Hit Rate (left) and Typicality and False Alarm Rate (right) for objects that were named (black circle, solid line) and objects for which preference was rated (white square, dashed line) during the study phase of Experiment 2.

Experiment 3

Typicality ratings were collected from a separate group of eight Vanderbilt University undergraduates (one male; mean age years 19.13 years). The procedure for acquiring typicality ratings was the same as Experiment 1.

Average typicality ratings for birds and dogs were 2.58 and 2.33, respectively. The difference in typicality ratings between birds and dogs was not significant.

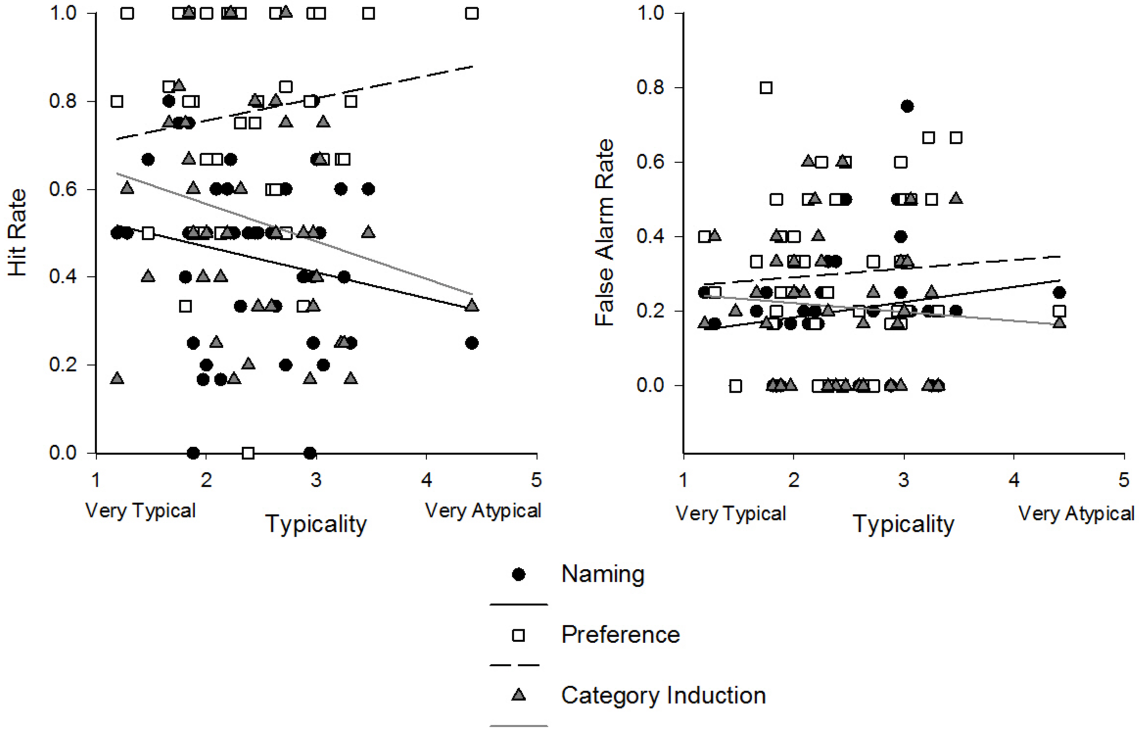

Figure 14. Correlation between Typicality and Hit Rate (left) and correlation between Typicality and False Alarm Rate (right) for objects presented during naming (black circle, solid black line), preference judgement (white square, dashed line) and category induction (gray triangle, solid gray line) study blocks in Experiment 3.

Correlations between typicality and hit rate and correlations between typicality and false alarm rate are plotted in Figure 14. Typicality did not correlate with either average hit rate, nor did typicality correlate with hit rate for any study condition (naming: r_{42} = -.188; preference: r_{42} = .139; category induction: r_{42} = -.225). When typicality was entered as a covariate in a general linear model to predict hit rates the interaction between study task and typicality was not significant ($p > .1$).

