Measurement of Semantic Knowledge and Its Contribution to Object Recognition Performance

By

Ana Elizabeth Van Gulick

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, 2014

Nashville, Tennessee

Approved:

Isabel Gauthier, Ph.D.

Thomas J. Palmeri, Ph.D.

Timothy P. McNamara, Ph.D.

James W. Tanaka, Ph.D.

To my parents, Robert and Virginia.

Thank you for your unconditional love and support. Your dedication to teaching me and your encouragement to be creative, hard working, and curious about the world has helped me learn and explore and has led me to where I am today. From creating the language of Parsinian, to backyard inventions and restaurants, perfecting the naked egg drop for Science Olympiad, and encouraging me to take my first cognitive psychology class and work in a lab, I could not have had better mentors and role models.

ACKNOWLEDGEMENTS

This work was supported by NSF (SBE-0542013), the Vanderbilt Vision Research Center (P30-EY008126), and the National Eye Institute (R01 EY013441).

Thanks to Jackie Floyd, Yimin Qiu, David Nelwan, and Amit Khandahadia for their assistance with data collection, and Simen Hagen, Jianhong Shen, Ryan Farrell, and Tom McKay for assistance recruiting birders. Thanks to Jenn Richler for comments on the work. Thanks to the members of my dissertation committee, Tom Palmeri, Tim McNamara, and Jim Tanaka, for their feedback on this project and their support. Thank you to all of the members of the Object Perception Lab and the CatLab who I have had the pleasure of working with these past five years. I am honored to have been a part of these labs and the Vanderbilt Psychology department. No matter where life takes me, my respect for science and the people who do it will always be influenced by your passion and dedication.

Thank you to my advisor, Isabel Gauthier, to whom I am so grateful for her extraordinary mentorship. Isabel, you have made this experience rich and wonderful. You are an inspiration to me in every part of your life. Your tireless efforts and dedication to teaching your students will stay with me forever, as will your genuine excitement for science. It is truly bittersweet to leave behind your advising and your lab that have felt like home.

Thank you to John Pyles for his love and support these last four years (and to the Perceptual Expertise Network for introducing us). John, this journey would not have been the same without you. From near and far you have been there for me both in times of struggle and success; your encouragement has meant so much. I look forward to our life together as Dr. and Dr. And lastly, thanks to Teddy the cat, the best feline companion a graduate student could ask for, even if he prefers sleeping on papers to writing them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 – Introduction

Introduction to the question

"What's in a name? That which we call a rose by any other name would smell as sweet." – William Shakespeare

Perhaps this is true, a rose would look and smell lovely to anyone, but if you had extensive experience with roses and were able to name it even more specifically as a particular species of rose then your recognition of that rose might be different. You might notice particular distinguishing features, that it was deep pink and strongly scented with cup-shaped blooms atop firm stems, and thus name it not just as a rose but an American Beauty. Your ability to name the rose by species might reflect your level of experience with roses; maybe you are a florist or a gardener. But the American Beauty is also a popular and well-known rose, it was featured in a movie with the same name, is the official flower of Washington D.C., and appears in the Lord & Taylor logo, so you may recognize it even if you have only a small amount of experience with roses. However, recognizing a rarer rose by name, say an Audrey Wilcox or Majestic rose, would suggest a high level of rose experience.

Our experience with objects in the world continuously shapes how we perceive, label, and act upon them (Bub, Masson, & Lin, 2013; Dehaene et al., 2010; Folstein, Palmeri, & Gauthier, 2013; Gauthier, Skudlarski, Gore, & Anderson, 2000; Goldstone, 1994; Goldstone & Styvers, 2001; K. H. James, James, Jobard, Wong, & Gauthier, 2005; Tanaka, Curran, & Sheinberg, 2005; Van Gulick & Gauthier, in press; A. C.-N. Wong, Palmeri, & Gauthier, 2009a; A. C.-N. Wong, Palmeri, Rogers, Gore, & Gauthier, 2009b; Y. K. Wong, Folstein, & Gauthier, 2012; Xu, 2005). In the field of high-level vision, we often measure performance on object recognition tasks in hopes of understanding the perceptual and cognitive processes that contribute to successful object recognition. When we quantify *performance* on these tasks, we sometimes want to know not only who performs best, but also why one person performs better than another. For the latter goal, we measure task *performance* to estimate an individual's underlying *ability*, the stable, underlying trait that contributes to observed performance. For example, we might measure visual recognition performance for birds. However, performance by itself is difficult to interpret. Performance, especially domain-specific performance, is likely driven by *experience* in that domain and a person's domain-general visual *ability*. Therefore, accounting for the effect of experience is important when we compare performance between two people to make inferences about their relative ability: if Liz performs better than Theo on our measure of bird recognition, is that because she has better general visual ability or more bird experience? If we wish to know someone's true object recognition ability, we must be able to quantify experience as well, in a way that will allow us to separate the contributions of experience and ability to observed performance.

Experience can be measured directly, for example through self-report, although this can be biased and is limited to introspection. Experience can also be estimated by looking at its effects on the acquisition of other kinds of knowledge. As category experience accumulates over time, experts are likely to acquire both visual and semantic category knowledge. The Vanderbilt Expertise Test (VET; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012b) was created to measure visual knowledge for a variety of categories, but there is currently no non-visual test of semantic knowledge for common object categories.

The goal of the following studies is to create a measure of semantic knowledge for specific categories, the Semantic Vanderbilt Expertise Test (SVET). This measure will be in a standardized format that can be used to test knowledge of object names for many different categories and that can be completed easily by subjects, including those with little semantic knowledge. We will evaluate the measure's reliability and validity for each category, and test whether it can be used to capture individual differences in knowledge across the range of experience from novice to expert. The resulting set of tests will be a novel and valuable tool to measure semantic knowledge that will enhance our ability to study object learning and expertise.

However, just as when measuring visual performance for birds, if Theo performs better than Liz on a test of semantic bird knowledge, we cannot say if this is because he has greater domain-general ability, such as verbal ability or general intelligence, or because he has more bird experience that led to the acquisition of domain-specific semantic knowledge. It is important to disentangle the contributions of domain-general ability and domain-specific experience to understand both visual and semantic performance. Thus, the SVET will offer a way to measure semantic knowledge and, when paired with another domain-specific measure such as a visual test (VET; McGugin et al., 2012b), will allow us to compare performance on two different tasks that are both influenced by domain-specific experience.

Testing many different categories is important for this goal so that we can assess the relationship between experience and performance within a category and compare an individual's within category performance to their average experience and performance across many categories. This comparison is needed to determine what contributions to performance might be domain-general versus domain-specific.

In the following work, we hope to begin to understand domain-general and domain-specific contributions to individuals' performance, with an emphasis on the common role of category-specific experience across tasks. We will measure both visual and semantic performance for eight different object categories using a newly designed test of semantic knowledge, the SVET. We will also measure domain-general visual and verbal abilities that contribute to visual and semantic performance. If we find that these general abilities are not correlated with each other, then we can hypothesize that any shared variance observed between visual and semantic performance for the same category may be the result of common category experience.

Background

Individual differences in visual performance. Individual differences have been almost completely overlooked in the study of general object recognition in neuro-typical populations. While some have measured object recognition ability in clinical populations (Barton, Hanif, & Ashraf, 2009; Duchaine & Nakayama, 2005; 2006; Germine, Cashdollar, Düzel, & Duchaine, 2011b), those studies generally seek to understand the extent of possible object recognition deficits rather than the variation in ability within a normal population. The one object domain for which individual differences have been extensively investigated is human faces. This is due in part to the development of a standardized test of face memory that has proven reliability and validity, the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006). It has been demonstrated that the CFMT can reveal fine differences in face recognition performance across the spectrum of performance from those with prosopagnosia and Asperger syndrome to so-called "super-recognizers" (Bowles et al., 2009; Germine, Duchaine, & Nakayama, 2011a; Hedley, Brewer, & Young, 2011; Russell, Duchaine, & Nakayama, 2009). This has allowed the CFMT to be used in work investigating the relationship between face recognition and other traits and abilities, such as holistic processing (Richler, Cheung, & Gauthier, 2011) and social anxiety (Davis et al., 2011). In a study of individual differences between pairs of homozygous and heterozygous twins (Wilmer et al., 2010), this measure, together with memory tests for art and words, suggested that face recognition ability is both highly heritable and specific (although see Gauthier et al., in press regarding specificity). While these studies inform our understanding of face processing, they measure performance with only one specific category, faces, with which most people have extensive experience. Because experience with faces is very high for most people, the CFMT may in fact offer a good estimate of domain-general visual ability, but it cannot inform us about the contribution of experience to performance.

A better way to tease apart domain-general ability from domain-specific experience in object recognition is to measure performance across a variety of different object categories. The Vanderbilt Expertise Test (VET) is an object memory test designed similarly to the CFMT that includes 8 different non-face object categories (birds, cars, planes, owls, wading birds, motorcycles, mushrooms and leaves). It allows measurement of both object recognition performance for specific categories as well as a better estimate of general object recognition tested across a range of categories (McGugin, et al., 2012b). Tests like the VET that compare performance across categories for which experience varies can help disentangle the domain-general and domain-specific contributions to individual differences in object recognition.

Measurement of category experience. In previous work, researchers have tried to quantify experience by asking subjects to report their category expertise on a numerical scale (1-9) as their level of “interest in, years exposure to, knowledge of, and familiarity with” objects from a specific category (Gauthier et al., in press; McGugin et al., 2012b). These self-report measures of object experience for specific categories individually have been shown not to be highly predictive of visual performance for that category (mean R^2=3.1% reported for eight categories by Gauthier et al., in press). However, an interesting result emerged when face and object recognition performance was considered in the context of self-reported object experience averaged across eight categories. Object experience did not predict face nor object recognition, but did influence the relationship between them: object recognition was more similar to face recognition with increasing object experience (Gauthier et al., in press). This is important because it suggests that face perception is not independent from object recognition, but that experience is a critical variable that differentiates visual performance for all domains, including faces. Because experience for faces is very high for most people, performance reflects mainly domain-general ability rather than experience; the same thing occurs for other object categories: increasing experience results in more similar performance between faces and other objects.

Self-reports are subjective and susceptible to bias. In particular, when subjects rate their experience with a category, which is a fairly general attribute, they may not have much information about how they rank relative to others. Self-reports might also differ in their relationship to performance based on how people commonly introspect about their abilities. For example, when asked to rate their expertise with an object category, subjects may consider how frequently they see an object, their interest in the category, or the extent of their vocabulary for the domain. The availability of each type of experience during introspection might vary and could be different for different domains. This might affect how self-reports are related to performance. For example, if domain-specific vocabulary is easier to introspect about than visual ability, then self-reports might be more related to semantic versus visual performance.

A recent metasynthesis (a qualitative interpretation of the results of many related studies) investigated the relationship between self-reports and objective measures of performance and found that on average there is a moderate correlation (mean r=0.29), with most abilities ranging between 0.2 and 0.4 (Zell & Krizan, 2014). However, the correspondence between perceived ability and performance was greater for specific abilities (e.g., hitting a baseball) than general abilities (e.g., playing baseball) and for simpler and more familiar abilities. Interestingly, they found that whether self-report occurred before or after performance had almost no effect, suggesting either no interaction between self-report and performance or an equal effect in either direction (e.g., saying you are good at something first makes you perform well, but also performing well makes you say you are skilled) (Zell & Krizan, 2014).

Measuring semantic knowledge. Performance on a measure of semantic knowledge also reflects experience for a given category. In a test of semantic knowledge, we might assume that performance is a reflection of both experience and domain-general ability, such as general verbal ability or fluid intelligence. There is currently no well-tested, non-visual measure of semantic knowledge available for a variety of everyday object categories.

One study that did employ a test of semantic knowledge used this measurement in a compelling way to estimate the level of visual performance that should match a specific level of semantic performance (Barton et al., 2009). This study aimed to quantify the true extent of visual object recognition deficits in a group of patients with prosopagnosia. Because of the wide variation in levels of experience with specific object categories, it can be difficult to dissociate the contributions of experience and ability to observed performance. In this case, it was difficult to determine if a particular patient's performance recognizing cars was reduced or not from his/her pre-morbid performance. Barton et al. (2009) developed a test of verbal semantic knowledge for cars in which subjects were given a list of the names of car models made between 1950 and 2005 and asked to write the manufacturer of each model, and a test of visual semantic car knowledge in which they saw an image of a car and were asked to name its manufacturer, model and decade of make. Then, using a group of healthy control subjects, they gathered data and found the relationship between performance on the verbal and visual tasks to be highly correlated. Based on the relationship found in controls, the verbal semantic knowledge performance of the individual prosopagnosics could be used to predict what their normal visual performance should have been, and determine if a patient's current performance demonstrated a deficit below expected performance. This comparison demonstrated that the prosospagnosics had significantly worse visual car recognition performance than would be expected based on their verbal semantic scores, an effect that was particularly strong for three of the prosopagnosics (one who self-reported high car experience) who had high semantic scores.

The tests of semantic knowledge developed by Barton et al. (2009) demonstrate the predictive power of semantic knowledge on visual recognition performance and the importance of accounting for experience in measures of object recognition. However, the weakness of these tests for measuring the relationship between visual and semantic performance is that both tasks require semantic knowledge of cars, which makes it perhaps unsurprising that tests are highly correlated, especially for naming cars at the most specific level (models). The correlation between measures in this case could be the result of shared verbal abilities as well as experience. Thus, for the goals of this dissertation, it is important that visual tests not require semantic knowledge or object naming and that tests of semantic knowledge not include images of objects so that the abilities tapped by each test are as distinct as possible.

Semantic knowledge has also been measured in studies of perceptual expertise, where categorization and naming at the subordinate-level as quickly and frequently as the basic-level has come to be a regarded as a hallmark of expertise (Tanaka & Taylor, 1991). Rosch et al. (1976) first demonstrated that subjects were fastest to categorize objects at the basic level (e.g., dog) compared to either the subordinate (e.g., golden retriever) or superordinate level (e.g., animal). However, Tanaka and Taylor (1991) found that the level of categorization for an object interacted with an individuals' amount of experience recognizing and naming objects from that category. In this study, a group of bird and dog experts were asked to complete several tasks with both their category of expertise (birds or dogs) and the other category, with which they did not have extensive experience. For each category, subjects were asked to list the features of basic and subordinate-level categories, freely name object images, and perform a category-verification task at both the basic and subordinate levels. Tanaka & Taylor (1991) found that for their domain of expertise, experts listed as many features for the subordinate- as the basic-level category, named objects at the subordinate-level as frequently as the basic-level, and verified subordinate-level labels as quickly as basic-level labels, demonstrating that semantic knowledge of object names and features was a result of extensive object experience. Having the vocabulary to describe object features and name objects at a specific level has even been shown to predict perceptual taste discrimination of specific wines (Hughson & Boakes, 2002).

Another example of non-visual semantic knowledge measurement are two tests that were created to measure print exposure: the Author Recognition Test (ART) and the Magazine Recognition Test (MRT) (Stanovich & Cunningham, 1992). These tests were created to provide a valid measure of print exposure that was less subject to bias than self-reports or activity diaries and to measure small individual differences in print exposure between literate subjects (rather than the large difference between illiterate and literate subjects). In each test, the names of real authors, mostly of novels, and magazine titles are presented in a list among foils and subjects must select as many targets as they recognize while avoiding foils. The SVET we created to pair with visual tests uses a similar format of recognizing target names among foils, but we will employ a triplet (one target, two foils) trial structure to reduce guessing and allow for more precise design of easy and difficult trials based on foil selection.

Beyond test creation, the study of print exposure by Stanovich and Cunningham (1992) is also an example how a theoretical framework can be used to relate performance and experience to make inferences about why people differ. The ART and the MRT were used with general measures of fluid intelligence and reading comprehension to demonstrate that print exposure predicts an array of verbal skills (spelling, vocabulary, verbal fluency) independent of fluid intelligence and reading comprehension. A similar framework has also been applied to understanding individual differences in knowledge of specific domains, for example knowledge of basketball (Hambrick, 2003) and current events (Hambrick, Meinz, & Oswald, 2007; Hambrick, Pink, Meinz, Pettibone, & Oswald, 2008). To understand why some individuals acquire different levels of knowledge, an array of domain-general ability and domain-specific interest and knowledge was tested, including general cognitive ability factors such as working memory and fluid intelligence, as well as personality traits and personal interest and previous knowledge in the domain. These studies suggest that acquisition of knowledge in a domain depends on independent contributions from both ability and non-ability factors. In this work, we will adapt this framework to ask how ability and experience contribute to both visual and semantic performance for the same category and across categories.

The need for a non-visual measure of semantic knowledge

The measure of verbal semantic car knowledge used by Barton et al. (Barton et al., 2009) correlated well with visual performance, however, that correlation is difficult to interpret because both tasks required knowledge of object names. Their verbal semantic task was also designed in a way that was specific to cars, and would be difficult to adapt for other object categories. Tanaka and Taylor's (1991) measures of feature listing and object naming were excellent for experts and for differentiating between basic- and subordinate-level category names. However, it is not adaptable to measuring object knowledge within a category for novices without enough knowledge to list any features, may be susceptible to expertise illusions (e.g., novices may be reluctant to list features in categories they do not believe they know much about), and may be relatively difficult to standardize.

For these reasons, in this dissertation we will develop a new test of semantic object knowledge, the SVET, which will be more versatile and standardized across a variety of object categories. There are several attributes that we will strive for in designing this new measure: it should measure semantic knowledge in a manner that is theoretically independent of visual knowledge, it should be easily implemented for a wide range of object categories, and it should have a fixed number of trials that will produce a standardized score, both for each category separately and across categories. The SVET will use knowledge of subordinate-level object labels and names as a measure of semantic knowledge for each category under the assumption that these are an important part of semantic networks. For example, an image of a specific cat might bring to mind the words “tabby,” “domestic short hair,” or “Teddy” (if we know the specific cat). Certainly the semantic network for many categories might include considerably more than names, however, using subordinate-level names will allow us to create tests that are comparable across many categories because names and labels are almost always an important part of semantic knowledge.

One could use the SVET as an independent measure of semantic knowledge either on its own or together with any variety of cognitive or perceptual measures or individual variables. We suggest that the SVET could also be used with another measure of performance for the same category, such as a visual task, to estimate the contribution of category-specific experience to both tasks. To estimate experience from measures of visual and semantic performance, we must first consider how these variables interact with each other and underlying abilities. Figure 1 presents a schematic model of the proposed relationships between these observed and latent variables. It is important to note that the performance measures are domain-specific, just as experience is domain-specific, but that the abilities are postulated to be domain-general. Every person has experience with all sorts of objects through interactions with the world in their jobs, hobbies, and everyday life; an individual may care about some objects a lot, and not care about others, and this is reflected in each individual's level of experience for a particular category.

We will consider two domain-general abilities that may be important for the tasks we use in this project. One is fluid intelligence (*Gf*), which can be thought of as general cognitive aptitude based on measures of problem-solving and reasoning. While crystallized intelligence refers to knowledge of specific facts and content, research and modeling have suggested that fluid intelligence is predictive of a general interest in learning and seeking knowledge, which results in acquiring crystallized intelligence in specific domains (Schmidt, 2014). Because fluid intelligence measures are mostly independent of prior experience, we hypothesize they may best capture individual differences that would influence a person's ability to acquire semantic knowledge in any domain.

The other general ability we will consider is a domain-general ability for visual learning (v), for example the ability to discriminate visually similar objects. The variance that is common to performance on many domain-specific tests of visual recognition might reflect v, although this is complicated by variability in domain-specific performance. Under the assumption that v is domain-general and includes for face recognition, measures of face recognition performance (such as the Cambridge Face Memory Test, CFMT; Duchaine and Nakayama, 2006) might provide a relatively direct estimate of v. This is because there is very little variance in amount of experience with faces between individuals (experience is high for everyone), so the standardized score on the CFMT is likely to be a very good estimate of an individual's true ability to recognize faces, which might reflect *v*. An estimate of *v* might also be obtained by looking at average VET performance across all categories (or all but one) because in principle aggregating across categories reduces any category-specific effects on performance. Importantly, studies of performance on the CFMT and measures of intelligence suggest that *v* and *Gf* are not correlated (Davis et al., 2011; Hedley et al., 2011; Wilhelm et al., 2010; Wilmer et al., 2010).

Domain-specific experience must interact with each of these domain-general abilities to produce domain-specific performance. If individuals' performance on visual and semantic tasks is correlated, and if *v* and *g* are independent of one another, the shared variance in performance may be the result of common category experience. Therefore, according to this proposed framework experience can contribute to at least part of the shared variance between visual and semantic performance.

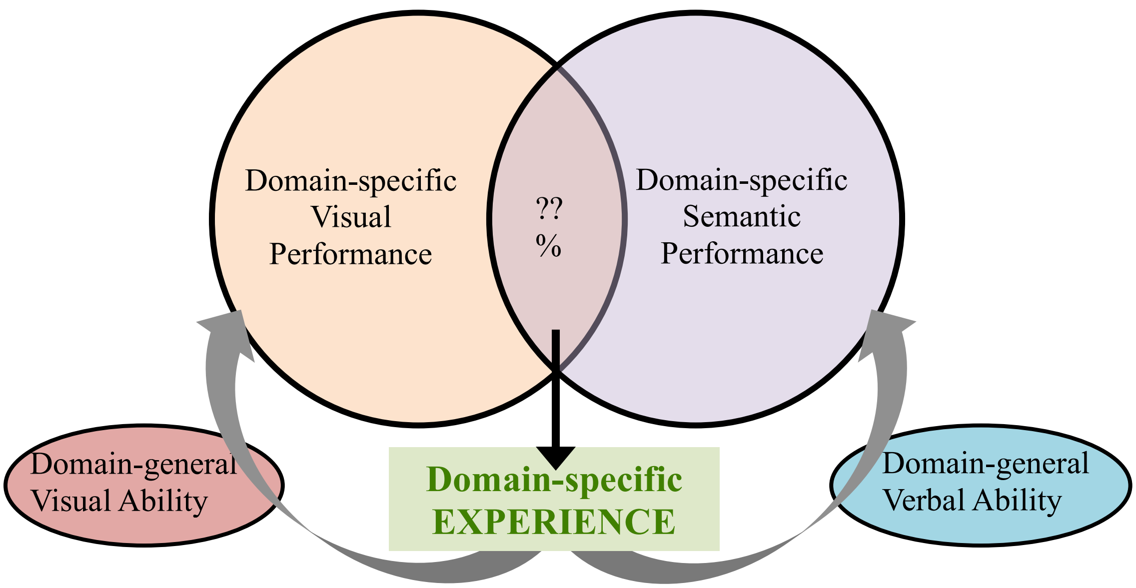


Figure 1. Diagram of proposed relationships between variables that contribute to domain specific performance. Visual ability and verbal ability are domain-general abilities that do not interact with one another, but each interact with domain-specific experience separately. The interaction of ability and experience shapes domain-specific performance. The shared variance between visual and semantic performance reflects common domain experience.