Typicality did not correlate with average false alarm rate, nor did typicality correlate with false alarm rate for any study condition (naming: $r_{42} = .154$; preference: $r_{42} = .070$; category induction: $r_{42} = -.083$). When typicality was entered as a covariate in a general linear model to predict false alarm rates the interaction between study task and typicality was not significant ($p > .5$).

Experiment 4

Typicality ratings were obtained in the same manner as Experiment 1 from a different group of eight Vanderbilt Undergraduates (4 male; mean age 21.3 years).

Average typicality ratings for chairs and lamps was 2.93 and 2.91, respectively. The difference in typicality ratings for chairs and lamps was not significant.

Correlations between typicality and hit and false alarm rates for category-matching and exemplar-matching are plotted in Figure 15. Typicality did not correlate with average hit rates ($r_{40} = .011$), nor did typicality correlate with hit rates for objects presented in the category-matching ($r_{40} = .259$) or exemplar-matching ($r_{40} = -.108$) blocks. When typicality was entered as a covariate in a general linear model to predict hit rates the interaction between study task and typicality was not significant ($p > .3$).

Typicality correlated with average false alarm rates ($r_{40} = -.345$, $p < .05$) such that false alarm rates were higher for more typical objects. This correlation was significant for objects presented during exemplar-matching (r_{40} = -.319, *p* < .05) and approached significance for objects presented during category-matching (r_{40} = -.295, *p* = .064). When typicality was entered as a covariate in a general linear model to predict false alarm rates, the interaction between study task and typicality was not significant (*p* > .7).

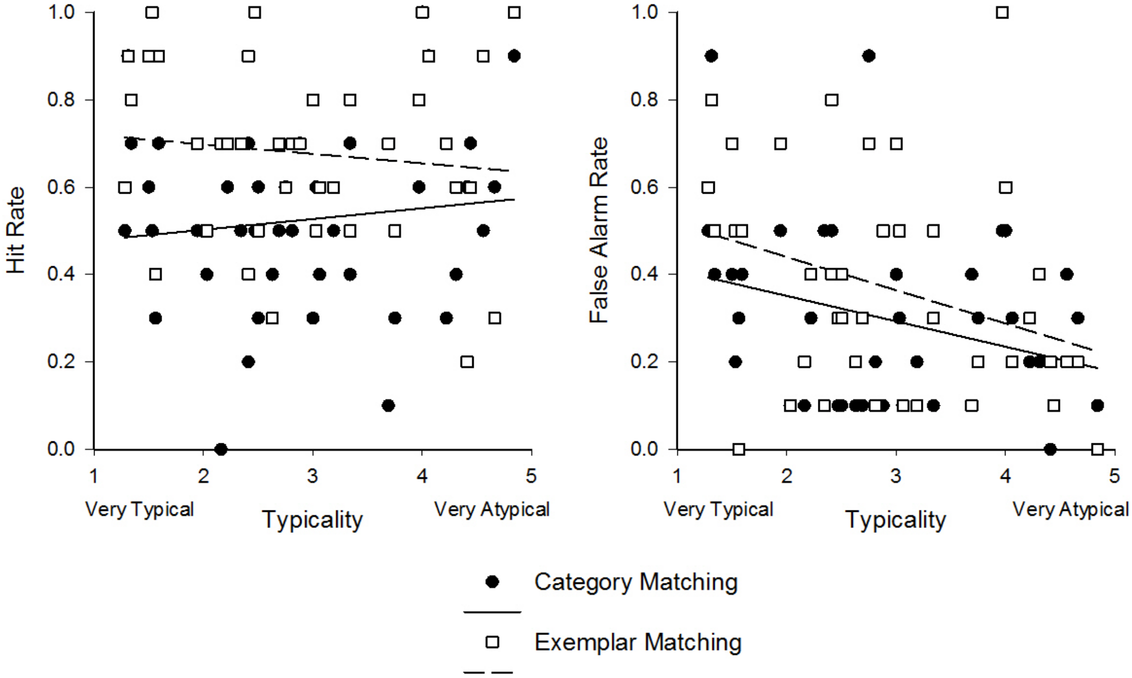

Figure 15. Correlations between Typicality and Hit Rate (left) and Typicality and False Alarm Rate (right) for objects presented during category-matching and exemplar-matching study blocks in Experiment 4.