Goals of the dissertation

The goals of this work are first to create a standardized, non-visual measure of semantic knowledge for object categories that is reliable by testing and refining the test with a large online sample (Study 1). In Study 2A, we will use the newly developed SVET together with measures of visual performance for each category and domain-general abilities in a large sample in the laboratory to understand the relationships between these measures and assess the contribution of category experience to performance across visual and semantic tasks. Studies 2B and 2C will continue our exploration of the SVET by whether name knowledge explains the visual-semantic relationship and testing the SVET with an expert population of birders. Lastly, in Study 3 we will demonstrate how the SVET might be a useful independent measure for understanding another cognitive phenomenon; we will explore the lateralization of visual object recognition performance as a function of semantic knowledge by asking if SVET performance predicts lateralized object recognition. Overall, the work will contribute to the new research area of individual differences in object recognition and provide a carefully tested and refined tool, the SVET, to the psychology community for future use in this field.

Chapter 2 – Study 1

Creating a Test of Semantic Object Knowledge: The Semantic Vanderbilt Expertise Test (SVET)

Overview

The goal of the first study was to create a valid and reliable measure of semantic knowledge for a variety of object categories, which will be called the Semantic Vanderbilt Expertise Test (SVET). We wanted to create a measure that could easily be used to assess an individual's level of semantic knowledge both for a specific category, and using a format that could be applied in the same way across many categories. There were several attributes we wanted the test to have that would make it flexible to use for many different categories and effective for easily testing a range of subject populations in combination with other measures. First, the test should have a concise, standardized format that can easily be used for many different object categories and that can be completed by subjects with both low and high semantic knowledge. Some methods of testing semantic knowledge such as listing names or features of objects in a category might favor high knowledge subjects but not discriminate low knowledge subjects who might be at floor; to avoid this, recognition of object names will be a better task than naming. Second, the test should not include any visual information so that it would measure performance relatively independently of visual skills. Ultimately we are seeking to pair the SVET with a visual test, such as the Vanderbilt Expertise Test (VET; McGugin et al., 2012b), to measure visual and semantic performance separately. Third, the test should be able to measure performance with as much precision as possible across the full range of ability, from novice to expert. Fourth, the test should provide a valid measure of domain-specific semantic knowledge. This can be tested by assessing whether semantic performance is correlated with self-reports of experience with that category and with visual performance for that category, more than the same measures for other categories. Finally, the test should have good internal consistency and be relatively unidimensional, such that differences in scores between individuals will reliably estimate differences in a single type of knowledge.

To create and refine the SVET for each category, and to determine which categories would be appropriate for measurement with the SVET, we recruited subjects online using Amazon Mechanical Turk (AMT) to complete single category SVETs. AMT allows us to collect data quickly and with minimal cost from a population that is more diverse in terms of age, experience, and education than we typically have access to on a university campus, a feature that is particularly important for creating these tests. We will also test reliability in a university population in Study 2 since that population is most likely to be used by other psychologists who use the SVET. In recent years AMT has been used to collect data for a wide range of psychology tasks including classic and complex cognitive and perceptual tasks (e.g. face recognition, Stroop, attentional blink) and studies have found that results obtained from AMT are consistent with those obtained in the laboratory when subjects are screened and carefully instructed online (on AMT: Crump, McDonnell, & Gureckis, 2013; or on other websites: Germine et al., 2012). The steps taken to refine each SVET based on online data will be outlined here, followed by the presentation and discussion of data collected on AMT with the final version of each SVET with 116 subjects who completed the SVET for all eight categories: cars, planes, Transformers, dinosaurs, shoes, birds, leaves, and mushrooms.

Test design

The SVET uses knowledge of subordinate-level names (Jolicoeur, Gluck, & Kosslyn, 1984; Rosch et al., 1976) as an indicator of how much knowledge an individual has about a particular object category. For many domains individuals may have more complex, relational, and hierarchical knowledge beyond naming, but the nature of this knowledge varies considerably between domains, (e.g. for dinosaurs: time periods, diets of dinosaurs, predator-prey relationships; for birds: migration patterns, habitat, songs, physical differences between males and females) whereas naming knowledge can be tested for nearly all domains. This measurement makes the assumption that the more a person knows about a domain, the more vocabulary he/she has acquired about that domain. We will validate this assumption by comparing SVET performance with self-reports of category experience to ensure a positive relationship.

Categories. We selected object categories for which subordinate-level names would be relevant and typically acquired with experience and interest in the category. Categories were also required to have subordinate-level objects that could be learned and recognized in the VET. We selected a set of categories that included domains of greater interest to men or women as previous work has found strong effects of sex on experience and visual performance with different categories (McGugin et al., 2012b), as well as a mix of living domains and artifacts. We created SVETs for 9 categories and pilot testing showed that only one of them would not be useful with a non-expert population; the SVET-Butterfly was excluded from further testing because of near chance performance in pilot testing. Our final set of categories in the SVET includes four male-interest categories—cars, planes, Transformers, and dinosaurs—and four female-interest categories—shoes, birds, leaves, and mushrooms. Previous work (McGugin et al., 2012b) demonstrated greater self-reports of experience and better recognition performance for men with non-living vehicles categories such as cars, planes, and motorcycles, and for women with living categories such as owls, wading birds, and leaves. Given the nature of these sex differences for object categories, men may have greater experience with more artifact than living categories, and women may have greater experience with more living than artifact categories.

Alternatively, this may be an accident of the categories used in McGugin et al., (2012b), because all the non-living categories in that study were vehicles (cars, planes and motorcycles). To achieve a better balance of living/artifact categories relative to the original VET (McGugin et al., 2012b), we included dinosaurs as a living category (or once-living and natural) that we believed men would have more experience with and shoes, specifically women's high heels, as an artifact that we believed women would have more experience with.

Trials. The SVET for each category consisted of 48 test trials and 3 additional catch trials. Each SVET trial presented three names: one target name, a real subordinate-level name of an object in that category (e.g. Honda Civic, 737, blue jay, birch), and two foil names, which are either names of objects in a category not tested in any of the SVETs (e.g., types of stone, grass, or viruses; never more than 3 per category were used in the same SVET) or were entirely made-up words. Compared to object naming, this trial format allows those with limited category knowledge to complete the test. Having three names in each trial reduces the chance level and allowed us to manipulate trial difficulty through foil properties, for example how similar a foil is to a real name. Catch trials were very easy trials that followed the same format as the test trials with real target names but much more obvious foil names (e.g. blue jay, JCPenney, lipstick). Including catch trials is particularly important for online testing to exclude any subjects who answer randomly or do not understand the task instructions. The trials and trial order were the same for all subjects to eliminate order effects as a contribution to individual differences. In the versions of the SVET used in the complete eight-category set (SVET 1.0) in the following studies, trials were ordered from easiest to hardest based on trial accuracy from previous pilot data. See Table 1 for examples of an easy, medium, and difficult SVET trial for each category (see Appendix A for all trials in each SVET).

Car names. Names in the SVET-Car consisted of make and model names (e.g., Honda Accord). All makes for both target and foil names were real car brands sold in the United States between 2000 and 2013 (e.g., Toyota, Ford, Audi). Target model names were real models of sedans sold in the United States between the years 2000 and 2012 (e.g., Camry, Focus, A6). Foil model names were either real words or non-words that (to the best of our knowledge) had not been used as 20th or 21st century American car names (e.g., Olympic, Alepo, Primo). Target names always consisted of both a real make and real model name combined to form a real car name (e.g., Honda Accord). Foil names were of two forms, with equal frequency: 1) Mismatched, a real make and a real model name but that do not form a real car name when combined (e.g., Honda Camry); and 2) Fake, a real make and a fake model name (e.g., Honda Napa).

Plane names. Target names in the SVET-Plane were names of model airplanes used in the United States in the past 20 years, with the exception of 8 trials that were plane models used in World War I or World War II. Although planes could be referred to by a manufacturer name, model name and sub-model name (e.g. Airbus A340-300), we aimed to use only the model name that a person with knowledge of planes would use (e.g. A340). Plane models were from several types of aircraft: commercial (16 trials), general aviation (7 trials), World War I (1 trial), World War II (7 trials), military aircraft including fighter, fighter trainer, transport, and bomber (12 trials), drones (3 trials), and business jets (2 trials). Foil names were created to match the format of real plane model names, such that some were all numbers, combinations of letters and numbers, or words.

Transformer names. Target Transformer names were the names of characters capable of changing form (not human characters or the names of other objects or locations) from the

Transformers entertainment franchise produced by Takara Tomy and Hasbro toy companies. Names were selected from *Generation 1, Generation 2*, and *Beast Wars* series of comics and television and from the recent film series (2007, 2009, 2011 movies). Foil names were created to match the style of real Transformer names and were words or non-words.

Dinosaur names. Target dinosaur names were the common names of discovered dinosaurs that are generally accepted by the scientific community. Names were taken with roughly equally sampling across time period (e.g., Jurassic, Cretaceous) and other dinosaur traits (e.g. herbivore vs. carnivore, bipedal vs. quadrupedal). Foil names were created to match the style of real dinosaur names, with some names based on physical attributes denoted by Greek roots (e.g. using roots tetra meaning four and cerato meaning horns) and others named after places or fictitious people who may have discovered them.

Shoe names. The VET-shoe and SVET-shoe both refer to knowledge of women's high-heeled pumps. Shoes are perhaps a particularly interesting category for name knowledge. Although individual shoes do have a model name on the box (e.g. Moxy, Delilah), these names change every season and would rarely be used to identify a shoe even by those who are very skilled at visual shoe recognition (see Study 2B). Instead, we used the brand (or designer) names of women's high-heeled pumps as the target names. We hypothesized that these were the names that one would acquire knowledge of as they become more experienced with shoes. All brand names are brands of women's high-heeled pumps currently sold in the United States at Nordstrom or Saks Fifth Avenue department stores. Foil names were created to match the style of real brand names (e.g. one or two words or the name of a designer) and were words or non-words.

Bird names. Target names for the SVET-bird were all common names of passerine, or perching, birds found in a large portion of North America. Real bird names were selected to sample across a variety of passerine families (e.g. flycatchers, orioles, jays, finches) and east and west coast birds. Foil names were created to match the style of real bird names and were words or non-words.

Leaf names. Target names for the SVET-leaf were all common names of deciduous trees found in a large portion of North America. Foil names were created to match the style of real leaf (tree) names and were words or non-words.

Mushroom names. Target mushroom names were all common names of mushroom species found in North America. An effort was made to avoid using multiple names that refer to the same species. Most of the mushrooms used are edible, although some (6 trials) are poisonous or potentially poisonous.

Table 1. Example trials from each SVET. Selected trials illustrate an easy, medium, and difficult trial for each category (lower trial numbers are easier). Names in orange are the real name and correct response.

	Name 1	Name 2	Name 3
CAR			
Trial 6	Volvo Focus	Mercedes-Benz C300	**Mercury Alero**
Trial 26	Suzuki Prestige	**Infiniti G37**	Pontiac S550
Trial 49	Saturn Fusion	**Acura TSX**	Saab S80
PLANE			
Trial 2	**737**	Serpens	Sheffield
Trial 24	8900	A2 Lobo	**Spitfire**
Trial 45	**Mosquito**	Western Lair	A480
TRANSFORMER			
Trial 4	Lavaman	Chromoburn	**Quickstrike**
Trial 28	**Sunstreaker**	Septawave	Proton
Trial 46	Waveracer	**Hound**	Sotter
DINOSAUR			
Trial 7	**Pentaceratops**	Eudontidectes	Microtarius
Trial 22	Stuthioceratops	Centaurisaurus	**Iguanodon**
Trial 48	Corposaurus	Monocyclosaurus	**Mussaurus**
SHOE			
Trial 4	**Nine West**	Rebecca Fox	Aloft
Trial 25	Zetta	Kalden White	**Franco Sarto**
Trial 47	Graham Wood	Gravelle	**Chinese Laundry**
BIRD			
Trial 5	**Savannah Sparrow**	Tufted Gemthroat	Green Huckaloo
Trial 23	**Scarlet Tanager**	Blue-stripe Binbeak	Tri-colored Wheatear
Trial 37	Spot-breasted Pixie	**McCown's Longspur**	Pale-eyed Baylin
LEAF			
Trial 3	Red Mountainwood	Venuswood	**American Sycamore**
Trial 19	**Yellow Poplar**	California Bargo	Feather Willow
Trial 47	Silver Aster	Valley Walnut	**Tulip Poplar**
MUSHROOM			
Trial 4	**White Truffle**	Milky Scaber	Sugar Siullus
Trial 25	Amber Stalk	Tavel	**Enoki**
Trial 46	**Crab Brittlegill**	Elephant Trunk	Glass Cap

Trial presentation. On each trial, three object names (one real name and two foils) were presented mid-height in the browser window, one in the center and one each to the left and right of center, on a white background (see Figure 2). The location of the correct response in each trial was counter-balanced so that it occurred with equal frequency in the left, right, and center locations. Subjects responded by clicking on a name and were given unlimited time to make their response on each trial. No feedback was provided and subjects could not return to previous trials.

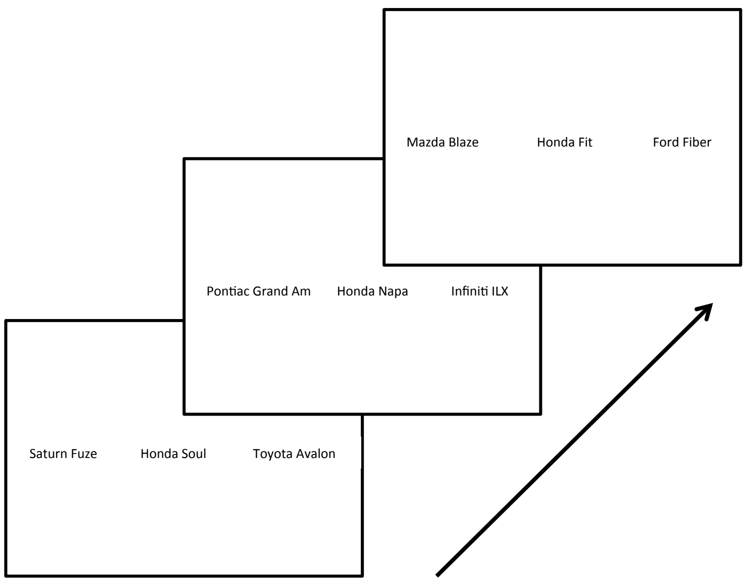


Figure 2. SVET trial format showing three trials from SVET-Car.

Test refinement

After designing an initial SVET for each category, we went through multiple iterations of data collection and revision for each test (between 1 and 4 revisions per category). Data were collected independently for each SVET with subjects recruited from AMT (https://www.mturk.com/, Amazon Web Services, Amazon.com Inc., Seattle, WA) with restrictions that they be English speakers residing in the United States (N=35-101 per SVET version; N=1,383 total).

For each SVET version, we went through several steps to assess and revise the trials. First, we looked at accuracy on test trials to evaluate if the frequency distribution of trial accuracy was spread evenly across difficulty. Ideally the tests would have an even distribution with some trials demonstrating high accuracy, almost all subjects got them correct, and others demonstrating medium or low accuracy suggesting that more category knowledge was required to answer the trial correctly. In addition to accuracy for each trial, we looked at the percentage of times each foil was chosen on a given trial. A foil that is rarely chosen will not be useful in differentiating knowledge levels and will in effect increase the chance level to .50 if subjects without any knowledge almost always know to ignore a certain implausible foil.

For each test we looked at the correlation between SVET performance and subjects' self-report of experience with that category. We found that these were always well-correlated, although to a greater extent for certain tests than others, which is likely due to the nature of that category and how subjects answered our "experience" question. More importantly, we expect when we test the same subjects with many categories that correlation of performance and experience will be greater for the same category than across categories (this will be reported for the last versions of each SVET).

For each SVET dataset, we computed Cronbach's alpha as a measure of internal consistency to assess test reliability. Our goal was at least acceptable test reliability ($\alpha > .7$), although given our relatively small sample size for each version we expected that some variability would be due to sampling noise. We also conducted factor analyses using polychoric correlations for each test to estimate how similar the trials were to one another, i.e., if they were measuring the same type of knowledge. Using the results of these factor analyses, we could look at the eigenvalues for each factor as well as the factor loadings for each trial. Our goal was to make each test as unidimensional as possible so that test scores would reflect differences in primarily just one type of knowledge across individuals. With a similar goal in mind for test revision, we also conducted some more exploratory analyses using 2-parameter Item Response Theory (IRT) models. We used IRT results to assess trial difficulty in the context of subject ability (theta), which we wanted to have an even distribution, and the discriminability of trials (slope), which we wanted to be positive and as high as possible to increase the usefulness of our measure in differentiating between individuals.

Using all of these data, we revised the tests to increase reliability and coverage over the range of ability in the normal population. We replaced target and foil names in some trials to smooth the accuracy distribution so that there was an even number of trials from easy to difficult and so that foils in the same trial were chosen with equal frequency for incorrect responses. Trials that had low Factor 1 loading or high Factor 2 loading were changed to reduce multidimensionality and trials with a negative slope in our IRT model were also revised. In a typical test revision, between 5 and 20 target or foil names were changed, but an entire trial was rarely dropped. Although it was sometimes possible to replace a more common name with a more obscure one or vice versa, the decision of how to replace names was often limited to our introspection on how someone would approach a trial. This was especially difficult for understanding why certain foils were more or less appealing, and for cars and planes we sought out experts with that category to get their feedback on the trials. A more data-driven approach was employed for several categories, in which we refined not just the most recent version (e.g. change version 2 trials to create version 3), but also looked at data from multiple test versions (e.g. version 1 its derivation version 2) and combined trials from the two tests to create a version a well-rounded test.

Testing all SVETs in a single online sample

After revising each SVET separately, we conducted an online study in which all subjects completed the final SVET for all eight categories. This dataset allowed us to compare performance between categories in the same subject and obtain data for the 1.0 version of each SVET with an online sample, which differs in age, education, and category experience from the Vanderbilt samples we will test in Studies 2 and 3.

Subjects. Subjects were recruited online on AMT. The SVET for each of the eight categories was completed as a separate HIT or task. The SVET-Car was posted first and subjects who completed the test without missing more than one catch trial and who obtained above chance accuracy (all but two) were sent a personal invitation offering them the chance to complete the other seven SVETs. The study was approved by the Vanderbilt IRB and subjects gave informed consent before the start of each test. Subjects were paid $0.75 for the SVET-Car and $0.10 for each of the other tests, but were awarded a $1.00 bonus for each of the seven additional tests if they completed the full set within 24 hours ($8.45 total). All eight SVETs were completed by 116 subjects (48 male) aged 18-67 (mean=35.82, SD=12.45) who are reported in the analyses. All but one of these subjects reported English as their native language, but all reported being English-speakers currently residing in the United States. Subjects who completed only a subset of the SVETs (N=9) or completed only the SVET-car but did not accept the invitation to complete more tests (N=22) were compensated without bonuses and are excluded from the analyses.

Procedure. At the beginning of each SVET, we asked subjects to provide a rating of their experience with the object category (as in Gauthier et al., in press; McGugin et al., 2012b). For example the self-report for cars asked, “Rate your expertise with: *Cars*. By expertise we mean your experience with, interest in, and knowledge about items in this category, relative to other people." Ratings were on a whole-number scale from 1 (very much below average) to 9 (very much above average). Task instructions to select the *real* name were adjusted for each category to be as clear as possible (e.g. for birds, "the real, common name of a bird species found in North America"). Each SVET trial was then presented, as shown in Figure 2, until subjects responded by clicking on one of the three names. There was a 1 sec inter-trial blank interval. Subjects completed 51 trials total for each SVET (48 test trials and 3 catch trials that were interspersed with the test trials); trial order was the same for all subjects with trials ordered approximately from easiest to hardest in each test. All subjects completed the SVET-car first, but could complete the other seven categories in any order they chose.

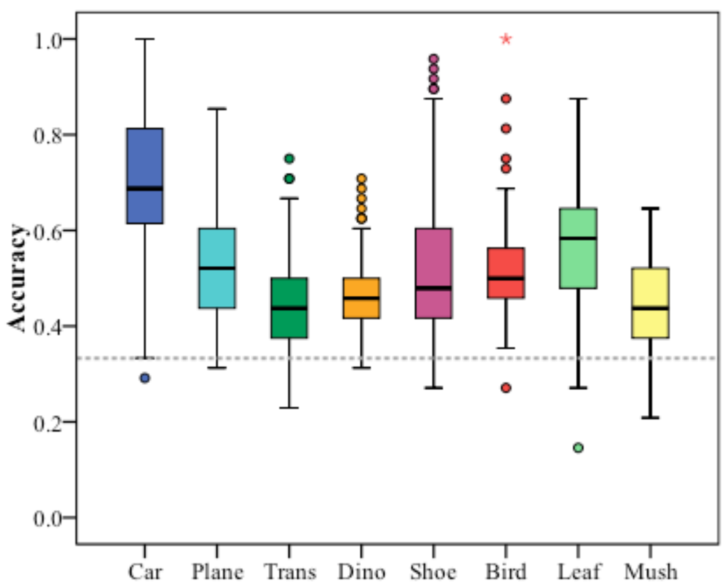


Figure 3. Boxplots of accuracy for SVETs completed by a single set of subjects on AMT (N=116). The bottom and the top of each box represent the first and third quartile, respectively, and the middle line of each box shows the second quartile, or the median. Whiskers show the highest and lowest scores between 1.5 and 3 (the interquartile range) while outliers beyond this range are represented by dots (or an asterisk in the case of one very extreme outlier). The grey dotted line shows chance (.33).

Results and discussion. No subjects were excluded due to catch trial performance (catch trial accuracy across all tests >.66), which was very good on all tests (.93-.99 on each test). As can be seen in Figure 3 and the first column of Table 2, mean SVET performance and variability differed between categories, with cars and mushrooms showing the highest and lowest accuracy, respectively. The greatest variability in accuracy was observed for cars and shoes, for which self-report of experience is the highest (Table 2, column 2), suggesting that greater amounts of experience with cars and shoes is reflected in the SVET score. For all categories, we observed a significant correlation between SVET performance and self-report of experience with that category (Table 2, column 3). To determine if this relationship between SVET performance and experience was greater within a category than across categories we calculated the average correlation between each SVET and self-reports of experience for the other seven categories (r=-0.05-0.10 for each category). We found that SVET performance and experience were always much more highly correlated within category (mean r=0.37) than across categories (mean r=0.05). In a further step we asked how each SVET predicts experience-Other (the average of experience ratings for the other seven categories) (mean r=0.08; Table 3, column 2) and also how category experience predicts SVET-Other (the average of SVET performance for the other seven categories) (mean r=0.07; Table 3, column 3). We found that within category SVET-experience correlation was already greater than both of these relationships with non-category SVET experience. This suggests that the SVET effectively measures category-specific experience with a category using non-visual naming knowledge. There may also be strong effects of sex or other variables underlying the differences between categories, both in amount of experience and, perhaps as a result, SVET performance. However, as the goal in Study 1 is to create and validate the SVETs we will hold off on exploring these effects in more detail in Study

2, where we will have a larger sample, a more nuanced measure of experience, and measures of both visual and semantic performance.

Table 2. Results of the SVET 1.0 for each category from a single group of subjects on AMT in Study 1. All correlations with experience (column 3) are significant ($r_{Crit}(114)=.18$, $p<.05$).

	Mean Acc (Std Dev)	**Experience (Std Dev)**	**Correlation with Exp, *r***	**Cronbach's Alpha**	**Factor 1 Eigenvalue**	**Factor 2 Eigenvalue**	**Factor 3 Eigenvalue**
Car	0.69 (0.15)	4.78(1.37)	0.37	0.94	14.88	6.84	3.95
Plane	0.53 (0.12)	2.78(1.38)	0.36	0.84	8.61	5.70	4.86
Transformer	0.45 (0.11)	2.84(1.65)	0.50	0.75	7.89	6.36	4.04
Dinosaur	0.47 (0.08)	3.90(1.48)	0.40	0.63	7.08	5.89	4.47
Shoe	0.53 (0.17)	4.07(1.66)	0.44	0.92	12.17	5.95	3.80
Bird	0.52 (0.10)	3.41(1.40)	0.33	0.75	6.93	4.97	4.58
Leaf	0.57 (0.13)	3.57(1.53)	0.27	0.87	10.32	4.68	3.61
Mushroom	0.44 (0.10)	2.63(1.40)	0.27	0.71	7.17	5.86	3.72

Table 3. Correlations (*r*) of SVET and experience self-report ratings within and between categories for a single group of subjects on AMT in Study 1. Column 1 shows the correlation between SVET and experience for the same category (column 3 in Table 2). Column 2 shows the correlation between SVET for the category and experience-other (average of self-reports on other 7 categories). Column 3 shows the correlation between experience for the category and SVET-Other (average of SVET performance on other 7 categories). Values in bolded red are statistically significant ($r_{Crit}(114)=.18$, $p<.05$).

	SVET-Category and Experience-Category	**SVET-Category and Experience-Other**	**Experience-Category and SVET-Other**
Car	**0.37**	0.06	0.13
Plane	**0.36**	0.15	0.06
Transformer	**0.50**	0.06	0.09
Dinosaur	**0.40**	**0.21**	**0.19**
Shoe	**0.44**	-0.09	**0.24**
Bird	**0.33**	0.14	0.04
Leaf	**0.27**	0.01	-0.15
Mushroom	**0.27**	0.08	-0.03

To assess the SVET as a useful measure of semantic experience, we need to look at several properties of each test in this dataset. First, we consider the range and frequency of trial difficulty. The frequency histograms of trial accuracy in Figure 4 illustrate that while some categories demonstrate a better spread than others, for example SVET-Car is skewed towards easy (or high accuracy) trials likely because it is the category with the greatest level of experience, all of the tests cover nearly the full range of difficulty without over-representing any difficulty level too much. Some of the spikes in frequency are likely due to noise in our sample, and while categories such as cars, shoes and leaves could be smoothed further in their difficulty distribution, we feel comfortable using these versions of the SVET going forward since no area of the difficulty distribution above chance (.33) is unrepresented.

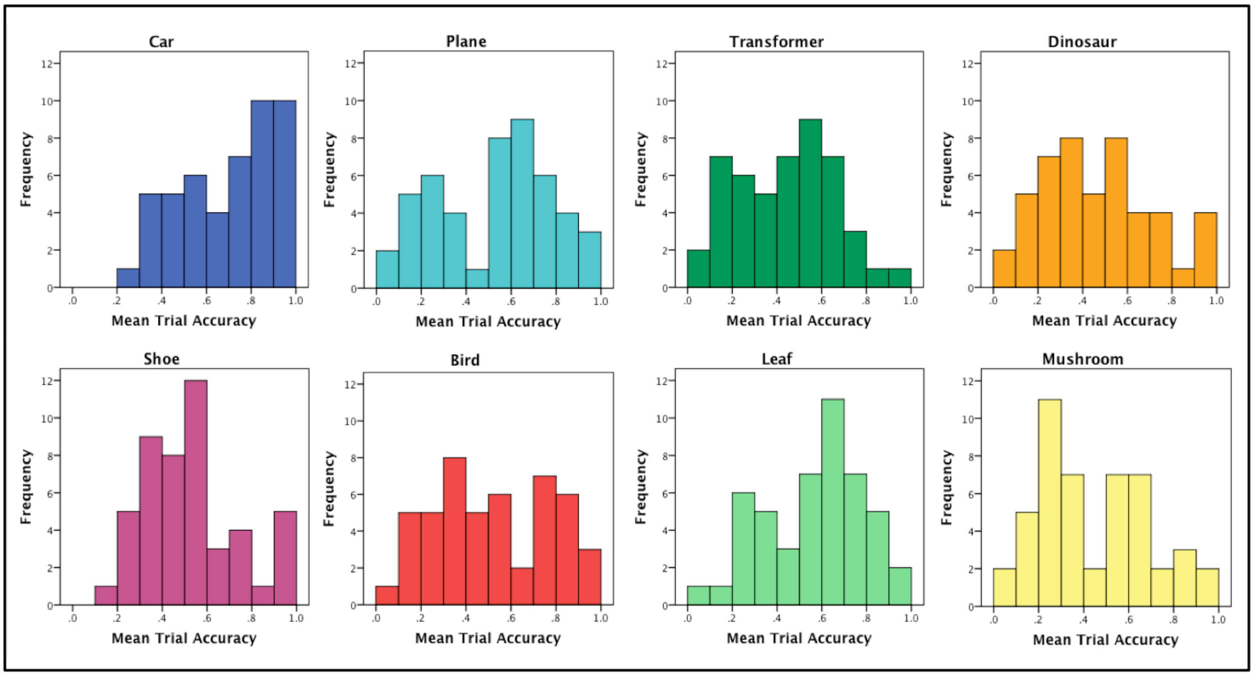


Figure 4. Frequency histograms of trial accuracy for each SVET as completed by a single group of subjects on AMT (N=116).

Next, we assessed the reliability and dimensionality of the SVETs. We computed Cronbach's alpha, a coefficient of internal consistency, to estimate the reliability of each test, which is a way of asking if all of the trials in a SVET are measuring the same thing, in this case semantic object knowledge. In Study 1, all tests had at least an acceptable level of internal consistency (α>.6) and many demonstrated higher consistency (α>.75) (Table 2, column 4). We can understand these differences in internal consistency by looking at the principal factor analysis for each test. Table 2 (columns 5-7) shows the eigenvalues for the first three factors for each of the SVETs, which represent the amount of variance in the test that is accounted for by each factor. Factor eigenvalues are traditionally plotted on a scree plot to appreciate how factor loadings drop off significantly after the first or several factors. Figure 5 shows example scree plots for a subset of the SVETs. As can be appreciated from the figure, the SVET-Car and SVET-Shoe are highly unidimensional with the first factor carrying a large portion of the test variance, while other SVETS, including plane and bird, are more multidimensional. Looking at the factor 1 and 2 eigenvalues in Table 2, it can be appreciated that categories with higher factor 2 values relative to factor 1 values, for example SVET-Dinosaur, also have lower internal consistency, suggesting that those SVETs may measure more than a single type of knowledge. Overall, our tests demonstrate good reliability, with some being more multi-dimensional than others. For categories that demonstrate multidimensionality, it may be useful in some analyses to restrict analyses to a subset of trials that load highly on a single factor, a strategy we will use in Study 2.

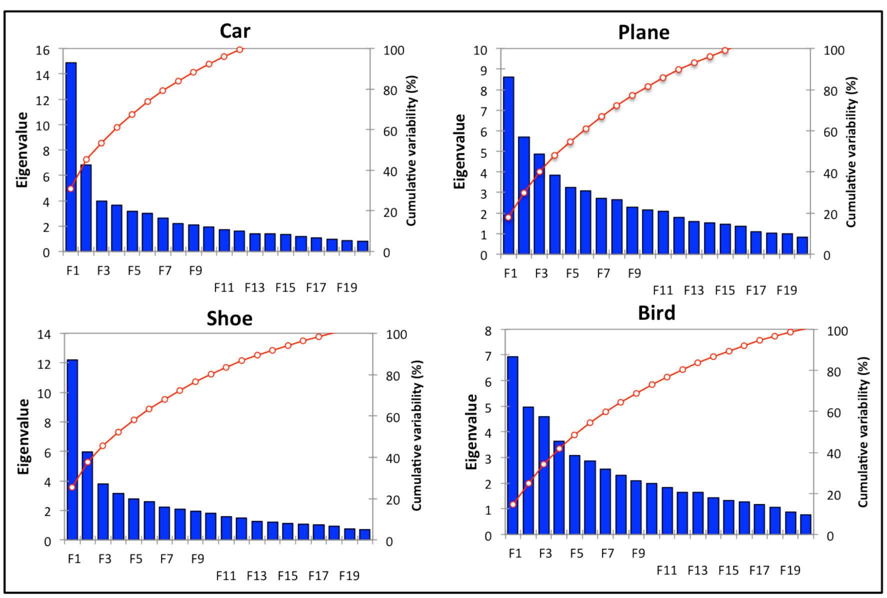


Figure 5. Sample scree plots for four of the final SVETs from data with a single group of subjects on AMT. The blue bars and the y-axis on the left show the eigenvalues for each of the first 20 factors, which measure test variance accounted for by that factor. The red curve and the y-axis on the right show cumulative variability accounted for by Factors 1-n.

Conclusions

In Study 1 we applied a multi-stage process of creating, testing, and refining the SVET for a final set of eight object categories. We successfully created a non-visual test of semantic object knowledge using subordinate-level names in a standard format that can be used for many different object categories and can be completed by both novices and experts. We demonstrated that each SVET is reliable, valid, and covers a range of performance.

We found that all eight SVETs show acceptable or good internal consistency. We also investigated the dimensionality of each test because unidimensionality is an assumption required for some analyses including basic IRT methods. Some tests were highly unidimensional while others were more multidimensional, with two factors carrying much of the test's variance, perhaps reflecting the knowledge structure for that category.

Each SVET consists of trials ranging from easy to difficult so that it can measure individual performance with a high level of precision across the full range of knowledge. We tested the frequency distribution of trial accuracy for each SVET and found excellent coverage across the range of performance.

We also validated the SVET by considering the correlation between SVET performance and self-reports of category experience. We found that each SVET was well-correlated with experience for that category, and that within category experience correlated much more than across category experience.

We tested over 1,500 subjects and addressed many aspects of test design usually overlooked in cognitive psychology. In Study 1, we sought to design the SVET as carefully as possible to be a reliable and valid measure of semantic knowledge that is standardized, informative, and reasonable in length. The attention and time spent rigorously testing our measures is critical for tests of individual differences in object knowledge and cannot be overlooked. The SVET will be available to psychologists as a tool for measuring object experience and understanding visual and semantic object performance.

In Study 2 we will use the SVET together with a visual memory test (the VET) for each category, as well as domain-general measures of visual and verbal ability to investigate visual object performance and the contribution of experience. This demanding individual differences approach to studying object recognition would not be possible without the careful design of all of these measures.

Chapter 3 – Study 2

Exploring the SVET and Understanding the Contribution of Experience to Performance

After establishing the basic psychometric properties of the SVET for eight object categories in Study 1, the next goal is to investigate the relationship between visual (VET) and semantic performance (SVET) for the same categories. In this study, all subjects will complete the VET and the SVET for eight object categories, allowing us to compare performance on both measures both within and across categories. In addition, adding measures of domain-general visual and verbal ability will help to determine how performance is influenced by domain-general versus domain-specific abilities. Using these measures together to understand the contributions to performance will allow us to understand how category experience influences an individual's visual performance. We will also consider the SVET in two special cases. We will measure visual and semantic performance for a category for which even experts rarely employ subordinate-level names to assess the extent to which visual and semantic performance may be related not because of common experience but because semantic information is automatically employed even in visual tasks where it is not required. We will also test the SVET-Bird and VET-Bird in a sample of expert birders to test the validity of the SVET for use in expert populations.

Study 2A:

Using the SVET and VET with Measures of General Visual and Verbal Ability to Understand Performance and the Contribution of Experience

Overview

Study 2A addresses the SVET's relationship with a number of other individual abilities and variables. Here we explore the relationship between visual and semantic performance for each of the eight categories for which we developed a SVET in Study 1. We ask how visual and semantic performance measures are related both within a category and across categories and how they are related to an individual's age, sex, and self-report of experience. Using domain-general measures of visual learning and fluid intelligence (*Gf*) we will also ask if underlying domain-general abilities contribute to domain-specific performance. With this set of measures, we hope to account for many of the variables that contribute to visual and semantic performance for a given category, which may then allow us to uncover whether experience with a given category is the main contributor of the domain-specific relationship between visual and semantic performance.

We address these questions with a dataset collected in the laboratory with subjects from the Vanderbilt and Nashville community, thereby also obtaining reliability measurements in a different population than Study 1 (younger on average, tested in the laboratory). We employed a progression of correlations, rotated factor analyses, Item Response Theory models, and regressions to understand the contribution of category experience to performance.

An important goal of Study 2A is to investigate the relationships between domain-specific performance, experience, and domain-general abilities. By collecting a large dataset in which every subject has completed all measures, including both the VET (visual) and the SVET (semantic) for eight object categories, we can ask a number of questions about the relationship between these variables. First, we will assess the relationship between both VET and SVET performance with an expanded self-report measure of category experience. The correlation between performance and experience within a category, more than across categories, will provide further evidence that these tests are valid in their measurement of category-specific knowledge. We will then consider the relationship between VET and SVET performance for each category with age, sex, domain-general visual learning, and *Gf*. The relationship between performance with each of these variables is interesting on its own, but it is also important to account for so that we can later consider the relationship between domain-specific visual and semantic performance independent of these variables. We will address the relationship between VET and SVET performance within compared to between categories and try to understand what variables contribute to shared variance between VET and SVET. Our hypothesis is that common VET-SVET variance reflects shared category experience. To estimate the contribution of experience, we will remove this contribution from other measured variables that are not tied to category-specific experience (age, sex, *Gf*, domain-general task performance assessed using aggregates of the other domains). The results will help us understand how category experience is expressed in visual and semantic performance.

Experiment design

Five measures were used in Study 2A, three domain-specific measures including a self-report of experience, the VET (McGugin et al., 2012b), and the SVET, and two domain-general tests of visual learning and fluid intelligence (*Gf*). Each domain-specific measure tested eight object categories in the following order: cars, birds, dinosaurs, shoes, planes, mushrooms, Transformers, and leaves.

Face recognition performance was measured using the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006). The CFMT is a well-established measure of visual memory for an object category outside of our SVET set, human faces. As suggested before, visual performance for faces is of particular interest because experience is high and variability is low, such that CFMT may be a relatively good measure of visual ability.

In this study we used three different tasks to measure fluid intelligence, as each task measures reasoning and problem-solving abilities in different domains that all contribute to a measure of *Gf* (adapted from Redick et al., 2012; see also Hambrick et al., 2007; 2008). All three tasks ask subjects to find a pattern or rule in a set of stimuli. Raven's Advanced Progressive Matrices (RAPM) (Raven, Raven, & Court, 1998) is a test of spatial ability, The Letter Sets task (Ekstrom, French, Harman, & Dermen, 1976) is a test of verbal fluid intelligence, and the Number Series task (Thurstone, 1938) is a test of numerical fluid intelligence.

Methods

Subjects. Subjects were recruited from Vanderbilt University and the Nashville community; they gave informed consent and received course credit or monetary compensation for their participation. The study was approved by the Vanderbilt IRB. All subjects reported normal or corrected to normal visual acuity, were native English-speakers, and had lived in the United States at least 10 years. Subjects were not specifically recruited for their interest in any of the tested categories; rather, we were interested in testing the range of experience and performance for these categories found in a typically varying sample, although this is constrained by the fact that many subjects were students at the university. Two hundred and seventeen subjects participated in the study. One subject was excluded for not completing all of the tasks and three subjects were excluded for below chance (.33) performance on two or more SVETs. Data reported here are for 213 subjects (86 male) aged 18-55 (mean=22.49, SD=6.31).

Equipment. The experiment was conducted in the laboratory on Apple Mac Minis (OSX 10.9.2, 2Ghz Intel core 2 duo) with 21.5-inch LCD monitors (1920x1080 resolution) using MATLAB R2009b (Mathworks, Natick, MA, USA) and the Psychtoolbox (http://psychtoolbox.org; Brainard, 1997). The experience questionnaire was completed using REDCap electronic data capture survey tools (http://redcap.vanderbilt.edu; Harris et al., 2009) hosted by Vanderbilt University. Subjects were allowed to sit a comfortable distance from the monitor (approximately 40cm).

Stimuli.

CFMT. Stimuli were grey-scale images of faces from varying viewpoints with and without added noise used by Duchaine and Nakayama (2006) in the Cambridge Face Memory Test (see Figure 8).

VET. Stimuli were similar to the grey-scale images of objects with real-world backgrounds used by McGugin et al. (2012b) in the Vanderbilt Expertise Test (VET) (See Figure 6). Some of the VETs used here (dinosaur, Transformer, shoe, and passerine bird) were not included in the original set (McGugin et al., 2012b) but were created in the same way to be paired with the SVET. Other VETs (car, plane, leaf and mushroom) have been revised (images and trials altered) from their original form to improve coverage of the range of performance and dimensionality. All images were 256 x 256 pixels and subtended a visual angle of approximately 5.2 degrees. In the VET for each category, 6 object identities (e.g. cars: Chevrolet Cobalt, Lincoln MKS, Acura RL; birds: Cedar Waxwing, Blue Jay, Horned Lark) were used as target objects. One exemplar of each target was used at study and in same-exemplar trials. Three other

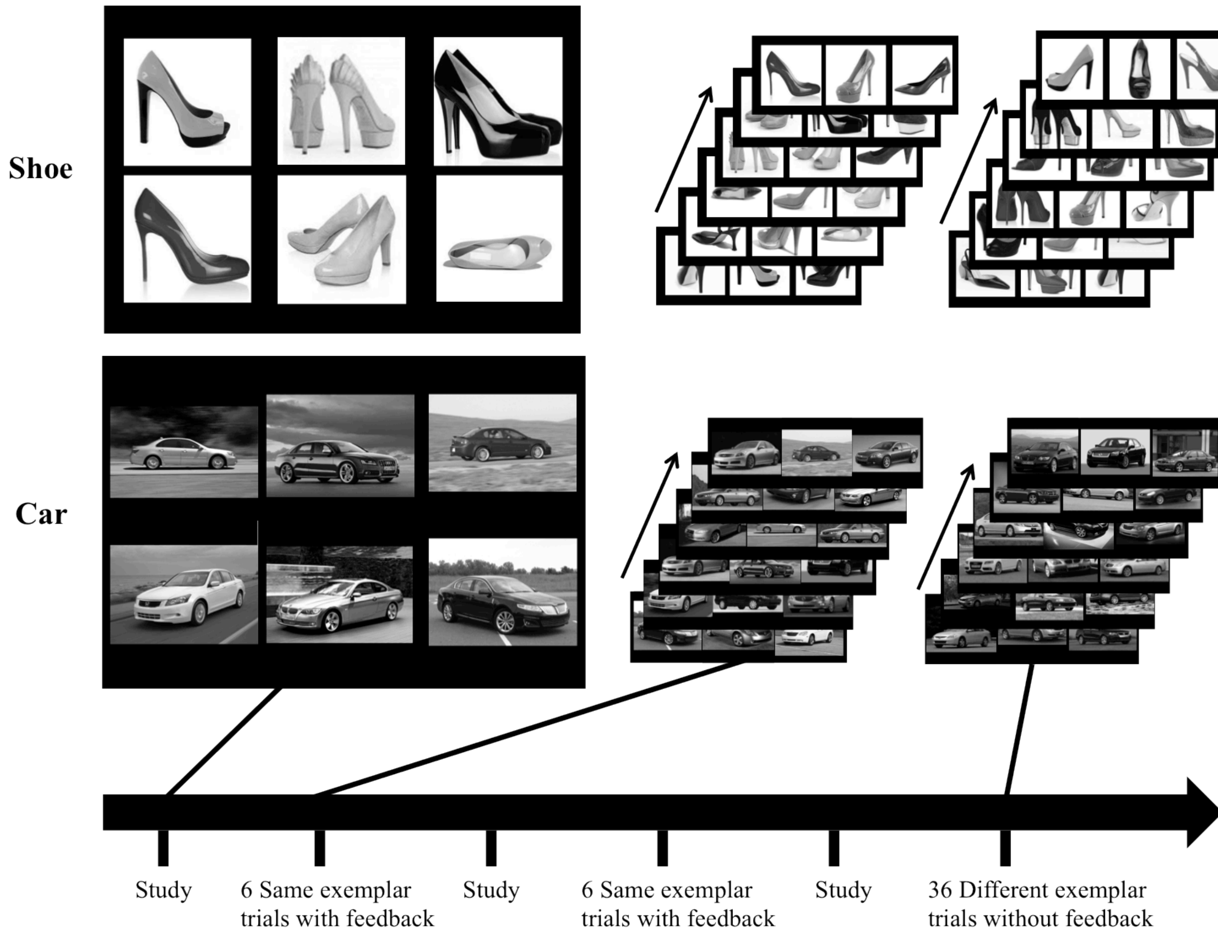


Figure 6. Stimuli and test design for VETs. Figure adapted from McGugin et al. (2012b).

exemplars of each target object were used for different-exemplar trials. The different exemplars of the targets differed from the studied exemplar in one or more ways including background, viewpoint, position, model year, color, and other non-diagnostic features, but were always the same species or model. The objects were selected for each category according to the same guidelines as the SVET, as described in Study 1, to be species or models found in North America when applicable; all birds were passerine birds and only male birds were shown, cars were sedans, leaves were from deciduous trees, shoes were women's high heels, Transformers were shown in multiple forms from any of the series used in the SVET, and planes were a mix of commercial and military planes. Foil images were objects from the same category (e.g. cars, planes, leaves) but of a different type (name as in model or species) than any of the six target objects. Different exemplars of the same foil object occurred between one and four times per VET, but never on the same trial. Catch trial foil images were obviously different from studied objects and were usually selected from a similar category or sub-category that was not studied (e.g. SUVs in the VET-Car, wading birds in the VET-Bird, sneakers in the VET-Shoe).

SVET. SVET names were the same as those described for version 1.0 of each SVET in Study 1.

Fluid intelligence. The stimuli used in these tasks were adapted from Redick et al. (2012) (Figure 7).