Experiment 5

Typicality ratings were the same as those used in Experiment 1

The correlations between typicality and hit rate and typicality and false alarm rates are plotted in Figure 16. Typicality did not correlate with average hit rate or average false alarm rate. Moreover, when typicality was entered as a covariate in a general linear model to predict hit rates, the interaction between study task and typicality was not significant ($p > .3$). Interestingly, however, in contrast to previous experiments, the correlation between hit rate and typicality was significant for preference objects ($r_{40} = .338$, $p < .05$) but not named objects ($r_{40} = .137$) or objects for which colour judgements were made ($r_{40} = -.046$) . False alarm rates were not correlated with typicality for any study task, nor was the interaction between study task and typicality significant when typicality was entered as a covariate to predict false alarm rates ($p > .9$).

.

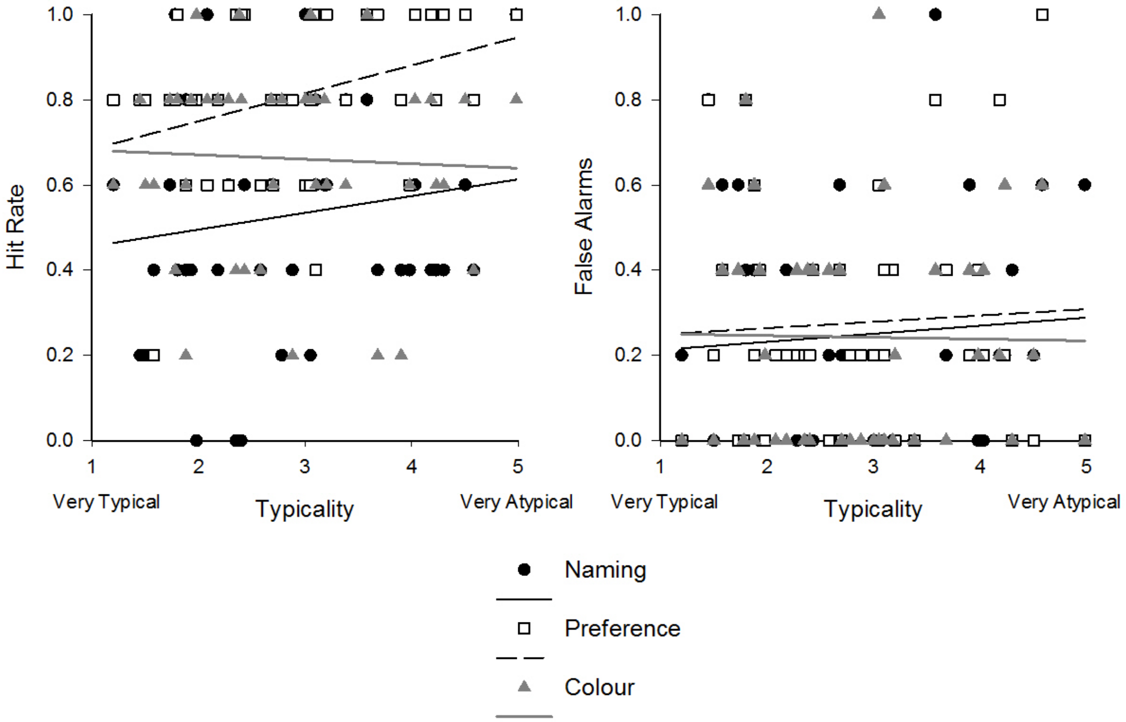

Figure 16. Correlations between Typicality and Hit Rate (left) and correlations between Typicality and False Alarm Rate (right) for objects presented during the naming (black circle, black solid line), preference judgement (white square, dashed line) and colour judgement (gray triangle, gray line) study blocks in Experiment 5.