Raven's advanced progressive matrices (RAPM). Stimuli were matrices with one missing piece (Raven et al., 1998). Each matrix was a 3 x 3 array of objects with the lower right-hand object missing. The features of the objects in the matrix varied systematically (e.g., object shape, number of lines, direction of lines) according to a pattern that subjects needed to deduce to correctly select which of eight objects would appropriately complete the matrix. Eight single objects were presented below the matrix and labeled with the numbers 1–8, from which subjects selected their response.

Letter sets. Stimuli in each trial were five sets of four letters (e.g. BCCB, GFFG, LMML, QRRQ, WXXW) (Ekstrom et al., 1976). Four of the five letter sets followed a specific rule regarding the order or composition of the set. The sets were displayed in a row in the center of the screen and labeled with the numbers 1–5, which subjects used to select their response. These rules were not based on the sounds or shapes of the letters, or whether the letters formed specific words, but were based on features such as alphabetical order, repetition, or the presence/absence of a specific letter.

Number series. The stimulus in each trial was an array of 5–12 one or two-digit numbers, selected and arranged so that when read from left to right they followed a particular rule (Thurstone, 1938). The pattern governing the number array could apply to single numbers alone or groupings of numbers in the series and could follow either a numerical order (e.g. 1 2 4 1 2 5 1 2 6 interpreted as 124, 125, 126) or a mathematic function (e.g. 2 5 8 11 14 17 interpreted as +3 to each number). Below the number array the numbers 1–5 were presented and labeled "Answer" from which subjects selected their response.

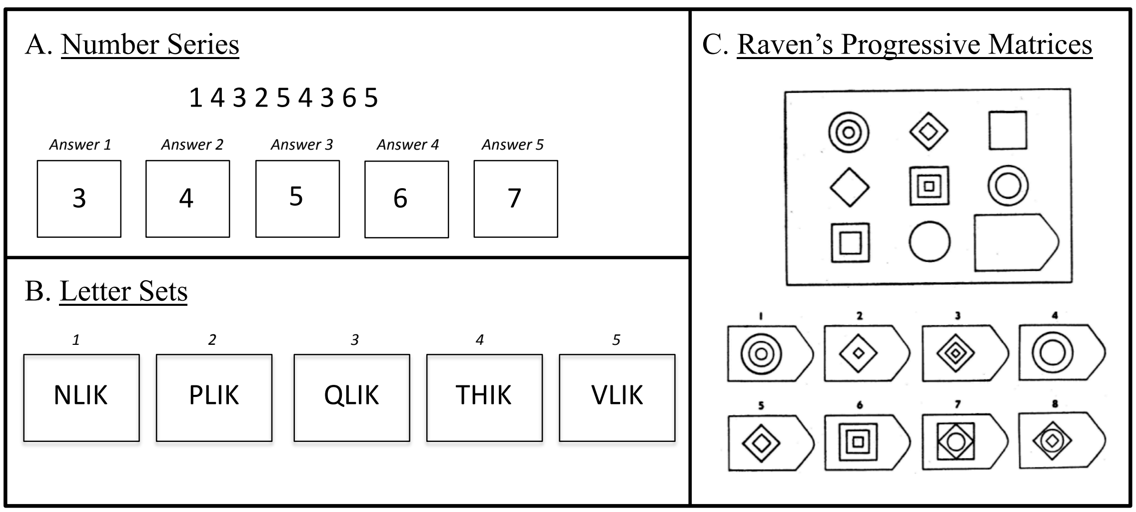


Figure 7. Stimuli and trials for fluid intelligence tests. Examples of a single trial for each task. Stimuli adapted from Redick et al. (2012).

Procedure. To prevent a contribution of order to individual scores, all subjects received the same order of tasks and trials. The experience questionnaire was given first before subjects could be influenced by their task performance. This was followed by the visual tasks, the CFMT, and then the VETs, which do not include any specific names. Subjects then completed a short bird and shoe image naming test, which will be described and reported in Study 2B, followed by the SVETs, and lastly the fluid intelligence tests.

Experience self-report. All questions in this self-report questionnaire of experience were answered as multiple choice questions with radio buttons in a website interface. The questions are the same as those used by Gauthier et al. (in press) to quantify category experience. They specifically ask about verbal, visual, and amount of experience. First, subjects answered 4 domain-general questions and then seven domain-specific questions, answering the same question for each of the eight categories before moving to the next question (see Table 4 for questions). Even-numbered subjects completed the domain-specific questions in reverse order from odd-numbered subjects to counter-balance any potential order effects. All questions were answered on a scale from 1 to 9 described for each questions (1 = very little, 9 = a lot) except for duration experience, which was answered from 1 (no interest) to 5 (6 or more years).

Table 4. Questions used for the self-report of experience questionnaire answered on a scale from 1-9.

General Experience
1. Generally speaking, how strong is your interest in classifying objects in their various sub-categories (such as learning about different kinds of insects, plants, vehicles, tools...)?
2. Generally speaking, how easily do you learn to recognize objects visually?
3. Generally speaking, relative to the average person, how much of your time at WORK or SCHOOL involves recognizing things visually?
4. Generally speaking, relative to the average person, how much of your FREE TIME involves recognizing things visually?

Domain-specific Experience
Note the following order is for odd subject numbers, even subject numbers had the reverse order; XXX = category (e.g. birds, cars)
1. Please rate yourself on your expertise with XXX considering your interest in, years of exposure to, knowledge of, and familiarity with XXX.
2. If you are interested in XXX, when did this interest begin?
3. How often do you look at IMAGES of XXX, in movies, television, or other kinds of documents (books, magazines, or online)?
4. How often do you read TEXT (in books, magazines, online) that contains information about XXX?
5. How important is the domain of XXX to you, relative to all the other things you are interested in?
6. If you saw a specific XXX in a TV show, how sure are you that you could recognize that item among similar images if you were tested the next day?
7. If you were asked to write an essay about different kinds of XXX, how extensive and detailed do you think your essay would be?

CFMT. In this task, subjects learned six target face identities and then had to select those target faces among distractors. The target was sometimes shown in a different orientation, or with different lighting or added noise compared to study (Figure 8). Thus, success on the CFMT requires generalizing across low-level image properties and viewpoints to identify faces. The CFMT has been shown to be a reliable measure that discriminates between individuals at all skill levels (Bowles et al., 2009; Germine et al., 2011a; Germine et al., 2011b; Wilmer et al., 2010; Woolley, Gerbasi, Chabris, Kosslyn, & Hackman, 2008).

First, subjects completed three practice trials with a cartoon face shown from different viewpoints; these are the only trials on which subjects receive feedback. Subjects then studied the sex target face identities for 20 sec. The target images shown at study were front-facing views of six unique individuals. This was followed by an 18 trial learning phase in which they saw three faces, one of which was the exact study image of a test face. In all test trials, they had to select the face that matched one of the six target identities. Every trial contained one target face and two unstudied distractor faces that were not taken from the target set. Next, there was a 30 trial no noise test phase, in which the target image was one of the target face identities, but was not the exact study image. In this phase, the target face images varied from the studied images in viewpoint and lighting conditions. Subjects were told that they would need to select the same identity even though images would not be exactly the same. Subjects were then shown the 6 target face study images again for 20 seconds. Lastly, they completed another 24 trial test phase in which Gaussian noise was added to all of the images. Subjects indicated their responses by pressing 1, 2, or 3 on the number pad to indicate the location of the target face on the screen. Images remained on the screen until subjects made a response. See Duchaine and Nakayama (2006) for additional details.

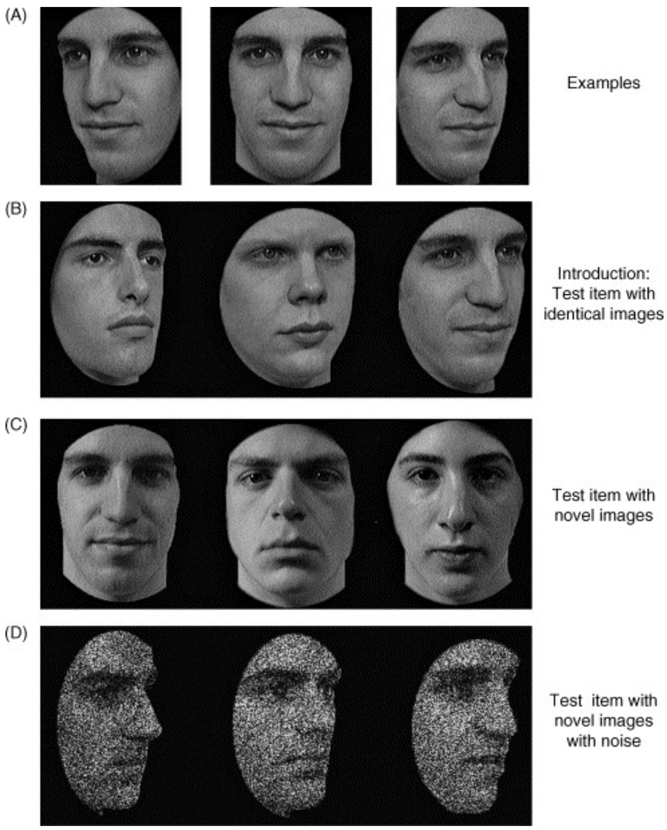


Figure 8. Stimuli for the Cambridge Face Memory Test. Figure from Duchaine and Nakayama (2006).

VET. Each VET task for each of the eight categories began with the presentation of a study screen showing an example of each of the six target objects (See Figure 6). Subjects were told to learn each of these objects and could study them for as long as they chose. No object names beyond the category name (e.g. birds, cars) was ever indicated. In the first 12 trials, the exact studied image of the target appeared along with two foil object images. The images were vertically centered on the screen, to the left of center, at center, and to the right of center. The position of the target was counterbalanced for each trial so that the location varied but each location occurred an equal number of times. Accuracy was stressed in the instructions and images remained on the screen until a response was selected by pressing 1, 2, and 3 on the keyboard for the left, center, and right image, respectively. Subjects received feedback after each of these trials indicating if their response was correct or incorrect. After these 12 exact image trials, the study screen with all six target objects was presented again for subjects to study as long as they desired. Subjects then completed 36 trials in which the target image was a different exemplar of the target than at study or the first 12 trials. They were told that the target would now be the same object but a different image and that they would need to generalize across viewpoint, background, color, size and/or model year, depending on the category. Subjects did not receive feedback on these trials. There were 3 unique exemplar images for each target that were different from the studied image, each of which were shown twice among different foils.

SVET. The procedure was the same as in Study 1, except that subjects indicated their response by pressing the 1, 2, and 3 keys (corresponding to the left, center, and right name, respectively) on the keyboard number pad.

Fluid intelligence. Trials and procedure for our set of fluid intelligence tasks were adapted from Redick et al. (2012) (see also (Hambrick et al., 2007; 2008). For all of the tests, subjects had a time limit in which to complete as many trials as possible. They were informed of this time limit in the instructions and told to focus on accuracy rather than speed in completing the tests. There was no time limit on any specific trial, but a response was necessary to advance to the next trial and subjects could not return to previous trials. In each of the tasks, trials got progressively more difficult. Subjects responded using the keyboard number pad. Practice trials (2-5 per task) were included at the start of each task with a short explanation of the correct response for each problem. No feedback was provided on any of the test trials.

RAPM. Subjects were instructed to determine the pattern that governed objects in the matrix and select which of the eight response objects would correctly complete the matrix. A subset of 18 trials selected from the full advanced matrices test were used (Raven et al., 1998). Subjects were given 10 minutes to complete as many trials as possible up to 18 trials.

Letter sets. Subjects were instructed to select which of the five sets did not follow the same composition rule as the others. Subjects were given 7 minutes to complete as many trials as possible, up to 30 trials.

Number series. Subjects were instructed to determine the rule that governed the number series from left to right in each trial and to select which of five numbers presented as possible responses would continue the pattern as the next number in the series. Subjects were given 5 minutes to complete as many trials as possible, up to 15 trials.

Measurement with item response theory. To further investigate the measurement properties of the VET and SVET and to enhance our ability to examine relationships between measure we will employ techniques from Item Response Theory (IRT). IRT models produce subject scores known as theta that offer several advantages over sum scores to measure performance. The goal of IRT is to produce a more meaningful scale than sum scores on which ability levels can be compared. IRT places items and people on the same scale. One assumption of IRT is that the only reason that item responses are correlated is that there is an underlying theta, which is the latent factor that represents the level of subjects on the ability measured by the test.

IRT takes into account how individual items (trials) function in terms of their discriminability on the model's construct of the underlying trait – different items can provide different amounts of information about an individual's ability, and so are weighted accordingly in the model. Each item is described by an item characteristic function, a mathematical expression that relates a subject's probability of success on an item to the trait measured by the set of items. Unlike sum scores that are based on all test trials equally, theta scores are standardized Z-scores that are based on a measurement model in which item responses are weighted differently to best estimate an individual's level on the construct (see Embretson, 1996; Embretson & Reise, 2000; Lord & Novick, 1968). As a result theta scores make better use of measurements, characterizing individual items on their relative difficult, their ability to discriminate subjects at a given ability level, as well as guessing rates in more complex models, in theory providing more useful scaling of individual differences, as well as valuable information about test items.

When looking to correlate performance on a test with another measure, using thetas may depart from sum scores as a function of the variability in discrimination among the trials on the test, because such variability will lead to different items being weighted differently. IRT scores offer the advantage of being normed based on the test items themselves, rather than on a normative sample, and of being scaled to linearly relate to the underlying ability, properties that facilitate their use when correlating across different measures (Embretson & Reise, 2000).

Basic IRT models assume that tests are unidimensional, meaning that a single factor carries that majority of variance in test performance. The usefulness of IRT measures with the CFMT (Duchaine & Nakayama, 2006), which demonstrate good unidimensionality, has been demonstrated using large datasets both online and in the lab (Cho et al., submitted; Wilmer et al., 2012). To investigate the dimensionality of each VET and SVET, we will conduct exploratory factor analyses (EFA) to determine how many factors carry significant portions of variance. If we find that many of the tests, either in the VET or the SVET set, are multidimensional, we will create versions of those tests by selecting only a subset of trials that load on a single factor. To do this, we will identify the factor out of the two or three that are most significant that correlates best with the other measure for the same category (VET for SVET and vice versa) and then select only trials that load highly on that factor. We will then use this "select trial" version of the test to conduct an IRT model. We will use this specific IRT-pipeline in later analyses to explore if these measurement models produce more easily interpreted patterns of relationships across our measures. For example, we will ask if using theta rather than sum scores increases the extent to which VET and SVETs are more related within then across categories.

Results and discussion

Accuracy, variability and reliability for each measure.

Visual and semantic tests. Accuracy on all of the visual and semantic tests was above chance (.33) and not at ceiling, and all tasks displayed variability between individuals (Figure 9 and Table 5). Skewness was negative for all VETs and positive for all SVETs, suggesting that the SVETs had a more difficult distribution of trials. When considering all CFMT trials, both with and without added noise, accuracy on the CFMT was high, which is consistent with previous face memory results in a typical population (Duchaine & Nakayama, 2006; McGugin et al., 2012b). In general, performance was higher and more variable on the VETs that the SVETs (paired *t*-test of mean VET and mean SVET, $p<.05$ for all categories). While this difference could the result of the difficulty of a visual versus semantic task, it is most likely only because the SVET unintentionally contained more difficult trials. Future revisions of the SVETs should add more easy trials at the beginning of the test to engage subjects and increase overall performance. SVET accuracy in this study largely matches what we observed in Study 1 with an older, online sample. Again, we see the greatest variance in accuracy on SVET-Car and SVET-Shoe, likely a reflection of our sample's greater variability in experience with those categories, especially between men and women.

Table 5. Accuracy on CFMT, VETs and SVETs in Study 2A. Columns show the mean sum score, 95% confidence interval (CI), median, interquartile range (IQR), and skewness.

		Mean	95% CI	Median	IQR	Skewness
	CFMT	0.80	(.78, .82)	0.81	0.17	-0.42
VET	Car	0.59	(.57, .61)	0.60	0.21	0.06
	Plane	0.68	(.66, .70)	0.69	0.21	-0.01
	Transformer	0.72	(.70, .74)	0.73	0.19	-0.19
	Dinosaur	0.70	(.69, .71)	0.69	0.17	-0.28
	Shoe	0.73	(.71, .75)	0.75	0.17	-0.42
	Bird	0.67	(.65, .69)	0.69	0.19	-0.41
	Leaf	0.57	(.55, .59)	0.58	0.15	-0.27
	Mushroom	0.60	(.59, .61)	0.60	0.13	-0.22
SVET	Car	0.62	(.60, .64)	0.60	0.19	0.21
	Plane	0.45	(.44, .46)	0.44	0.13	0.32
	Transformer	0.42	(.41, .43)	0.42	0.13	0.54
	Dinosaur	0.47	(.46, .48)	0.46	0.08	0.80
	Shoe	0.54	(.52, .56)	0.50	0.23	0.43
	Bird	0.47	(.46, .48)	0.46	0.10	0.48
	Leaf	0.53	(.51, .55)	0.52	0.17	0.24
	Mushroom	0.40	(.39, .41)	0.40	0.13	0.30

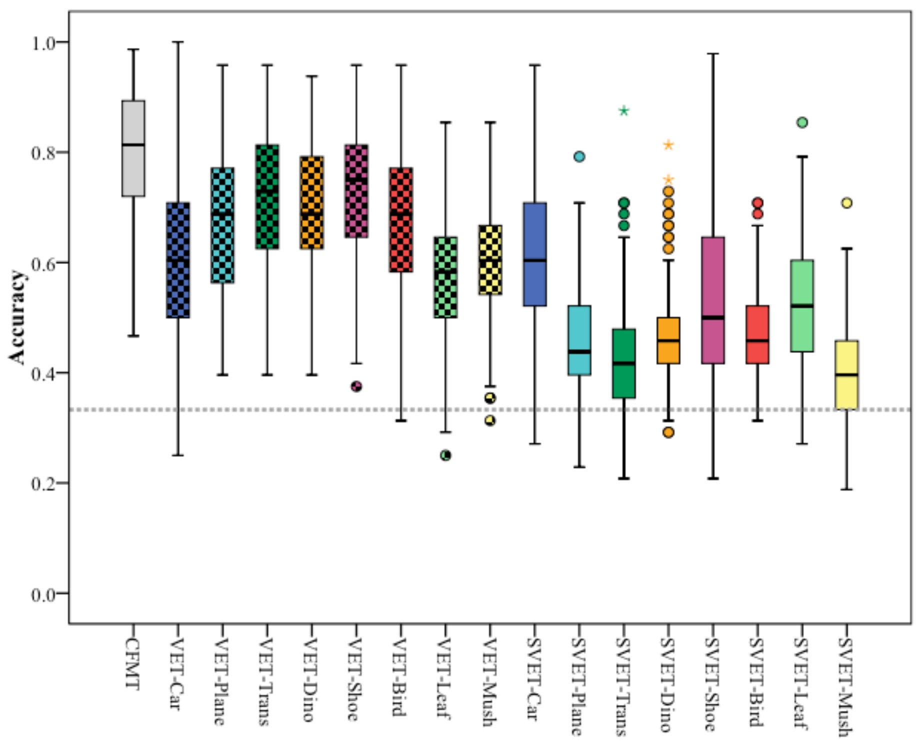


Figure 9. Boxplots of accuracy performance in Study 2A on CFMT (grey), VETs (checkered), and SVETs (solid). The bottom and the top of each box represent the first and third quartile, respectively and the middle line of each box shows the second quartile, or the median. Whiskers show the highest and lowest scores between 1.5 and 3 times the interquartile range while outliers beyond this range are represented by dots (or an asterisk in the case of very extreme outliers). The grey, dotted line shows chance (.33).

The visual measures, CFMT and VETs, all showed high reliability measured as internal consistency with Cronbach's alpha (values reported in Table 8). In general, the reliability of the SVETs was also excellent, with the exception of the SVET-Bird ($\alpha = .52$) and SVET-Mushroom ($\alpha = .66$), which were less reliable. In Study 2A the reliability of the SVET-Dinosaur was higher ($\alpha = .73$), than in Study 1 ($\alpha = .63$). Exploratory factor analyses with the maximum number of factors and polychoric correlation for each of the SVETs in Study 2A show that SVET-Bird, -Mushroom, and –Dinosaur exhibited the most multidimensionality, which is manifest here in lower test reliability. Overall, our set of measures in Study 2A appear to be robust measurement tools.

Experience self-report. In reporting their experience classifying and recognizing objects generally for all categories, subjects reported the highest mean rating for the question asking how easily they recognize objects visually (mean=6.87, SD=1.33) and the lowest mean rating but most variability for the question asking about their interest in fine-level object classification (mean=4.88, SD=1.93). Cronbach's alpha for the four domain-general experience questions was 0.65, suggesting that they measured related experience.

Similarly, self-reports of category experience based on each of the seven domain-specific experience questions differed somewhat but demonstrated high reliability across questions (Cronbach's alpha ranged from 0.83-0.93, mean α =0.88). Table 6 shows the correlations between each of the each of the category-specific experience questions averaged across category (mean r=0.55). The questions were all well-correlated with one another and this was generally consistent across categories, as can be seen from the range reported. Birds demonstrated the least consistency (mean r=0.44) and shoes demonstrated the greatest consistency (mean r=0.69).

Table 6. Correlations (*r*) between each of the seven category-specific experience questions and averaged across the eight object categories; range of correlations shown in parentheses.

		1	2	3	4	5	6
1	Overall Expertise	-					
2	Duration Interest	0.60 (.49-.77)	-				
3	Image Frequency	0.53 (.40-.68)	0.44 (.30-.64)	-			
4	Text Frequency	0.58 (.49-.70)	0.52 (.44-.69)	0.61 (.50-.70)	-		
5	Importance	0.63 (.57-.77)	0.57 (.43-.73)	0.57 (.43-.69)	0.67 (.54-.77)	-	
6	Visual Memory	0.56 (.38-.73)	0.45 (.32-.63)	0.44 (.26-.60)	0.45 (.27-.60)	0.54 (.28-.67)	-
7	Essay	0.68 (.59-.78)	0.51 (.40-.66)	0.47 (30-.62)	0.56 (.49-.67)	0.63 (.52-.77)	0.51 (.34-.66)

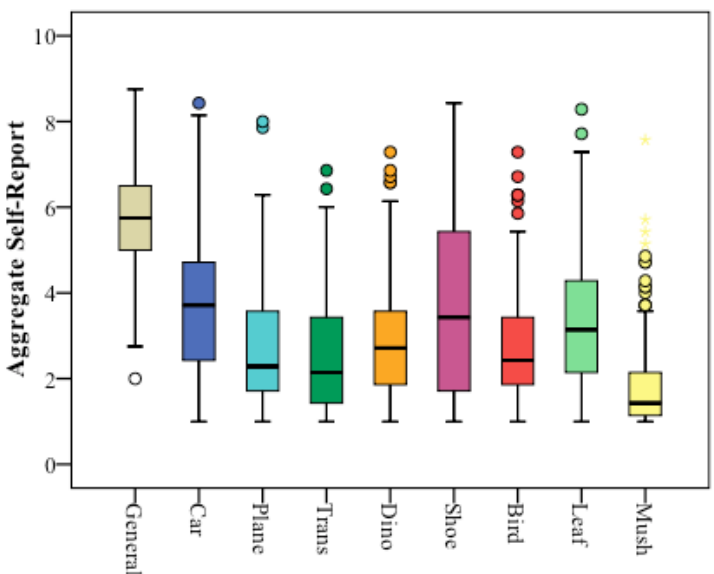


Figure 10. Boxplots showing the general and category-specific experience aggregates for Study 2A.

We computed nine different experience aggregate scores for each subject for use in later analyses: A single general experience aggregate that was the average of the four general experience ratings, and category-specific experience aggregates for each of the eight categories, which in each case was an average of the seven category-specific experience questions.

Figure 10 shows the results of the nine experience self-report aggregates. Subjects reported the highest levels of experience with cars and shoes, although, as in Study 1, those categories also show the highest variance as a result of the interaction between sex and interest for cars and shoes. Reported experience was similar for the other six categories with the lowest average experience reported for mushrooms, although with a notable number of outliers.

Fluid intelligence. Performance on the three measures of fluid intelligence was calculated as the number of trials correctly answered within the time limit and was comparable to that reported in the literature (Redick et al., 2012) (Number series: mean=10.27, SD=2.75; Letter sets: mean=18.08, SD=4.12; RAPM: mean=10.85, SD=2.91). Each of the tests demonstrated high reliability calculated with Cronbach's alpha (Number series: α=0.85; Letter sets: α=0.87; RAPM: α=0.92). Considering all trials from the three different fluid intelligence tests together as a single measure also exhibited high internal consistency (α=0.92). This allowed us to calculate a fluid intelligence aggregate score (that we will refer to as *Gf*) for each subject as the mean number of correct trials on each of the three tests, which we will use in later analyses.

Correlation between VET and SVET with experience. Using the aggregate scores of self-reported experience for each category, as well as general object classification, we can look at the correlations between experience and accuracy on the VET and the SVET for each category (Table 7). First, we consider the relationship with general experience. VET performance is generally not strongly correlated with general experience, although more so for Transformers and birds. SVET performance in general was more related to general experience than VET, with SVET-Transformer, -Bird, -Car, -Dinosaur-, and -Leaf demonstrating a significant correlation. However, these relationships with general experience are modest and while they are specific to category-specific visual or semantic performance, they could reflect domain-general interests or skills. For example, looking at all four general experience questions for Transformers and birds, the greatest correlation values with the VET are with the first two questions (see Table 6) that ask about classification and visual recognition, skills which may be more related to the structure of these categories compared to others. Second, we consider the relationship with category-specific experience, which shows a significant correlation with both the VET and the SVET for all categories, with the exception of the VET-Mushroom. Notably, all of the category-specific correlation values are higher for the SVET (mean r=0.44) than the VET (mean r=0.27).

Table 7. Correlations (r) of general experience aggregate and category-specific experience aggregate for each category with VET and SVET accuracy as well as VET-Other and SVET-Other accuracy (the average of accuracy on the other 7 categories) in Study 2A. Values shown in bolded red are statistically significant ($r_{Crit}(211)=.132$, $p<.05$).

	Correlation (r) of General Experience Aggregate with:		**Correlation (r) of Category Experience Aggregate with:**			
	VET	SVET	VET	SVET	VET-Other	SVET-Other
Car	0.11	**0.14**	**0.42**	**0.52**	0.03	0.13
Plane	0.10	0.09	**0.17**	**0.32**	-0.06	0.06
Transformer	**0.16**	**0.13**	**0.23**	**0.55**	-0.02	0.05
Dinosaur	0.06	**0.17**	**0.32**	**0.50**	0.08	**0.32**
Shoe	-0.02	-0.01	**0.50**	**0.62**	0.01	**-0.29**
Bird	**0.24**	**0.16**	**0.26**	**0.42**	0.06	**0.26**
Leaf	0.05	**0.14**	**0.15**	**0.36**	0.10	**0.24**
Mushroom	0.00	0.07	0.11	**0.28**	0.13	0.13

As expected, we found that both VET and SVET were related to self-reported experience, the fact that this relationship was stronger for SVET suggests that subjects based their self-ratings of experience on semantic knowledge more than perceptual performance. This could be because people have more access to other people's semantic knowledge (through its verbal expression) than to other people's perceptual knowledge. These relationships seem to be specific to each category, and not general effects across all objects: the correlation between VET and SVET with category experience is always greater within category (with VET and SVET for the same category) than across category (with VET-Other and SVET-Other, the average of accuracy on the other seven categories for each test – see Table 7). Overall, VET and SVET both demonstrate category-specific validity as they are well correlated with self-reports of category-specific experience.

Variance accounted for by age and sex. Next we considered the correlations between VET and SVET performance for each of the eight categories and age and sex (Table 8). Age is only strongly correlated with *Gf* and a subset of the SVETs. The negative correlation between *Gf* and age is very likely due to a sampling bias, as the majority of our subjects were from the university community and under 30 years of age, and the few older subjects we had in this study may not represent the same population in *Gf* performance. SVET performance was positively correlated with age, significantly for cars, birds, leaves, and mushrooms, which is likely the result of older subjects having more experience and thus more name knowledge for these categories. However, especially for the categories of birds, leaves and mushrooms, age may also be correlated with a particular interest in these living categories as a hobby.

Women performed better than men on the CFMT, a sex advantage that has been shown before using this face measure (Bowles et al., 2009; Duchaine & Nakayama, 2006). Because of previously observed sex effects in visual performance (McGugin et al., 2012b), we specifically selected a mix of categories for which we predicted men or women would perform better. This prediction for our four male-interest and four female-interest categories was largely borne out in our data, however the effect of sex was larger for male-interest categories for SVET than VET (VET: male-interest mean r=0.05, female-interest mean r=-0.22; SVET: male-interest mean r=0.30, female-interest mean r=-0.10). For both VET and SVET, only shoes demonstrated a strong sex effect for which women performed better than men. The sex effects we observed for VETs show the same pattern reported by McGugin et al. (2012b), although their sex effect for visual performance with male-interest categories, all vehicles in their study (cars, planes, and motorcycles), were considerably stronger than what we observed for the same categories.

Correlation between VET and SVET and visual and verbal ability. Next we consider the relationship between VET and SVET performance for each category with face recognition (CFMT) and verbal (*Gf*) abilities (Table 8). Note that even though the CFMT is a domain-specific test, because faces are generally thought to be a particularly distinct domain in high-level recognition, shared variance between VET and CFMT might be interpreted as a domain-general visual ability. As predicted based on previous findings (Gauthier et al., n.d.; McGugin et al., 2012b), CFMT was positively correlated with VET performance for all categories, suggesting that a common visual ability contributes to performance on all of these visual tests. As found before (Gauthier et al., in press), the CFMT is not particularly distinct from all the other visual tests (VETs): the average correlation between CFMT and VETs was r=0.26 (range = 0.19-0.38) (Table 8), while the mean pairwise correlation among VETs was r=0.33 (range = 0.09-0.46) (Table 9). This suggests that the modest correlation between face and object recognition tasks is only one example of a general principle whereby such visual measures seem to primarily reflect domain-specific variance. Interestingly, the correlation between CFMT is stronger with the average performance on all eight VETs (VET-All) (r=0.40) than it is with any single category VET, which may demonstrate that aggregating the VET across categories reduces categories-specific contributions producing a measure that more closely reflects domain-general visual ability. The relationship between CFMT and SVET performance (mean r=0.08), a non-visual measure, was on average weaker than for CFMT-VET correlations (mean r=0.26, two-sided Fisher's Z=-2.69, p=0.007). However, for two categories (dinosaurs and shoes), visual ability and semantic performance were significantly correlated.

Table 8. Correlations of test accuracy with age, sex, *Gf* and CMFT in Study 2A. The first column shows the reliability of each measure as Cronbach's alpha. Columns 2-5 show the correlation (r) of each accuracy on each measure with age, sex, *Gf* accuracy, and CFMT accuracy. Values shown in bolded red are statistically significant ($r_{\text{Crit}}(211)=.132$, $p<.05$).

		Correlation (*r*) of test with:			
	Cronbach's α	**Age**	**Sex**	***Gf***	**CFMT**
Gf	0.92	**-0.25**	0.06	-	
CFMT	0.92	-0.07	**-0.15**	**0.14**	-
VET-Car	0.90	-0.07	0.06	0.01	**0.21**
VET-Plane	0.89	0.08	0.09	**0.28**	**0.22**
VET-Transformer	0.89	-0.01	0.06	**0.29**	**0.28**
VET-Dinosaur	0.88	0.07	0.00	**0.29**	**0.26**
VET-Shoe	0.89	0.00	**-0.54**	0.10	**0.38**
VET-Bird	0.93	0.01	-0.05	**0.29**	**0.24**
VET-Leaf	0.84	-0.04	**-0.16**	**0.25**	**0.28**
VET-Mushroom	0.71	0.03	-0.12	**0.21**	**0.19**
SVET-Car	0.89	**0.22**	**0.28**	-0.08	0.06
SVET-Plane	0.72	0.10	**0.36**	**0.14**	-0.01
SVET-Transformer	0.74	0.02	**0.30**	**0.15**	0.07
SVET-Dinosaur	0.73	0.09	**0.24**	**0.25**	**0.14**
SVET-Shoe	0.91	0.04	**-0.55**	-0.11	**0.18**
SVET-Bird	0.52	**0.29**	0.02	0.04	0.06
SVET-Leaf	0.77	**0.39**	-0.01	0.04	0.06
SVET-Mushroom	0.66	**0.16**	0.13	**0.14**	0.10

We observed that *Gf* and CFMT were positively correlated. While this correlation is small, it is important because this relationship has not be observed previously (Davis et al., 2011; Hedley et al., 2011; Wilhelm et al., 2010; Wilmer et al., 2010). *Gf* was positively correlated with VET accuracy for most categories, although not significantly for cars and shoes, and positively correlated with SVET accuracy for planes, Transformers, dinosaurs, and mushrooms. Only the SVET-Car and SVET-Shoe were slightly negatively related to *Gf*, suggesting that the strong sex differences in experience and performance with those categories in our sample may influence their relationship with other variables.

Of course, when calculating the correlation between two measures it is important to consider the reliability of each measure (Cronbach's alpha in Table 8), as test reliability limits correlation. Reliability is high for nearly all of these measures, so performing a correction for attenuation on these correlations is largely not a concern. However, as an example, the corrected Pearson's *r* of SVET with *Gf* for three SVETs that demonstrate medium reliability are 0.06, 0.05, and 0.18 for SVET-Bird, -Leaf, and -Mushroom, respectively. These correlation coefficients after correcting for attenuation are only slightly larger than the uncorrected correlations reported in Table 8, suggesting that our measures are reliable enough to safely interpret the pattern of correlations without adjustment.

Correlation between VET and SVET within and between categories. We considered the relationship between performance on the VET and SVET within and between the eight categories using correlations between accuracy on each VET and SVET (Table 9). Panel A of Table 9 shows that nearly all VETs are positively correlated with each other, with the exception of VET-Bird and VET-Mushroom with VET-Car. The pair-wise correlations between VET accuracy was r=0.32 for all male-interest categories (cars, planes, Transformers, and dinosaurs) and r=0.37 for female-interest categories (shoes, birds, leaves, mushrooms). However, these correlations were not on average much stronger than the correlations between VET categories across sex-interest r=0.31(e.g.VET-Car with VET-Shoe). Contrary to previous findings with the VET (McGugin et al., 2012b), we did not observe strong sex effects in VET performance.

Looking at the relationship between SVETs (Panel C of Table 9), we see a strong positive correlation between all of the male-interest categories (mean r=0.27), although for female-interest categories, SVET-Shoe performance appears to be unrelated to the living, female-interest categories, birds, leaves, and mushrooms, which are positively correlated with each other (mean *r*=0.31). SVET-Shoe performance is also unrelated or even negatively related (planes and Transformers) with the male-interest categories. This suggests that sex is but one of several factors that are related to experience with object categories.

In considering the relationships between VETs and SVETs, we were most interested in the relationship between VET and SVET accuracy for the same category, which might indicate how common category experience contributes to both visual and semantic performance (Panel C of Table 9, same category shown in outlined diagonal). We observed a positive correlation between VET and SVET performance for all categories, with nearly all relationships being significant. Leaves and mushrooms are the only two categories that were not significant ($p<.05$). We will continue to explore these SVET-VET relationships within each category. Off the diagonal in Panel B of Table 9, we observed both positive and negative relationships between VET and SVET between categories that may reflect related interest, or lack of interest, across categories (e.g. VET-Plane with SVET-Car and VET-Bird with SVET-Mushroom are positively correlated, but VET-Shoe and SVET-Plane are negatively correlated). To determine if the VET-SVET correlation was stronger within category versus between categories, we compared the relationship between the VET and SVET for the same category versus SVET-Other (the average of performance on the other seven SVET categories). For six of the eight categories, excluding birds and leaves, we found that the VET-SVET correlation was stronger within than between categories, a difference that was statistically significant using an independent two-sided Steiger's Z ($p<.05$) for three of the categories (cars, Transformers, and shoes). We also compared the relationship between SVET and VET within category versus VET-other (average performance on the other VET categories), where we observed an even more robust pattern of results. We found that for all categories except leaves and mushrooms, the SVET-VET correlation was greater within category than between categories, an effect that was statistically significant using an independent two-sided Steiger's Z (p<.05) for all six categories (cars, planes, Transformers, dinosaurs, shoes, and birds). This suggests that the relationship between VET and SVET performance is largely category-specific and supports our hypothesis that shared VET-SVET variance primarily reflects common category experience.

Table 9. Correlations (r) of VET and SVET accuracy for each category in Study 2A. Panel A shows the correlations between each of the VETs. Panel B shows the correlations between each SVET and VET with within category correlations outlined along the diagonal. Panel C shoes the correlations between each of the SVETs. Values shown in bolded red are statistically significant ($r_{Crit}(211)=.132$, p<.05).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		A. VETs															
1	VET-Car	-															
2	VET-Plane	**0.26**	-														
3	VET-Transformer	**0.25**	**0.46**	-													
4	VET-Dinosaur	**0.15**	**0.46**	**0.36**	-												
5	VET-Shoe	**0.24**	**0.28**	**0.25**	**0.25**	-											
6	VET-Bird	0.09	**0.39**	**0.44**	**0.44**	**0.26**	-										
7	VET-Leaf	**0.25**	**0.43**	**0.44**	**0.38**	**0.41**	**0.44**	-									
8	VET-Mushroom	0.09	**0.36**	**0.41**	**0.28**	**0.27**	**0.37**	**0.46**	-								
		B. SVET and VET								**C. SVETs**							
9	SVET-Car	**0.45**	**0.25**	0.09	0.02	-0.08	-0.09	0.03	**-0.14**	-							
10	SVET-Plane	0.08	**0.32**	**0.16**	**0.21**	**-0.18**	0.11	0.04	0.02	**0.35**	-						
11	SVET-Transformer	0.05	**0.22**	**0.30**	**0.26**	-0.07	**0.25**	0.05	-0.03	**0.17**	**0.28**	-					
12	SVET-Dinosaur	0.03	**0.14**	**0.18**	**0.31**	-0.11	**0.19**	0.02	-0.07	**0.17**	**0.25**	**0.38**	-				
13	SVET-Shoe	**0.18**	-0.07	-0.10	-0.02	**0.40**	-0.05	0.03	-0.05	0.07	**-0.18**	**-0.22**	-0.02	-			
14	SVET-Bird	-0.13	0.03	0.02	**0.17**	0.02	**0.16**	-0.01	0.01	0.06	**0.16**	0.03	**0.29**	-0.09	-		
15	SVET-Leaf	0.01	0.11	0.11	**0.21**	0.11	**0.16**	0.08	-0.02	**0.31**	**0.22**	**0.19**	**0.30**	0.06	**0.38**	-	
16	SVET-Mushroom	0.10	**0.24**	0.11	**0.29**	-0.01	**0.18**	0.07	0.12	**0.15**	**0.22**	**0.28**	**0.32**	0.01	**0.28**	**0.26**	-

SVET-Select scores from the IRT pipeline. Finally, we used multiple regression to ask what variables contribute to the relationship between VET and SVET performance for each category. We were interested in looking at this relationship both with the sum scores we have discussed thus far as well as with theta scores from an IRT model of each test. We used a specific IRT-pipeline defined here to explore whether such an approach produces more domain-specific effects that are more likely related to category-specific experience.

To use a 2-parameter IRT model on these tests, it is important that the assumption of unidimensionality be met. To investigate the dimensionality of each VET and SVET, we first performed exploratory factor analysis (EFA) with the maximum number of factors. For VETs we only used data from Study 2A subjects (N=213), but to maximize power for SVETs we used data from subjects in both Study 2A and Study 1 (N=116 form AMT; N=319 total). We found that all of the VETs were unidimensional but that many SVETs were multidimensional and included trials that loaded more on factor 2 or 3 than factor 1, or that loaded on more than one factor equally. To create a SVET trial set that was unidimensional for use with IRT, we selected a subset of SVET trials, SVET-Select, that reflect a single factor. When several factors were strong, we chose the one that was most correlated with VET performance. Note that this choice could inflate the VET-SVET correlations within category using the IRT pipeline scores, but it should not affect relationships with any of the other variables.

To create the SVET-Select subsets, we first looked at the data for all test trials of each SVET in a polychoric principal factor analysis of three factors under rotation to reduce the cross-correlation between factors. We chose to use three factors because two or three factors accounted for most of the variance on all the SVETs (as in Cole et al., 2012). To determine which rotation was most appropriate for our data set, we tested several methods on two categories and then applied the best rotation to all SVETs. Because we did not believe that the factors contributing to performance on these tests were necessarily independent, we limited testing to oblique rotations. We tested several oblique rotations including Oblimin and Promax and found similar patterns of results, so we selected the Oblimin rotation, which is a common and well-accepted rotation that allows for non-independent factors. Using the results of the Oblimin-rotated factor analysis we identified trials that loaded strongly ($\leq$-0.3 of $\geq$0.3) on each of the three factors. To ensure that we selected the SVET factor that was most related to VET performance, we looked at the correlation between the subset of trials loading most strongly for each factor with VET performance for the same category. For all categories, the highest correlation between sub-SVET and VET was for SVET factor 1, with the exception of cars, for which SVET factor 2 was more correlated with VET performance. For mushrooms and leaves no factor was strongly correlated with VET so we selected factor 1 since that was the dominant factor for the SVETs overall. To create a unidimensional "SVET-select" measure for each category, we selected only these trials that loaded strongly on the VET-related factor, between 14 and 30 trials for each SVET. Table 10 shows the properties of SVET-Select for Study 2A subjects and demonstrates unidimensionality of the sub-test with high eigenvalues for factor 1 relative to factor 2 and 3.

To obtain theta scores for SVET-Select, we ran a two-parameter IRT model with the selected trials using data from Study 2A and subjects from Study 1 who took only one SVET each (these additional subjects in each model help reduce measurement error). We then extracted the theta scores of Study 2A subjects for each SVET-Select. We also ran a 2-parameter IRT model for each VET using only Study 2A subjects and all VET trials to obtain VET thetas.

Table 10. Properties of SVET-Select trials chosen as a subset of each SVET in Study 2A. Shown are the number of trials (out of 48 total on each SVET) included in SVET-Select, mean sum scores and standard deviation in parentheses, correlation (*r*) of SVET-Select with VET, and eigenvalues for Factors 1, 2 and 3. Mean and correlation are for Study 2A subjects only, eigenvalues are from an exploratory factor analysis with Study 1 and Study 2A subjects.

SVET-Select	Number of Trials	Mean (SD)	Correlation (*r*) with VET	Factor 1 Eigenvalue	Factor 2 Eigenvalue	Factor 3 Eigenvalue
Car	20	0.59 (.18)	0.42	5.10	1.25	0.92
Plane	19	0.38 (.15)	0.31	5.35	1.38	1.19
Transformer	18	0.45 (.21)	0.33	4.74	1.04	1.03
Dinosaur	16	0.34 (.15)	0.21	3.52	1.04	0.93
Shoe	30	0.57 (.22)	0.40	9.82	1.66	1.25
Bird	14	0.51 (.16)	0.33	3.37	0.82	0.78
Leaf	20	0.53 (.20)	0.11	5.63	1.47	1.11
Mushroom	16	0.41 (.15)	0.07	3.37	1.10	0.89

Contribution of experience to within category VET-SVET correlations. One of our goals in measuring visual and semantic performance for object categories was to determine which abilities and variables might contribute to performance, and in particular if common category experience can explain the shared variance between visual and semantic performance. We collected or measured several variables, including age, sex, and *Gf*, not because we were necessarily interested in their contribution to the VETs and SVETs *per se*, although those contributions are notable, but because we wanted to isolate the contribution of common category experience by accounting for variance due to other variables.

To understand the contribution of category experience to the correlation between VET and SVET performance, we performed several multiple regressions that progressively account for more sources of correlation between VET and SVET, before ultimately testing for the effect of shared experience. These analyses were conducted with both sum scores of accuracy for all trials, as used in previous analyses, and also theta scores for both tests based on our IRT pipeline. These two different ways of quantifying test performance show largely the same pattern of results.

The zero-order correlations between VET and SVET performance for each category are shown in Table 11, column A. The results are similar using our IRT-pipeline; most categories show the same or a greater positive relationship between VET and SVET, with the exception of planes, which are slightly more correlated using sum scores. Column B presents the partial correlations between VET and SVET performance residuals for each category with age, sex, and *Gf* entered simultaneously as predictors. Using both sum scores and the IRT-pipeline scores, we observed a slight decrease in the correlation coefficient, although planes increased slightly, once the contributions of age, sex, and *Gf* were removed. The notable exception to this trend is the

VET-SVET-Shoe relationship, which was substantially reduced from column A to column B. This reduction is almost entirely due to the contribution of sex, which is strongly correlated with both VET-Shoe and SVET-Shoe performance (as seen in Table 11). Overall, age, sex, and *Gf* do not account for much of the shared variance between VET and SVET performance.

Table 11. Correlations and partial correlations (r) of VET and SVET performance for each category in Study 2A – at top using sum scores and all SVET trials and bottom using theta and SVET-Select trials. Column A shows the zero-order correlations (also shown in *Table 9*, Panel B, diagonal boxes). Column B shows the partial correlations with age, sex, and *Gf* regressed out. Column C shows the partial correlations with age, sex, and *Gf* as well as VET-Other and SVET-Other performance regressed out. Column D shows the partial correlations with age, sex, *Gf* , VET-Other, and SVET-Other performance, as well as self-report category experience aggregate regressed out.