Experiment 6

Typicality ratings were the same as those used in Experiment 1.

Due to the manner in which object sets were assigned to conditions, there were not enough data points per object to calculate correlations for the naming/preference secondary tasks and the location secondary task separately. When hit and false alarm rates were combined for both secondary tasks for each primary task group (see Figure 17), the correlation between average hit rate and typicality approached significance ($r_{40} = .299$, $p = .061$). Although the correlation between typicality and hit rate was significant for the primary naming group ($r_{40} = .325$, $p < .05$) but not the primary preference group ($r_{40} = .178$), when typicality was entered as a covariate in a general linear model to predict hit rates, the interaction between primary task and typicality was not significant ($p > .3$).

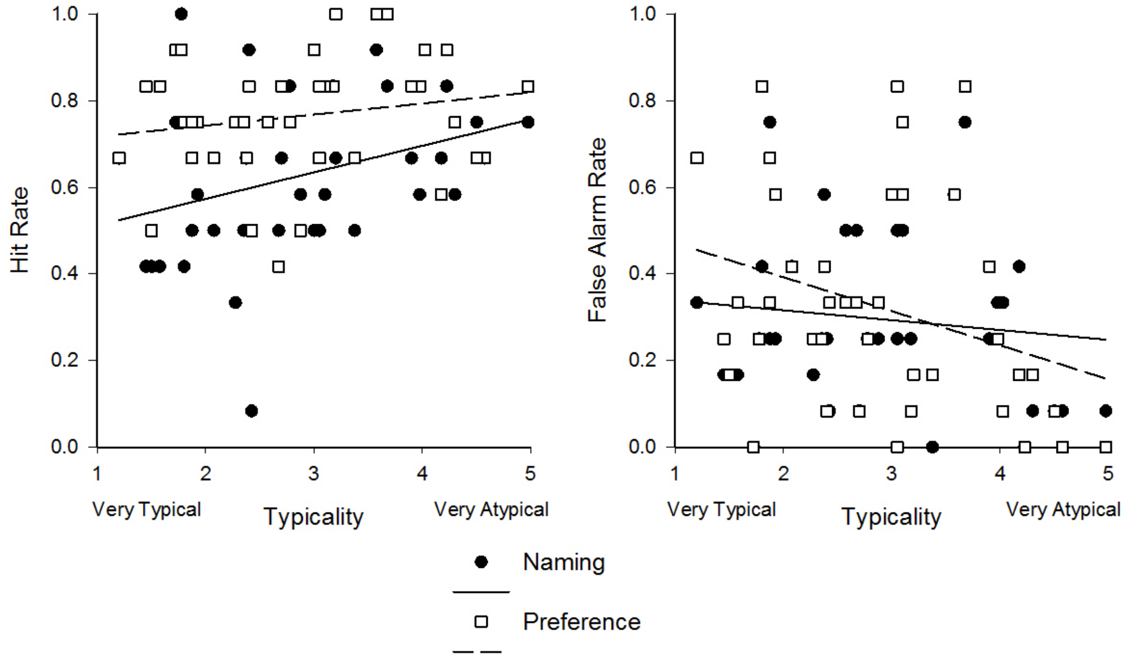

Figure 17. Correlation between Typicality and Hit Rates (left) and correlation between Typicality and False Alarm Rate (right) for objects where the primary study task was naming (black circle, solid line) or where the primary study task was preference judgements (white square, dashed line) in Experiment 6.

Average false alarm rate did not correlate with typicality. Although the correlation between typicality and false alarm rate approached significance for the primary preference group ($r_{40} = -.309$, $p = .053$) but not the primary naming group ($r_{40} = -.112$), the interaction between typicality and primary task group was not significant when typicality was entered as a covariate in a general linear model to predict false alarm rates ($p > .2$).