		A. Zero-Order	B. Age, Sex, *Gf* Out	C. B + VET-Other, SVET-Other Out	D. C + Category Experience Out
	α=.05:	r_{Cnt}=.13	r_{Cnt}=.13	r_{Cnt}=.14	r_{Cnt}=.14
Sum Score	Car	**0.45**	**0.49**	**0.49**	**0.37**
All Trials	Plane	**0.32**	**0.27**	**0.24**	**0.19**
	Transformer	**0.30**	**0.27**	**0.21**	0.08
	Dinosaur	**0.31**	**0.25**	**0.22**	0.13
	Shoe	**0.40**	**0.16**	**0.19**	0.12
	Bird	**0.16**	0.03	**0.17**	0.08
	Leaf	0.08	0.07	0.06	0.01
	Mushroom	0.12	0.09	0.11	0.09
Theta	Car	**0.53**	**0.54**	**0.54**	**0.41**
SVET Select	Plane	**0.28**	**0.22**	**0.18**	**0.15**
	Transformer	**0.31**	**0.27**	**0.24**	0.11
	Dinosaur	**0.31**	**0.27**	**0.23**	**0.15**
	Shoe	**0.42**	**0.20**	**0.24**	**0.18**
	Bird	**0.35**	**0.30**	**0.25**	**0.16**
	Leaf	**0.16**	0.13	0.06	0.01
	Mushroom	0.13	0.12	0.10	0.08

Next, we asked how much of the shared variance between VET and SVET for a category might be the result of a domain-general ability reflected in performance across all tests. Perhaps some individuals are just very good at these types of tasks with any object category, and we would not want to consider that effect to be domain-specific. For each VET and SVET we created a non-category score for each subject (VET-Other and SVET-Other), which was the average performance on the other seven categories for that test (e.g. SVET-Other for cars is the average of performance on all SVETs except SVET-Car). We then used VET-Other and SVET-

Other as regressors to remove domain-general task performance from the VET-SVET relationship. Table 11, column C shows the partial correlations between VET and SVET performance residuals for each category with age, sex, *Gf* aggregate, VET-Other, and SVET-Other performance all entered as simultaneous predictors in a multiple regression. Moving from column B to C, now accounting for domain-general test performance in addition to the other variables, we observe very little change in the correlation between VET and SVET performance with either sum scores or theta with SVET-Select. While a more substantial change can be seen for birds using sum scores, this change is not seen using theta scores. This can also be seen in the change between column A and B, suggesting that, especially for birds, the IRT pipeline seems to produce more robust measurements. Overall, there was little if any influence of domain-general test performance on the shared variance between VET and SVET. At this point looking at the correlations in column C, we hypothesize that the remaining shared variance between VET and SVET comes from domain-specific experience.

To test this hypothesis, we performed another regression including domain-specific experience as a predictor (Table 11, column D). Theoretically, if our hypothesis was correct *and* our measure of category experience contained no measurement error (an assumption of mediation analyses that is almost universally violated; Baron and Kenny, 1986), we would expect the correlation between VET and SVET to be completely eliminated by regressing out experience.

Column D shows the partial correlations between VET and SVET performance residuals for each category with age, sex, *Gf*, VET-Other, SVET-Other performance, self-report category experience aggregate removed in a simultaneous regression. Between the zero-order correlations (column A) and column C, we observed little meaningful change in the correlation between VET and SVET as we accounted for age, sex, *Gf*, and domain-general test performance. As we account for the contribution of category experience (column D), we now observe a substantial decrease in the VET-SVET partial correlation for several of the categories, in some cases resulting in non-significant correlations.

To estimate the contribution of experience more directly, we performed a 2-step, stepwise regression on VET scores. In step 1, we used age, sex, *Gf*, and non-category VET and SVET as predictors. In step 2, we entered both SVET and self-report category experience aggregate. We report the amount of variance (R^2) accounted for by experience in VET performance after controlling for SVET (Table 12). We found that category experience contributed a significant portion of variance for cars, planes, Transformers, dinosaurs, birds, and leaves. These results suggest that, as hypothesized, shared variance between visual and semantic performance can be explained, at least in part, by domain-specific experience.

It is interesting to note that the correlation coefficients in Table 11, column D are not zero, as theoretically predicted. One reason for this could be the imprecision of our experience measure, which relies on self-reports, and which may not accurately reflect real-world levels of experience (Zell & Krizan, 2014). There could also be qualitatively different aspects of domain-specific experience that we did not access with this measure. Alternatively, there could be other subject characteristics that are relevant to specific domains besides age, sex, *Gf*, and reported experience. We can honestly only speculate about what those might be, perhaps individual differences in attention or memory or unreported knowledge or skills that would influence both visual and semantic performance for a given category.

Table 12. Variance (R^2) in VET performance explained by experience after controlling for SVET and significance (*p*). These are shown at top for sum scores and all SVET trials and on bottom for theta and SVET-Select trials.

		Experience Variance in VET(SVET)	
		R^2	*p*
Sum Score	Car	0.04	0.003
All Trials	Plane	0.02	0.033
	Transformer	0.04	0.005
	Dinosaur	0.03	0.007
	Shoe	0.01	0.319
	Bird	0.05	0.001
	Leaf	0.02	0.032
	Mushroom	0.01	0.324
Theta	Car	0.03	0.019
SVET Select	Plane	0.03	0.009
	Transformer	0.03	0.022
	Dinosaur	0.03	0.007
	Shoe	0.00	0.743
	Bird	0.04	0.002
	Leaf	0.03	0.012
	Mushroom	0.01	0.251

Conclusions

In Study 2A we tested the SVET in the laboratory in a large sample from the Vanderbilt and Nashville community. We tested the SVET with a visual measure (VET) for each category as well as measures of domain-general visual and verbal abilities and an expanded self-report measure of category experience.

Here again, the SVET produced reliable measurements of semantic experience in a different sample collected in person that differed in age from our previous online sample in Study 1. We also obtained good reliability and similar performance as previously reported for the VETs with some of the same categories (McGugin et al., 2012b). Using an expanded questionnaire of seven category-specific questions about experience, we were able to demonstrate validity of the VET and SVET. Our seven-item questionnaire of domain-specific experience demonstrated good internal consistency, which is notable because there are not standard ways of measuring experience across various categories, and this at least offers a viable option. Using this measure we showed that the correlations between VET and SVET performance and experience was stronger within category than across categories.

We then examined the relationships between VET and SVET and age and sex, demonstrating better SVET performance with increasing age and finding sex effects in VET and SVET performance for different categories that we predicted to be *male* or *female* categories. Interestingly, we found the strongest sex effects for male-interest categories on the SVET, but also for SVET-Shoe for women. We observed patterns of sex effects for the VET that were qualitatively similar to those reported previously (McGugin et al., 2012b), although they were not as strong. The relationship between VET and CFMT performance demonstrated a common visual ability that contributes to performance for faces and each of our eight categories; however, the CFMT was less related to semantic performance, as would be expected given that it is a non-visual task. Fluid intelligence was positively related to performance on all VETs and most SVETs. The *Gf* relationship was on average stronger for VETs than SVETs, which might be because the VET is a learning task while the SVET tests mostly previously acquired knowledge. These results demonstrate the contribution of domain-general abilities that should be accounted for when examining category-specific performance.

With the addition of the VET in Study 2A we were able to examine performance for the same category on two different tasks, a visual memory test and a non-visual semantic test. Importantly, by looking at the relationships between all eight VETs and SVETs we observed that the relationship between visual and semantic performance within a category was always greater than the relationship between visual performance for a category and semantic performance for other categories. This further demonstrates that the VET and SVET measure category-specific performance and provides evidence for our hypothesis that performance across tasks reflects common category experience. We explored this VET-SVET relationship in greater depth to determine what other factors might contribute. We found that the relationship was still significant after removing contributions of age, sex, *Gf*, and non-category task performance, but that experience as measured by our self-report experience aggregate accounted for a significant portion of the variance for five of the categories. This finding demonstrates the critical importance of considering both domain-general ability and domain-specific experience when interpreting individual differences in domain-specific performance.

Study 2B: The Case of a Category Without Known Semantic Labels

Overview

We hypothesized that the underlying abilities that support the acquisition of visual and semantic knowledge may be independent, and therefore predicted that the only common contribution to visual and semantic performance for a given category would be experience with that category. One reason for this hypothesis is that face recognition has been found to be independent from general intelligence (*Gf*) and related measures (Davis et al., 2011; Hedley et al., 2011; Wilmer et al., 2010), and face recognition can be argued to be a relatively pure measure of visual ability (*v*), since there should be minimal contribution from differences in experience for this category. Therefore, if *Gf* and *v* are independent, then experience in a domain could be the main source of a correlation in *performance* on visual and non-visual knowledge acquired through these abilities. However, this is a conjecture, and the generalization to a domain-general *v* also depends on the assumption that what we learn from face recognition ability speaks to non-face recognition ability. In addition, it is worth considering whether there may be other possible sources of correlation between tasks like the VET and SVET.

One other possible reason for the correlation between visual and semantic performance could be the use of labels to help encode and remember the objects in a visual task like the VET. Even when labels are not provided anywhere in the test, some subjects with semantic knowledge may automatically use object names in a visual task that does not require them. Therefore, it is important to consider the extent to which performance on the visual tasks is potentially contaminated by verbal strategies.

Importantly, a verbal strategy may not be equally available for all categories. There are categories for which every exemplar has a name that is likely available to experts, such as cars. In this case, naming of visual images could be an automatic and common strategy, provided that the subject knows the object names. However, for other categories, semantic knowledge, at least individual object names, might not be as readily available, even to an expert. This would reduce the potential overlap between semantic knowledge and performance on a visual task. Shoes (in our case women's high heels) are an example of such a category. Individuals highly familiar with women's high heels might be very good at recognizing diagnostic visual features of women's high heels, such that they would do very well on the VET-Shoe in which they need to generalize across non-diagnostic features (color, material, viewpoint) but not diagnostic features (toe shape, heel design, heel height) to recognize different exemplars of the same pump. Yet, these subjects may not know the labels for specific shoes. While they might be able to recognize the style of some shoe designers, specific shoes models change frequently and those names are rarely used to identify shoes beyond the immediate shopping experience.

In Study 2B we will measure subjects' ability to name images of shoes and birds shown in the VET. For shoes, we expect that few subjects, if any, will provide a specific brand or model name, while for birds, we expect some will be able to name the birds by common species names. If these naming results are found, these two categories will serve as examples of categories in which knowledge of names might influence visual performance (birds), and in which semantic knowledge would not include individual names that would allow labeling in a visual task (shoes). We can then look at the correlation between visual and semantic performance for these two categories. If visual and semantic performance are correlated, even when subjects with high VET and SVET scores are unable to name shoes by model name, this would provide evidence that the VET-SVET relationship cannot be attributed to overlapping use of object names in both tasks.

Methods

Subjects. Two hundred and ten subjects (86 male; age: mean=22.49, SD=6.40) who participated in Study 2A completed a bird and shoe image naming task in addition to the other tasks in Study 2A.

Tasks.

VET, SVET, and experience questionnaire described in Study 2A. Study 2B looks at a subset of the data collected in Study 2A, which includes the VET, SVET, and self-report of category experience for shoes (women's high heels) and birds. Subjects completed the tasks in the following order so that the VET and image naming test were completed before the SVET to avoid subjects learning or remembering any object names from the SVET: category experience questionnaire, VET, CFMT, bird and shoe image naming, SVET. In study 2B will use the bird and shoe experience aggregate computed in Study 2A, and the theta scores for VETs and SVETs (SVET-Select trials).

Bird and shoe image naming. Images used for the naming test were grey-scale images of different birds and shoes used as foils in the VET-Bird and VET-Shoe, respectively. There were 18 trials for each category. Subjects completed the naming test as an online survey using REDCap electronic data capture survey tools (http://redcap.vanderbilt.edu; Harris et al., 2009) hosted by Vanderbilt University. Each image was presented with a blank textbox below it in which subjects were instructed to type the most specific name they had for each object or "NA" if they did not have a name for the object. All of the bird trials were shown on a single webpage first followed by all of the shoe trials on another page.

Results and discussion

Shoe naming. No subjects, even those with high VET-Shoe and SVET-Shoe scores, provided brand or designer names for shoes, as are used in the SVET, or specific shoe model names to name each shoe image. Instead, all shoe "name" responses were descriptions of the pictured shoe. These descriptions were almost always either very general category names (e.g., pump, stiletto, peep-toe, platform) or descriptions of the shoe's physical attributes including shape, color, fabric, and style (e.g., pointy-toe heel, ornate open-toed pumps, scalloped pumps, black bow, round-toe, beige suede). While some subjects included elaborate descriptions of shoes, suggesting an understanding of diagnostic shoe features, a specific subordinate-level name for the shoe was never given.

To determine if being able to give a detailed description of the shoe in terms of style, features, and type was related to shoe performance, we qualitatively scored subjects according to whether they provided a detailed description. For each subject we looked at all 18 shoe trials and assigned a single score based on their naming responses across trials. If a subject put "NA" or a single, general word (heel, shoe, pump) for more than half of the trials, they were scored as 0.

Subjects were scored as 1 if they listed a more detailed description on more than half of the trials (N=133, 101 female). This descriptive measure was significantly and positively correlated with all shoe measures: experience (shoe experience aggregate; $r(208)=0.37$, $p\leq 0.0001$), VET-Shoe ($r(208)=0.29$, $p\leq 0.0001$), and SVET-Shoe ($r(208)=0.28$, $p\leq 0.0001$). This suggests that the detail of shoe descriptions reflects common shoe experience. However these descriptions are longer and less unique than typical subordinate-level names, and so are less likely to be useful in the VET in the way that knowledge of individual object names might in another category.

Bird naming. To score the names subjects provided for each of the bird images, we counted any name that was a complete match or a more general partial match of the common species name of each bird as correct (e.g. for barn swallow: barn swallow, swallow, and swallow with a different sub-species descriptor, such as tree swallow or cliff swallow, were all counted as correct). More than half of subjects (N=112) did not correctly name any birds. Overall, scores ranged from 0-7 birds correctly named out of 18 bird trials (mean=0.84 birds correct, SD=1.27).

Correlation of naming with experience, VET, and SVET for birds. Performance on the bird naming task was significantly correlated with all other bird measures: self-report bird experience aggregate ($r(208)=0.46$, $p\leq .0001$), VET-Bird ($r(208)=0.37$, $p\leq .0001$), and SVET-Bird ($r(208)=0.42$, $p\leq .0001$). Figure 11 shows the scatterplots of bird naming score with VET-Bird and SVET-Bird accuracy. These plots illustrate that while many subjects could not name any birds, those who did correctly name even a few birds performed better on both the visual and semantic bird tests. These naming data provide further evidence of the convergent validity of our measures and suggest that with greater levels of bird experience, people typically acquire greater knowledge of subordinate-level bird names.

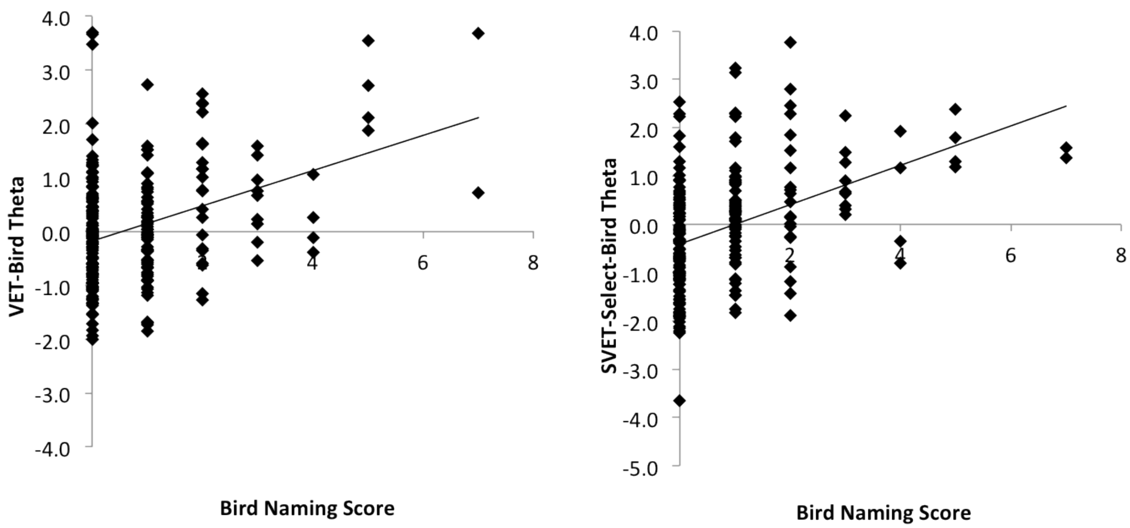


Figure 11. Scatterplots showing the relationship between the VET-Bird (left) and SVET-Bird (right) theta and bird naming score in Study 2B.

Correlation between VET and SVET with and without object names. A positive correlation between VET and SVET performance was found for a category for which expertise affords the ability to name objects (birds: $r(208)=0.35$, $p\leq0.0001$), but also for a category for which objects cannot be named at the subordinate-level by experts, as demonstrated by our naming survey (shoes: $r(208)=0.42$, $p\leq.0001$). While the bird and shoe domains may differ in many ways, it is worth noting that the correlation is numerically higher for shoes. This provides some evidence that subordinate-level names are not required for a category to demonstrate shared variance between visual and semantic performance.

To investigate if bird naming contributes to the correlation between VET and SVET for birds, we return to the partial correlations we calculated in Study 2A and remove factors that might contribute to the VET-SVET relationship (Table 11). In this dataset for birds, the partial correlation between VET and SVET while controlling for age, sex, *Gf*, general task performance (VET-Other and SVET-Other) and bird experience aggregate was $r=0.17$, $p=0.01$. The shared VET-SVET variance was 6.2% before entering bird experience into the model, and dropped to 2.8% after accounting for experience. Whatever is left is minimal, albeit statistically significant, and the present results suggest that it could be a contribution from naming. Adding the bird naming scores as another regressor indeed reduced the partial correlation of VET-Bird and SVET-Bird further (r=0.12, p=0.08, shared variance 1.4%), rendering the correlation no longer significant. This suggests that bird naming performance may contribute to both visual and semantic performance in a way that is independent from our bird experience measure and may be specific to how labels could be applied to birds in the VET in a manner that is not controlled by our experience measure.

Similarly, we also tested the contribution of shoe description to the VET-SVET relationship for shoes. In this dataset, the partial correlation between VET-Shoe and SVET-Shoe after controlling for age, sex, *Gf*, general task performance (VET-Other and SVET-Other) and shoe experience aggregate was r=0.18, p=0.01. The shared VET-SVET variance was 4.9% before entering shoe experience into the model, and dropped to 3.1% after accounting for experience. We then entered our binary score of subjects' shoe responses (1 for a detailed description of shoe style or 0 for a single word or no description) into the model. The shoe description score did not account for any additional VET-SVET variance (the shared variance remained at 3.1%, r=0.18, p=0.01). This suggests that for shoes, unlike for birds, naming, or describing shoe features verbally, did not contribute to VET performance.

Conclusions

In this dataset, shoes provide a really interesting example of common category experience. Clearly those with shoe experience have knowledge of shoes that can be measured in both semantic and visual domains. Performance on the visual (VET) and semantic (SVET) tasks is correlated, but we were able to demonstrate that shoe experts do not use the designer labels from the SVET, nor do they have specific subordinate-level model names, for the shoes in the

VET. Therefore, this appears to be a case where the VET-SVET correlation cannot be explained by the use of labels during the VET. There are likely other categories that would exhibit the same properties, perhaps watches, stand mixers, cell phones, or sunglasses. These categories might offer other opportunities to demonstrate shared variance between visual and semantic tasks that cannot be explained by labeling images in the visual task and could mainly reflect the role of experience for the independent acquisition of verbal and semantic knowledge.

In contrast, we found that for birds, bird naming performance contributed to the VET-SVET correlation in a way not captured by our self-report measure of bird experience or domain-general factors and abilities. For birds and other categories for which specific names may be strongly tied to object recognition performance, object naming may be used in both the VET and the SVET, and thus reflects a particular type of bird experience not completely captured in our bird experience measure.

Study 2C: Testing the SVET-Bird in Expert Birders

Overview

The goal of Study 2C was to provide further validation of the SVET by comparing performance of a sample of “experts” to a sample from the general population. We used online data collection to test birders with the SVET-Bird and VET-Bird and we also collected a more extensive questionnaire of experience that was specifically tailored to measure the extent of experience with birds. Our goal was to investigate if the SVET was capable of capturing individual differences in semantic knowledge even among high-level experts for a category, and to determine if we could validate these SVET results with other category-specific self-reported metrics of experience. We were also interested in whether we could replicate the relationship between visual and semantic performance found in Study 2A in experts.

Methods

Subjects. We recruited 64 subjects by email to participate in the online study. Email contact was facilitated by colleagues who had tested these specific birders previously or through acquaintances with personal relationships to regional birding clubs or specific birders. All subjects self-reported an interest and substantial experience in bird watching as a hobby or profession. The study was approved by the Vanderbilt IRB. As compensation, subjects who completed all parts of the study were entered in a lottery to win $50 with 1:10 odds of winning. Two additional subjects who began the study but chose not to complete all parts were not included in the analyses. One subject who completed all parts was excluded from the analyses because they misunderstood the VET instructions, resulting in below chance performance. The data reported here are for 63 subjects (29 male) aged 23-82 (mean=50.86, SD=15.16). All subjects reported that passerine birds were a type of bird they had experience recognizing. All subjects reported speaking English and living in the United States or Canada. As can be appreciated from Figure 12, our subjects reside in many different locations across North America.

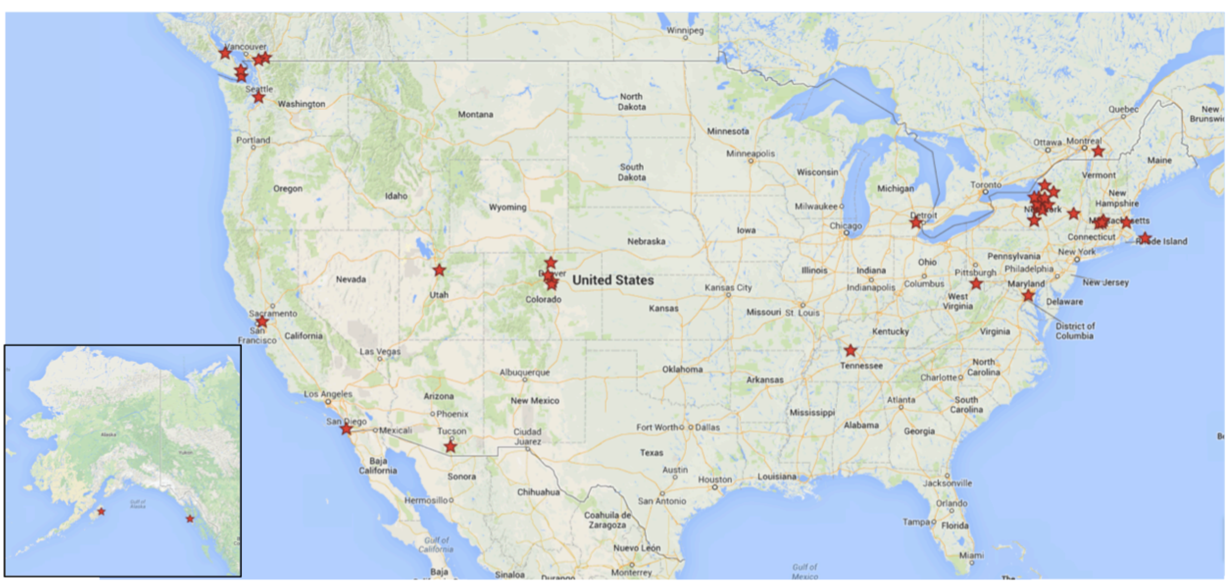


Figure 12. Map of North America, with Alaska shown in left inset, with red stars depicting the locations of birders who participated in Study 2C.