REFERENCES

Anderson, J.R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451-474.

Anderson, J.R., & Reder, L.M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General, 128,* 186-197.

Baddeley, A.D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review, 85,* 139-152.

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

Banks, W.P., & Barber, G. (1977). Color information in iconic memory. *Psychological Review, 84,* 536-546.

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.

Booth, A.E., & Waxman, S. (1997). Object names and object functions serve as cues to categorize for infants. *Developmental Psychology, 38,* 948-967.

Brady, T.F., Konkle, T., Alvarez, G.A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105,* 14325-14329.

Bukach, C.M., Vickery, T., Kinka, D., & Gauthier, I. (under revision). Individuation without naming is worth it.

Busch, N.A., Herrmann, C.S., Muller, M.M., Lenz, D., & Gruber, T. (2006). A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. *Neuroimage, 33,* 1169-1177.

Cant, J.S., Arnott, S.R., & Goodale, M.A. (2009). fMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Experimental Brain Research, 192,* 391-405.

Cabrera, A., & Billman, D. (1996). Language-driven concept learning: Deciphering Jabberwocky. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 539-555.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Memory & Cognition, 21,* 539-545.

Coltheart, V. (1999). Comparing short-term memory and memory for rapidly presented visual stimuli. *International Journal of Psychology, 34,* 293-300.

Colunga, E., & Smith, L.B. (2002). What makes a word? In W.D. Gray & C. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 214-219). Austin: Cognitive Science Society.

Colunga, E., & Smith, L.B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112,* 347-382.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75-84.

Craik, F.I.M. (1979). Human memory. *Annual Reviews Psychology, 30,* 63-102.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal behavior, 11,* 671-684.

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268-294.

Curby, K.M., Hayward, W.G., & Gauthier, I. (2004). Laterality effects in the recognition of depth-rotated novel objects. *Cognitive, Affective & Behavioral Neuroscience, 4,* 100-111.

Davidoff, J. (2001). Language and perceptual categorization. *Trends in Cognitive Sciences, 5,* 382-387.

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science, 20,* 252-257.

Ehret, G. (1987). Categorical perception of sound signals: Facts and hypotheses from animal studies. In S. Harnad (Ed.), *Categorical Perception* (pp. 301-331). New York: Cambridge University Press.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology, 91,* 451-471.

Franklin, A., Drivonikou, G.V., Clifford, A., Kay, P., Regier, T., & Davies, I.R.L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Science, 105,* 18221-18225.

Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Science, 105,* 3221-3225*.*

Fulkerson, A.L., & Waxman, S.R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition, 105,* 218-228.

Gauthier, I., James, T.W., Curby, K.M., & Tarr, M.J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology, 20,* 507-523.

Gelman, S.A., & Markman, E.M. (1986). Categories and induction in young children. *Cognition, 23,* 183-209.

Gelman, S.A., & Markman, E.M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58,* 1532-1541.

Gilbert, A.L., Regier, T., Kay, P., & Ivry, R.B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Science, 103,* 489-494.

Gilbert, A.L., Regier, T., Kay, P., Ivry, R.B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language, 105,* 91-98.

Gliga, T., Volein, A., & Csibra, G. (in press). Verbal labels modulate perceptual object processing in 1-year-old infants. *Journal of Cognitive Neuroscience.*

Gliozzi, V., Mayor, J., Hu, J.-F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science, 33,* 709-738.

Goldstein, J., & Davidoff, J. (2008). Categorical perception of animal patterns. *British Journal of Psychology, 99,* 229-243.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123,* 178-200.

Grezes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R.E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *European Journal of Neuroscience, 17,* 2735-2740.

Grill-Spector, K. & Kanwisher, N. (2005). Visual Recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152-160.