Procedure. Subjects completed four tasks in the following order: an extended bird experience questionnaire for birders, VET-Bird, SVET-Bird, and the bird image naming task. The VET-Bird and SVET-Bird were the same as Study 2A and the bird image naming task was the same as Study 2B. The birder experience questionnaire first asked the same questions used in Study 2A, four domain-general and seven bird specific; all were on a scale from 1-9 except for the duration of interest in birds for which they were asked to enter the number of years. To extend our experience questionnaire to measure differences in amount of bird experience between birders, we added 11 additional bird-specific questions, for example how frequently they go birding, how often they plan vacations around birding, if they belong to birding groups, and approximately how many different types of birds they have observed in person while birding during their lifetime (for full set of extended birder questions see Appendix B).

These tasks were completed using two online platforms: REDCap survey data collection tools were used for the experience questionnaire and bird naming test as in Studies 2A and 2B, and our own secure testing website was used for the VET and the SVET through a test-specific email link. The VET and SVET were presented one trial at time as in Study 2A, but subjects indicated their response by clicking on an image or name as in Study 1.

Results and discussion

Accuracy on VET-Bird, SVET-Bird, and bird naming. Our group of birders performed very well on the VET-Bird, SVET-Bird, and bird naming (see mean and SD in Table 13). Performance on catch trials was high (mean=0.99); no subjects were excluded due to catch trials. While performance was high, variability between subjects remained such that we were able to see individual differences between birders, although these differences were smaller than in Study 1 and Study 2A samples. Note that in this dataset, age and sex happen to be correlated

(r=0.31, p=0.02). This renders any relationships between performance and age or sex more difficult to interpret, so we will not consider these variables here. Because the birders in Study 2C represent a notably different population than the general population we tested in Study 2A, it would not be prudent to combine the data in an IRT model (without verifying if the test functions in the same qualitative manner in the two populations) and we did not have enough power with birders alone to conduct an IRT model to obtain theta scores. Therefore, we will use sum scores as our measure of performance for VET-Bird and SVET-Bird in our analyses.

Self-reported bird experience. Responses on the extended bird experience questionnaire for birders demonstrated high levels of self-reported experience with birds including many years birding, much time spent birding, looking at birds, and reading about birds (Table 13). Based on reports of the number of birds sighted during their lifetimes, frequency of birding including on bird-related trips and vacations, and involvement in birding organization and events, we can be fairly certain that we sampled a group of truly experienced birders.

The seven bird-specific questions used in Study 2A again demonstrated high internal consistency (average correlation between questions: r=0.34). We computed a bird-experience aggregate for each subject (basic bird aggregate) as the average of the Z-scored reports for each of the questions. The eleven extended bird questions specifically for birders also demonstrated good internal consistency (average correlation between questions: r=0.28) and were all well-correlated with the basic bird experience aggregate (mean r=0.40).

Table 13. Results from birders in Study 2C. The first column shows the mean and standard deviation for accuracy on each task, age, and self-reports of experience. The second and third columns give the correlations (r) between VET-Bird and SVET-Bird accuracy and each measure. Correlation coefficients shown in bolded red are statistically significant ($r_{Crit}(62)=.25$, $p<.05$).

	Mean (SD)	**VET-Bird**	**SVET-Bird**
VET-Bird	0.96 (0.08)	-	-
SVET-Bird	0.96 (0.07)	**0.43**	-
Bird Naming	14.95 (3.77)	**0.67**	**0.55**
General Experience (1-9):			
General Experience Aggregate	6.79 (1.09)	0.20	0.22
Bird-specific Experience (1-9):			
Overall Expertise	6.90 (1.36)	**0.48**	**0.45**
Importance	7.81 (1.27)	0.13	0.16
Duration Interest (years)	27.17 (17.95)	0.08	0.08
Visual Memory	7.00 (2.26)	0.10	0.12
Image Frequency	7.90 (1.59)	**0.32**	**0.30**
Text Frequency	7.79 (1.85)	**0.37**	**0.26**
Essay	6.21 (2.06)	0.24	**0.33**
Bird Experience Aggregate (Z-score)	-	**0.48**	**0.45**
Birder Extended Questions:			
Age Started (age)	20.33 (13.43)	**-0.44**	-0.14
Age Intense (age)	26.46 (13.53)	**-0.42**	-0.16
Birding Frequency (1-7)	5.86 (1.28)	**0.25**	**0.41**
Travel (1-5)	3.60 (1.36)	**0.26**	**0.41**
Vacation (1-6)	3.71 (1.68)	-0.13	**-0.26**
Log of Sightings (1-3)	2.49 (0.69)	**0.30**	**0.33**
Birds Sighted (number)	714.37 (811.89)	0.23	**0.32**
Local Expertise (1-7)	4.14 (1.29)	**0.32**	**0.28**
Periodicals (number)	1.46 (1.38)	0.25	-0.01
Organizations (number)	2.16 (1.61)	0.20	0.10
Events (1-7)	2.43 (1.28)	0.20	**0.28**
Extended Bird Experience Aggregate (Z-score)	-	**0.44**	**0.43**

We found that one question, age at which your interest in birds/birding became intense, was poorly correlated with the other birder questions, and that a similar question, age at which you became interested in birds/birding, was more consistent, so we removed the age intensity question from our later analyses and did not include it in the aggregate measure. Using the other 10 birder-specific questions and the 7 basic bird experience questions we computed an extended bird experience aggregate score for each subject as the average of the Z-scored reports for each of the 17 questions.

Interestingly, the domain-general experience aggregate score, calculated as the average of ratings on the four domain-general experience questions as in previous studies, was highly correlated with the extended bird experience aggregate (r=0.69, $p \leq 0.0001$)(Figure 13). This suggests that with an expert population, asking about their general experience with all objects may lead subjects to reflect on experience with their primary category of expertise, although this may not be a good measure of experience with birds given that the domain-general question does not predict VET or SVET scores.

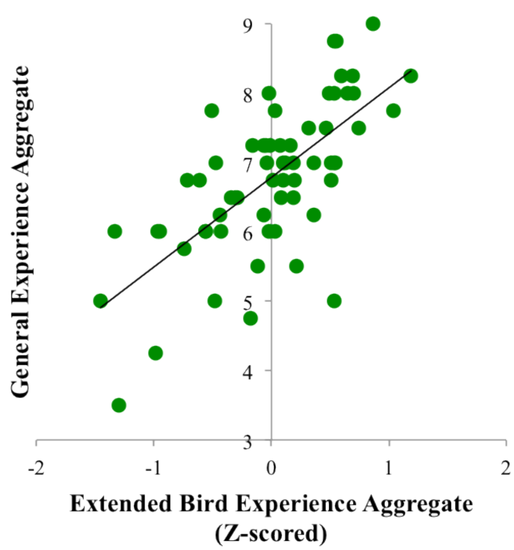


Figure 13. Scatterplot showing the relationship between general experience aggregate and extended bird experience aggregate for birders in Study 2C.

Correlations between SVET-Bird and VET-Bird and experience. Many but not all self-report questions of bird experience were correlated with SVET-Bird and VET-Bird performance (Table 13).

The domain-general experience aggregate was positively correlated with performance, but it was not significantly correlated with either SVET or VET performance. Several of the seven category-specific experience questions that we used in Study 2A were related to performance individually, especially questions about the frequency that someone read about or viewed picture of birds and the essay question, which was strongly correlated with SVET performance. The bird-specific experience aggregate for these seven questions demonstrated a high positive correlation with both SVET and VET. Interestingly, the category-specific overall expertise question ("Rate your expertise with XXX considering your interest in, years of exposure to, knowledge of, and familiarity with XXX") was the most highly correlated with both SVET-Bird and VET-Bird of any measure including the birder specific questions. This is consistent with previous findings (Gauthier et al., in press; McGugin, et al., 2012b) that this omnibus question of expertise is actually remarkably informative given its simplicity, here even in experts. Figure 14 shows SVET-Bird accuracy and expertise ratings for birds for both the birders in this study and the larger online sample in Study 1; considering both groups together (N=179), the correlation was very high, $r(177)=0.80$, $p<.0001$.

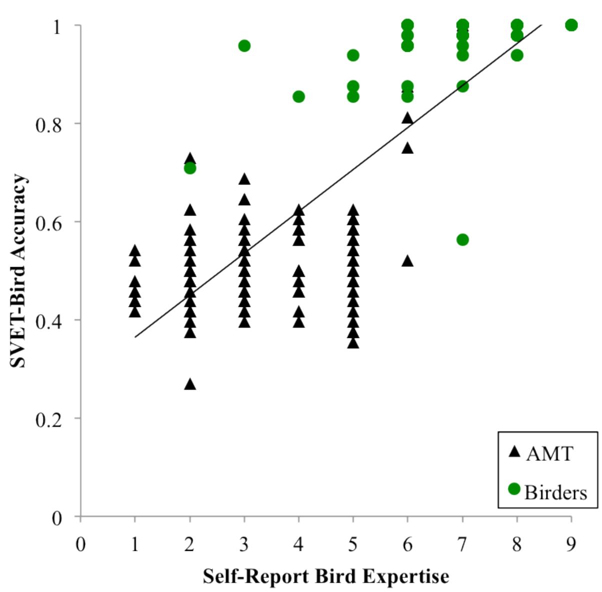


Figure 14. Scatterplot showing the relationship between SVET-Bird accuracy and self-reported bird expertise in AMT subjects from Study 1 (black triangles) and Birders from Study 2C (green circles).

Many of the extended bird experience questions that we added in this study just for birders were correlated with VET and SVET performance, especially how young someone was when they became interested in birds, how frequently they go birding or travel to see birds, and the estimated total number of bird species they have observed in their lifetime (Table 13). It is perhaps of interest that different experience questions seem to predict VET and SVET best: while age at which birding started predicts VET more than SVET (r=0.44 vs. 0.14), birding frequency predicts SVET more than VET (r=0.41 vs. 0.25). These patterns could be investigated in a larger study to test hypotheses regarding whether the same aspects of experience influence the acquisition of visual and semantic knowledge. The extended bird experience aggregate created from 17 bird-specific questions demonstrated a robust positive correlation with both VET and SVET. The significant correlations between these specific self-report measures of birder experience questions and our visual and semantic tests provide further validation that we are in fact measuring visual and semantic knowledge that would be acquired through extensive experience with that category.

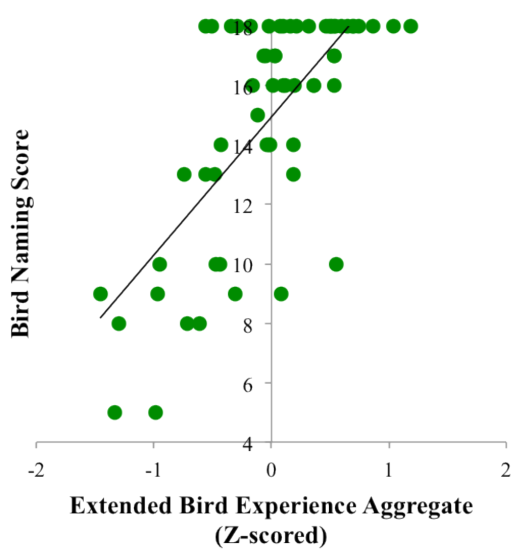


Figure 15. Scatterplot showing the relationship between bird naming score extended bird experience aggregate for birders in Study 2C.

Correlations between VET-Bird and SVET-Bird and bird naming. Results of the bird naming test can also be used as further validation of our measures, especially of the SVET. For birders in Study 2C, the number of birds correctly named was significantly positively correlated with both VET-Bird and SVET-Bird performance (Table 13). Bird naming was also strongly correlated with extended bird experience aggregate scores (r=0.72, p≤0.0001) (Figure 15). This demonstrates that for our sample of birders, greater levels of experience resulted in better bird naming performance, a skill that may also influence performance on the VET-Bird and SVET-Bird. To investigate the possible overlap between naming and these visual and semantic measures, we asked if VET and SVET contribute independently to naming performance for birds. We performed a simultaneous multiple regression predicting bird naming performance with VET-Bird and SVET-Bird performance (N=59 after removing three subjects who had very large externally studentized residuals (>2.5) in the correlation between VET and SVET; see blue X's in Figure 17). The results shown in Table 14 suggest that VET-Bird and SVET-Bird each make independent contributions to bird naming, and together account for 60.8% of the variance. Note that adding the bird aggregate experience scores to this model leads to an impressive R^2 adjusted of 71%.

Table 14. Results of a simultaneous multiple regression predicting bird naming performance with VET-Bird and SVET-Bird performance for birders (N=59) in Study 2C.

Model and predictor	β	*SE*	*t*	*p*
Bird Naming (R^2 adj = 60.8%)				
Intercept	-53.079	7.127	-7.450	≤0.0001
VET-Bird	25.412	7.942	3.200	0.002
SVET-Bird	44.637	8.824	5.060	≤ 0.0001

It is also useful to consider this expert data together on a continuum with non-expert data to observe the spread of performance from novice to expert. Image naming is a task that has been used previously (Barton et al., 2009) to quantify experience, and so it is important as we develop the non-visual SVET that we test its relationship with image naming as well. Figure 16 shows VET-Bird and SVET-Bird accuracy with bird naming scores for both the large sample of subjects collected at Vanderbilt University in Study 2A and the birders collected online in Study 2C. Considering both groups of subjects together (N=275), the correlations between each measure and naming were high (VET-Bird and bird naming: $r(273)=0.74$, $p<.0001$; SVET-Bird and bird naming: $r(273)=0.93$, $p<.0001$), although the strength of these relationships may be driven primarily by large group differences between those who could not name any birds and those who could name birds. Nevertheless, it is interesting to observe the variability in performance for both a general and an expert population. While we do observe some variability within experts on all measures, it clear that if we were interested in refining a measure specifically to investigate individual differences between experts we would need to extend the VET, SVET, and naming tests to include a larger set of more difficult trials to reduce ceiling effects and better discriminate among experts. A larger sample of bird experts would also allow us to compute an IRT measurement model, which should increase the information provided by these tests.

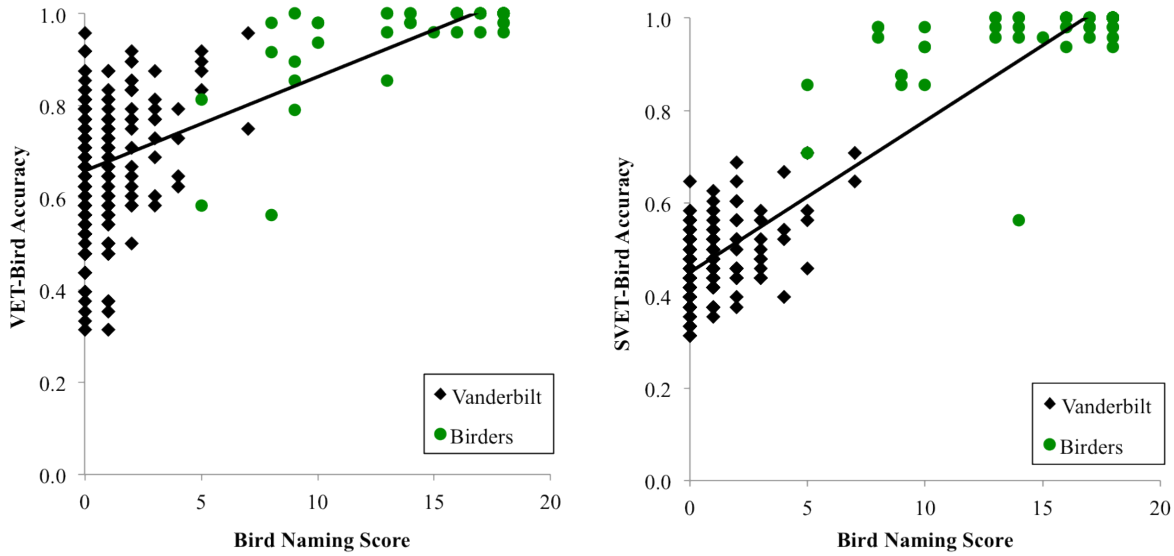


Figure 16. Scatterplots showing the relationship between the VET-Bird (left) and SVET-Bird (right) accuracy and bird naming score in Vanderbilt subjects from Study 2B (black diamonds) and Birders from Study 2C (green circles).

Correlations between VET-Bird and SVET-Bird. In our expert birder sample, performance on the SVET-Bird and VET-Bird was again positively correlated, $r(61)=0.43$, $p=0.005$, replicating the results of Study 2A in an expert population and allowing for the possibility that common variance between VET and SVET is the result of common category experience. We can also consider this relationship for non-expert subjects in Study 2A together with birders in Study 2C to gain an understanding of the relationship across greater variability in experience. Considering both groups together (N=278), the SVET-VET correlation for birds was high, $r(276)=0.70$, $p<.0001$ (Figure 17).

However, to compare with the results we observed in Study 2A in which experience accounted for shared VET-SVET variance, we were interested in the contribution of experience to the VET-SVET relationship in the birder sample. For this analysis, we began by looking at the correlation between VET-Bird and SVET-Bird ($r=0.62$, $p\leq0.0001$) after removing three subjects (N=59) who had very large externally studentized residuals (>2.5) (subjects marked with blue X's in Figure 17). We then performed a simultaneous multiple regression predicting VET-Bird with SVET-Bird and extended bird experience aggregate. The partial correlation between VET-Bird and SVET-Bird after partialing out bird experience and age (we did not include sex to avoid multicolinearity with age) was reduced, but still quite sizeable (r=0.56, p≤0.0001). In our birder dataset, unlike in Study 2A, we do not have measures of domain-general abilities or task performance for other categories, so we cannot say if the shared VET-SVET variance after removing experience is the result of domain-general or category-specific abilities. In novices in Study 2A we observed with that removing experience rendered the VET-SVET partial correlation non-significant for birds (from r=0.17 to r=0.08; Table 11), suggesting that for novices experience carries much of the variability in performance. However, for experts, while experience contributes somewhat to the VET-SVET relationship, much more variance still remains after partialing out experience. This result with expert birders is analogous to what has been suggested for face recognition performance: experience may account for less variability in performance in a sample/category for which experience is generally high (Gauthier et al., in press). However, adding bird naming to the model further reduced the correlation between SVET and VET to r=0.30, p=0.03), suggesting that subjects who could name birds better (and had better SVET scores) may have also used a naming strategy on the VET.

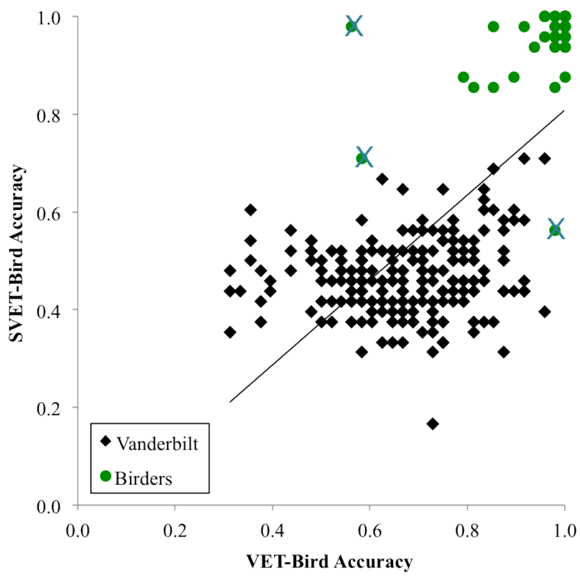


Figure 17. Scatterplots showing the relationship between SVET-Bird and VET-Bird accuracy for Vanderbilt subjects from Study 2A (black diamonds) and Birders from Study 2C (green circles). Three birders marked with blue X's denote subjects with very high externally studentized residuals (>2.5).

Conclusions

In Study 2C we provided validation of the SVET in an expert population by testing the SVET-Bird in a sample of experienced birders together with VET-Bird, a test of bird naming, and an extended measure of bird experience. We demonstrated that while birders' performance was high on all of the bird measures, there was sufficient individual variability that allowed us to observe relationships between performance on each test and experience. For future work testing performance with experts, we would recommend creating extended versions of VET, SVET, and the naming test with additional difficult trials to reduce ceiling effects for experts and to better resolve individual differences at the high end of performance. However, using the standard VET-Bird, SVET-Bird, and bird naming test we demonstrated high correlations between all bird measures and between each of those measures and the extended measure of bird experience, providing further evidence that the SVET and the correlation between VET and SVET reliably capture real-world experience with a specific object category. We also demonstrated that while experience accounts for some shared variance between VET-Bird and SVET-Bird for birders, it may account for less variance than in a novice population where experience varies much more. This study provides further evidence that the SVET is a useful tool that can be used together with a variety of other visual measures and that can capture expertise at high levels of performance.

Chapter 4 – Study 3

Putting the SVET to the Test – Can Semantic Knowledge Predict Lateralized Object Recognition Performance?

Overview

In Studies 1 and 2 we created the SVET, and validated the test by looking at its relationship with measures of experience, visual performance, and domain-general abilities. We also looked at the validity of one category tested with a SVET, birds, by testing a sample of bird experts. We found that the SVET and VET both measure dimensions that can discriminate experts from the general population, and provide information that discriminates between experts. Thus, we have presented the SVET for eight different object categories and can make a case that it is a reliable and valid measure of category-specific semantic knowledge that can be used to help understand other perceptual and cognitive phenomena.

In Study 3 we present an example of how the SVET might be used in combination with other measures to answer questions regarding how visual and semantic object information is represented in the brain. More specifically, we will use the SVET in combination with the VET to address the lateralization of expertise effects in the brain.

Introduction to the question

While visual processes are a large part of object recognition, this process also interacts with an array of other systems, such as cognitive, emotional, and linguistic systems, that can influence perception. In the previous studies we observed a positive relationship between category experience and semantic performance measured with the SVET. Semantic knowledge acquired for familiar objects can affect perception of those objects, and has been shown to change lateralization of visual processing (Curby, Hayward, & Gauthier, 2004). In Study 3 we will investigate if the degree and direction of lateralization for visual recognition of an object category, in particular performance in the left hemisphere, can be predicted by semantic knowledge of that category measured by the SVET.