Gruber, T., Muller, M.M., & Keil, A. (2002). Modulation of induced gamma band responses in a perceptual learning task in human EEG. *Journal of Cognitive Neuroscience, 14,* 732-744.

Harnad, S. (1987). *Categorical perception: The groundwork of cognition.* New York: Cambridge University Press.

Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

Jolicoeur, P., Gluck, M.A., & Kosslyn, S.M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16,* 243-275.

Keates, J., & Graham, S.A. (2008). Category markers or attributes: Why do labels guide infants' inductive inferences? *Psychological Science, 19,* 1287-1293.

Kikutani, M., Roberson, D., & Hanley, J.R. (2008). What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review, 15,* 787-794.

Klein, K., & Saltz, E. (1976). Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 671-679.

Koutstaal, W. (2006). Flexible remembering. *Psychonomic Bulletin & Review, 13,* 84-91.

Koutstaal, W., & Cavendish, M. (2006). Using what we know: Consequences of intentionally retrieving gist versus item-specific information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 778-791.

Koutstaal, W., & Schacter, D.L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37,* 555-583.

Koutstaal, W., Reddy, C., Jackson, E.M., Prince, S., Cendan, D.L., & Schacter, D.L. (2003). False recognition of abstract versus common objects in older and younger adults: Testing the semantic categorization account. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 499-510.

Lehrner, J.P. (1993). Gender difference in long-term odor recognition memory: Verbal versus sensory influences and consistency of label use. *Chemical Senses, 18,* 17-26.

Lehrner, J.P., Gluck, J., & Laska, M. (1999). Odor identification, consistency of label use, olfactory threshold and their relationships to odor memory over the human lifespan. *Chemical Senses, 24,* 337-346.

Lockhart, R.S., & Craik, F.I.M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology, 44,* 87-112.

Logan, G.D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39,* 367-386.

Lumeng, J.C., Zuckerman, M.D., Cardinal, T., & Kaciroti, N. (2005). The association between flavor labeling and flavor recall ability in children. *Chemical Senses, 30,* 565-574.

Lupyan, G. (2005). Carving nature at its joints and carving joints into nature: How labels augment category representations. In A. Cangelosi, G. Bugmann & R. Borisyuk (Eds.), *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop* (pp. 87-96). Singapore: World Scientific.

Lupyan, G. (2008). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General, 137,* 348-369.

Lupyan, G. (2008b). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108,* 566-577.

Lupyan, G., Rakison, D.H., & McClelland, J.L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science, 18,* 1077-1083.

Lupyan, G., & Spivey, M.J. (2008) Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology, 18,* R410-R412.

Lupyan, G., Thompson-Schill, S.L., & Swingley, D. (in press). Conceptual penetration of visual processing. *Psychological Science.*

Mack, M.L., Gauthier, I., Sadr, J., & Palmeri, T.J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review, 15,* 28-35.

Maki, S.L., McDaniel, W.F., Boyce, K.C., Brown, C.M., Crane, S.M., Cundey, J., Kell, R.G., & Marcengill, P.R. (2001). Visual categorical perception by rats with temporal, striate, or sham ablations. *Cognitive Neuroscience and Neuropsychology, 12,* 3425-3431.

Malmberg, K.J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57,* 335-384.

Mandler, G., Pearlstone, Z., & Koopmans, H.S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8,* 410-423.

Marks, W. (1991). Effects of encoding the perceptual features of pictures on memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 566-577.

Masharov, M., & Fischer, M.H. (2006). Linguistic relativity: Does language help or hinder perception? *Current Biology, 16,* R289-R291.

Meyer, A.S., Belke, E., Telling, A.L., Humphreys, G.W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14,* 710-716.

Meyer, A.S., & Damian, M.F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35,* 494-503.

Mitterer, H., Horschig, J.M., Musseler, J., & Majid, A. (2009). The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35,* 1557-1562.

Mondy, S., & Coltheart, V. (2006). Phonological similarity effects on detecting change in simple arrays. *Memory & Cognition, 34,* 1744-1753.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

Morse, P.A., & Snowdon, C.T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics, 17,* 9-16.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 555-563.

Moscovitch, M., & Craik, F.I.M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 447-458.

Musen, G. (1991). Effects of verbal labeling and exposure duration on implicit memory for visual patterns. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 954-962.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53,* 359-377.

Nelson, D.A., & Marler, P. (1989). Categorical perception of a natural stimulus continuum: Birdsong. *Science, 244,* 976-978.

Noizet, G., & Pynte, J. (1976). Implicit labeling and readiness for pronunciation during the perceptual process. *Perception, 5,* 217-223.

Ozgen, E. (2004). Language, learning, and color perception. *Current Directions in Psychological Science, 13,* 95-98.

Plunkett, K., Hu, J-F., & Cohen, L.B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106,* 665-681.

Posner, M.I. (1978). *Chonometric Explorations of Mind*. Englewood Heights, NJ: Erlbaum. 271 pp.

Reder, L.M., & Anderson, J.R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology, 12,* 447-472.

Reder, L.M., Donavos, D.K., & Erickson, M.A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition, 30,* 312-323.

Rehder, B., & Hastie, R. (2001). Casual knowledge and categories: The effects of causal beliefs on categorization, induction and similarity. *Journal of Experimental Psychology: General, 130,* 323-360.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28,* 977-986.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorization: Neuropsychological evidence for a dissociation in explicit categorization tasks. *Cognition, 71,* 1-42.

Roberson, D., Pak, H., & Hanley, J.R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition, 107,* 752-762.

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

Ross, B.H., & Murphy, G.L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 736-753.

Seamon, J.G., Luo, C.R., Schlegal, S.E., Greene, S.E., & Goldenberg, A.B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory and Language, 42,* 120-146.

Schiano, D.J., & Watkins, M.J. (1981). Speech-like coding of pictures in short-term memory. *Memory & Cognition, 9,* 110-114.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic information processing: I. Detection, search and attention. *Psychological Review, 84,* 1-66.

Schooler, J.W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16,* 989-997.

Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 17,* 36-71.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Simons, D.J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301-305.

Sloutsky, V.M., & Fisher, A.V. (2004). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science, 15,* 553-558.

Sloutsky, V.M., & Robinson, C.W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science, 32,* 354-377.

Spivey, M.J., Tyler, M.J., Eberhard, K.M., & Tanenhaus, M.K. (2001). Linguistically mediated visual search. *Psychological Science, 12,* 282-286.

Symons, C.S., & Johnson, B.T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121,* 371-394.

Tan, L.H., Chan, A.H.D., Kay, P., Khong, P.-L., Yip, L.K.C., Luke, K.-K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Science, 105,* 4004-4009.

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

Treisman, A., & Gormican, S. (1988). Feature analysis in vision: Evidence from search asymmetries. *Psychological Review, 95,* 15-48.

Tucker, M., & Ellis, R. (2000). Action priming by briefly presented objects. *Acta Psychologica, 116,* 185-203.

Waxman, S.R., & Markow, D.B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29,* 257, 302.

Wilson, M. & DeBauche, B.A. (1981). Inferotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia, 19,* 29-41.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science, 104,* 7780-7785.

Wong, A.C.-N., Palmeri, T.J., & Gauthier, I. (2009). Conditions for facelike expertise with objects: Becoming a Ziggerin expert – but which type? *Psychological Science, 20,* 1108-1117.

Woodward, A.L., & Hoyne, K.L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70,* 65-77.

Wyttenback, R.A., May, M.L., & Hoy, R.R. (1996). Categorical perception of sound frequency by crickets. *Science, 273,* 1542-1544.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85,* 223-250.

Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science, 16,* 372-377.

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

Yoshida, H., & Smith, L.B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16,* 90-95.

Zelinsky, G.J., & Murphy, G.L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science, 11,* 125-131.