Associating non-visual semantic features with scenes and objects, especially relevant and salient descriptions, can influence visual object recognition (Wiseman, MacLeod, & Lootsteen, 1985) and discrimination (Gauthier, James, Curby, & Tarr, 2003) by making objects with associated semantic knowledge easier to remember and recognize. Associations between shape information and non-visual semantic content can be created rapidly, creating multimodal representations in the brain that are automatically activated by the visual presentation of an object (James & Gauthier, 2003; 2004). While in some studies researchers have taught subjects to associate specific semantic knowledge with objects, the SVET can be used as a measure of semantic knowledge of an object category learned through real-world experience, and used to study the influence of this knowledge on perception.

One hypothesis is that greater semantic knowledge for a category will increase the recruitment of left hemisphere brain areas even during a visual task. From early lesion studies to modern fMRI studies, the left hemisphere has been found to be responsible for the majority of linguistic and verbal processing in the brain (Kann, 1950; Wagner, Desmond, Demb, Glover, & Gabrieli, 1997). In particular, the left inferior frontal gyrus (LIFG) is activated by semantic processing not only for words, but also for images and objects (Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996; Vuilleumier, Henson, Driver, & Dolan, 2002; Wagner et al., 1997). It also been theorized that different subsystems underlie object processing in each hemisphere, so that more abstract or conceptual visual processing is left-lateralized, while fine visual discrimination is more right-lateralized (Marsolek, 1999).

In this study, we will use the SVET to quantify semantic knowledge of a category acquired from real-world experience to predict the degree of visual processing lateralization for that category. There is evidence suggesting that semantic knowledge affects the laterality of object perception in the brain. A behavioral study with novel objects found that when subjects were trained to associate specific adjectives (e.g., fast, hollow, strong) with particular novel objects, recognition of those objects in a sequential matching task was facilitated for processing in the left versus right hemisphere (Curby et al., 2004). These results seemed to be due to associating semantic information with the objects, because subjects in a different training condition who performed visual similarity judgments during training rather than learning semantic information did not demonstrate any difference in performance between the right and left hemispheres.

A surprising difference in laterality was also observed in an fMRI study of perceptual expertise with cars and planes in face selective areas. The response to cars in the fusiform face area (FFA) was correlated with behavioral car expertise bilaterally, while the response to planes in FFA was only related to behavioral expertise for planes in the right hemisphere (McGugin, Gatenby, Gore, & Gauthier, 2012a). This, however, may not represent a true categorical difference. In this study, subjects were recruited to maximize differences in car expertise. About half the subjects self-reported expertise with cars, but none reported expertise with airplanes. However, despite the fact that only three of the subjects reported any above-average experience with planes, performance on visual tasks with planes and with cars were correlated. According to the results from the previous chapters of this dissertation, people who self-report as experts in a domain are likely to have more semantic knowledge in that domain. But because most of these subjects did not report any special knowledge of planes, it is unlikely that they had much semantic knowledge for that category. This led McGugin et al. (2012a) to suggest that the difference in laterality between the categories might be the result of semantic knowledge; high semantic knowledge for cars recruits both left and right hemispheres for object recognition, while lack of semantic knowledge for planes leads to activity that is correlated with visual performance in the right hemisphere only.

In Study 3, we will use the SVET as a measure of semantic knowledge for specific object categories to investigate the effect of semantic knowledge on individual differences in lateralized object processing. We predict that subjects with high SVET scores in a domain may have access to expert representations in both hemispheres when they recognize objects in that domain, whereas subjects with low SVET scores may mainly perform visual judgments based on right-hemisphere representations. To test this hypothesis, we had subjects perform a lateralized visual matching task with four different object categories. They also performed VETs and SVETs for each category. We will essentially attempt to predict VET performance based on the matching scores in each hemisphere, and ask if this is modulated by SVET scores.

We tested four objects categories selected from the set of eight tested in the previous studies: cars, planes, shoes, and birds. These categories were selected for several reasons. First, we chose two male-interest and two-female interest categories based on the results of the previous studies, not because we were interested in sex effects *per se*, but because we wanted to sample a range of individual experience for each of these categories, and experience level for many categories corresponds with sex. We also selected these four categories because we found robust effects of a category-specific VET-SVET relationship for these categories in Study 2A. We were also particularly interested in cars and planes because they were used in the previous fMRI study (McGugin et al., 2012a). Lastly, we chose to test shoes because we expected that the pattern of results might differ from the other categories based on the results of Study 2B, which showed that while VET-Shoe and SVET-Shoe are robustly correlated, shoe naming does not contribute to the relationship. While our prediction for cars, planes, and birds is greater recruitment of left hemisphere with increased semantic knowledge, this effect may not be observed for shoes, for which visual and semantic knowledge from common experience do not seem to overlap.

Methods

Subjects. One hundred and twenty subjects were recruited from the Vanderbilt University and Nashville, TN community; they gave informed consent and received course credit or monetary compensation for participation. The study was approved by the Vanderbilt IRB. All subjects reported normal or corrected to normal visual acuity, were native English-speakers, and had lived in the United States at least 10 years, except for one subject who had lived in the U.S. for four years. Data from one subject were excluded for below chance (.33) performance on three SVETs and data from one subject were left in the dataset but they completed the lateralized matching task on a different day. Thus, data are reported here for 119 subjects (47 male) aged 18-46 (mean=21.54, SD=3.55).

Equipment. The experiment was conducted in the laboratory on Apple Mac Minis (OSX 10.9.2, 2Ghz Intel core 2 duo) with 21.5-inch LCD monitors (1920x1080 resolution) using MATLAB R2009b (Mathworks, Natick, MA, USA) and Psychtoolbox (http://psychtoolbox.org; Brainard, 1997). The experience questionnaire was completed using REDCap electronic data capture survey tools (http://redcap.vanderbilt.edu; Harris et al., 2009) hosted by Vanderbilt University. Subjects were seated approximately 60 cm from the monitor and used a chin rest during the lateralized matching task to maintain this distance.

Tasks. Subjects performed four tasks in the following order: experience questionnaire, lateralized matching, VET, and SVET. Each task tested four object categories with trials always blocked by category, and blocks occurring in the order: cars, birds, shoes, and planes.

Experience questionnaire. Subjects completed the same questionnaire as in Study 2A with four domain-general experience questions and seven category-specific questions for each category.

VET and SVET. The VETs and SVETs for cars, birds, shoes and planes were the same as those used in Study 2A and were administered in the same fashion.

Lateralized matching. The lateralized matching task was a sequential same-different object recognition task in which the second image was presented in the periphery to the left or right of fixation to selectively recruit one hemisphere of the brain in the perceptual judgment.

Stimuli. Stimuli were grey-scale images of cars, birds, shoes, and planes. Cars, birds, and planes were shown on natural backgrounds as they are commonly seen, and shoes were shown on white backgrounds as they might be seen in magazines or online. Images were padded by flanking solid-color rectangles, if necessary, to make them square; flankers were a consistent color (grey, black, or white) for each category. These images were selected to be similar to the images used in the VET and followed the same criteria used for the VET and SVET in Studies 1 and 2 (e.g. found in North America, car models from 2000 to present, male passerine birds, etc). While some of the same objects (e.g. Honda Civic, blue jay) appear in both tasks, no exact images are used in both tasks. Each trial consisted of a pair of images shown sequentially. The first image was presented in the center of the screen and the second image was presented 200 pixels to the left or right of center. The first image was 180 x 180 pixels and subtended a visual angle of approximately 4.8 x 4.8 degrees. The second image was 100 x 100 pixels and subtended a visual angle of approximately 2.6 x 2.6 degrees.

Test and trial structure. The block of trials for each category consisted of 110 trials: 100 matching trials and 10 fixation catch trials. Of the 100 matching trials, 50 were same trials and 50 were different trials. Same trials presented the same object in each image, but only 10 same trials in each block showed an identical match; we will refer to these as 'same: exact' trials (Figure 18). The other 40 same trials in each block were 'same: different example' trials which presented two different examples of the same object but might differ in non-diagnostic features of the object or image, for example, color, model year, viewpoint, or background. Each of the 50 same trials presented a unique object. Subjects were carefully instructed that same responses indicated the same object (car, plane, or shoe model, or bird species) regardless of whether a same or different example was presented in the second image. On different trials there were two different objects in each image. The images used in different trials were not used in same trials and were only used once each, but object identity sometimes overlapped with an object presented in a same trial up to twice per block. Apart from Same vs. Different, none of the other distinctions will be explicitly analyzed; trials were merely selected to cover a range of difficulty, so that the test as a whole would discriminate domain-specific performance across the entire spectrum of ability.

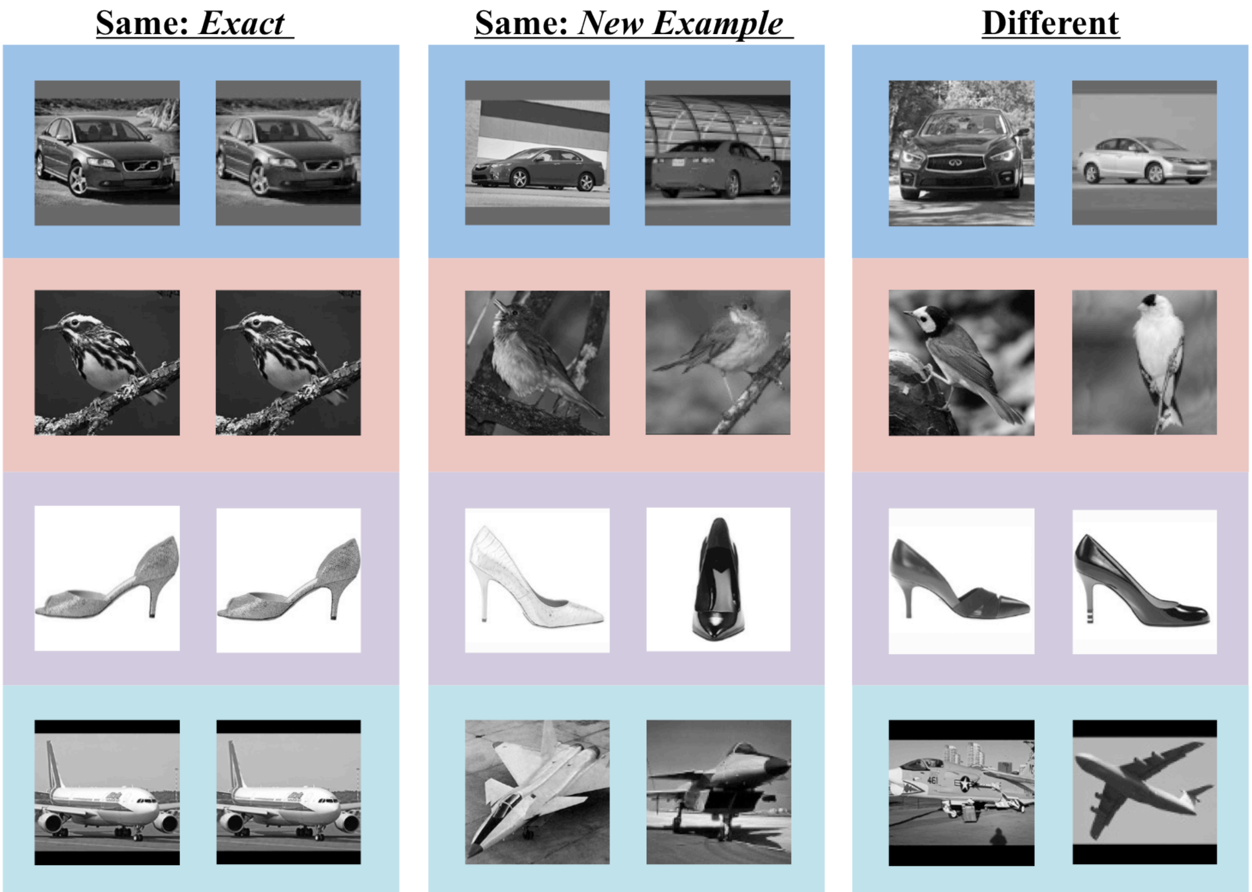


Figure 18. Examples of each trial type, same: exact, same: new example, and different, used in the lateralized matching task in Study 3 for cars, birds, shoes, and planes, shown in each row.

Trials. Each trial began with a red fixation cross for 1,000ms, followed by the first image presented centrally for 1,500ms, a black cross shown for 500ms, the second image presented to the left or right of fixation for 200ms, and then a mask (specific to the test category) shown for 250ms (Figure 19). Subjects responded with a keyboard button press ('1' for same or '2' for different on the keypad) and could make a response as soon as the second object appeared. The screen remained blank until a response was made and then advanced to the next trial after a 500ms ITI.

Procedure. Subjects were instructed to fixate the center of the screen at all times throughout the trial even when images appeared off center, which is critical for lateralized processing. We used fixation catch trials consisting of a perceptual detection task in the center of the screen to ensure fixation. On fixation catch trials (10 trials, presented at pseudo-random times during the block with greater frequency at the start of each block), the fixation cross and first image appeared, as on a matching trial, but then there was a subtle shift in the black fixation cross such that the horizontal line shifted 5 pixels to the left or right (five left and five right trials per block) for 150ms followed by a blank screen. On fixation catch trials, subjects were required to make a different keypress response indicating which side of the horizontal line in the cross become longer ('Z' for left, 'X' for right).

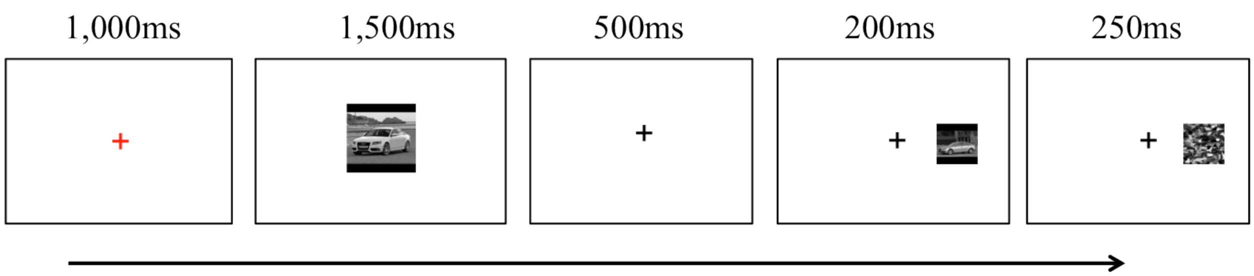


Figure 19. The trial structure of a single trial in the lateralized matching task in Study 3. Each trial began with a red fixation cross, followed by a first image presented centrally, a brief cross, a second image presented briefly to the left or right (right in this example) of fixation and then a mask. Presentation times are shown for each part of the trial.

Results and discussion

Task performance.

Experience ratings. Subjects reported a range of general and category-specific experience for each category comparable to responses in Study 2A. As in the previous study, we looked at the consistency of the set of questions for each category and of the general experience question using the average correlation between all questions and Cronbach's alpha. The four general questions were the least correlated with one another (mean r=0.32, α=0.64). An aggregate score computed as the average of the four general questions had median=5.50 and SD=1.16. The seven category-specific questions for each category demonstrated higher correlations and excellent internal consistency (Car: mean r=0.76, α=0.93; Plane: mean r=0.61, α=0.89; Shoe: mean *r*=0.73, α=0.95; Bird: mean *r*=0.53, α=0.87). Because the questions were highly correlated, we created a category-specific aggregate score for each subject (average of the 7 questions). Figure 20 shows the aggregate of category-specific experience for each category. As in the previous study, subjects reported greater experience with cars and shoes, although this experience was also more variable than that for planes and birds.

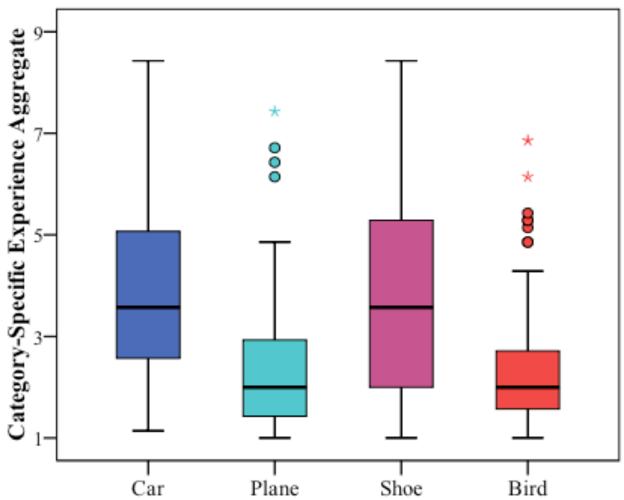


Figure 20. Boxplots showing the aggregate of self-reported category-specific experience ratings for four categories in Study 3.

VET and SVET performance. Accuracy computed as the sum score on the VET (VET-Car: mean=0.59, SD=0.15; VET-Plane: mean=0.65, SD=0.13; VET-Shoe: mean=0.72, SD=0.11; VET-Bird: mean=0.64, SD=0.13) and SVET (SVET-Car: mean=0.64, SD=0.16; SVET-Plane: mean=0.45, SD=0.09; SVET-Shoe: mean=0.57, SD=0.16; SVET-Bird: mean=0.47, SD=0.07) was above chance and within the expected range based on previous studies. No subjects were excluded due to catch trial performance on the VET or SVET. One subject was excluded for below chance (.33) performance on three SVETs.

We used theta VET and SVET scores, as in Study 2A. For VET, we computed theta scores from a two-parameter IRT model based on all the data we had for these categories to provide a better measurement model. Therefore, the scores for individuals in Study 3 were computed in separate models for each category using these subjects (N=119) together with those from Study 2A (N=213; Total N=332). Likewise, for SVET performance we computed theta using only the SVET-Select trials identified in Study 2A to produce a unidimensional measure and the factor most correlated with the VET. SVET theta scores for individuals in Study 3 were computed from a two-parameter IRT model for each category together with subjects from Study 1 (N=116), Study 2A (N=213), and Study 3 (N=119; Total N=448). Figure 21 shows the distributions of these theta scores for each VET and SVET for the Study 3 subjects.

Table 15 shows the correlations between each of the VETs and SVETs. Overall, VETs were more correlated with one another (mean r=0.26; Panel A) than SVETs (mean r=0.04; Panel C), a pattern also observed in Study 2A, and which is likely due to the common contribution of domain-general visual ability on all VETs. The VET-SVET correlation was also again observed to be stronger within category (mean r=0.33) than between category (VET-SVET for different category, mean r=0.04; two-sided Fisher's Z test: Z=4.39, $p \leq 0.0001$; Panel B), producing evidence in another sample that performance for both tests is category-specific.

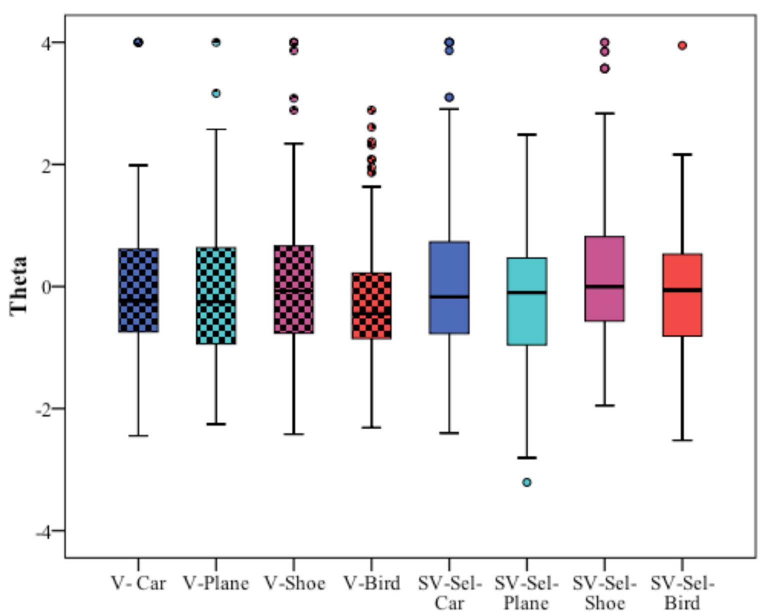


Figure 21. Boxplots showing the theta scores for subjects on each VET (V-) and SVET-Select (SV-Sel) for each category in Study 3.

Lateralized matching performance.

Most subjects performed well on the fixation catch trials (mean=0.93, SD=0.09). Despite instructions to report which side of the cross became longer, some subjects instead reported which direction the horizontal line appeared to move (the opposite response). Subjects with very low accuracy on fixation (<0.2, many with perfectly negative accuracy) were considered to have swapped responses and their scores were inverted (e.g. 0.1 to 0.9 accuracy). Three subjects performed nearly at chance (0.5) on fixation trials, but further inspection of their data revealed that they had high accuracy (≥0.7) in at least one block and switched the response mapping part-way through the task in other blocks, resulting in poor average performance. Therefore, no subject was excluded for performance on fixation catch trials, especially since performance on the fixation task was not correlated with performance on the matching tasks.

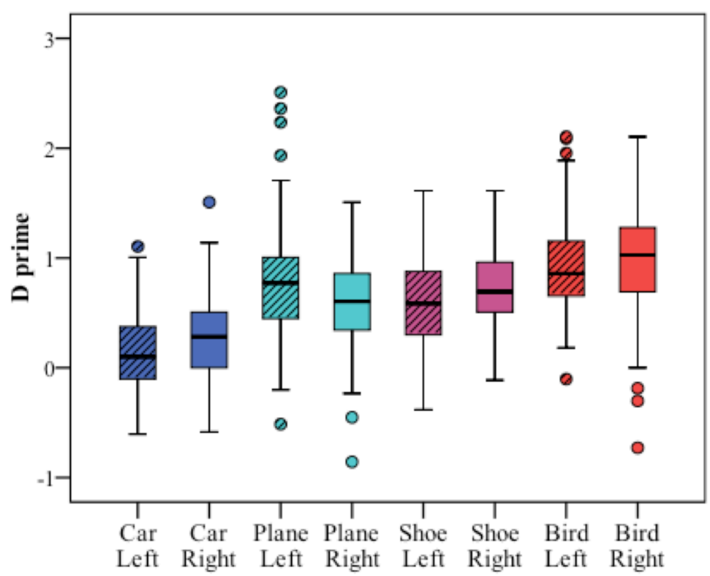


Figure 22. Boxplots showing performance on the lateralized matching task in Study 3 as d prime for each category and hemisphere (e.g. 'Car Left' shows performance when cars were presented to the right of fixation such that visual processing would be in the left hemisphere.)

The lateralized matching task was challenging given that the second image in each trial was presented very briefly in the periphery, and most of the same trials required complex object recognition beyond image matching. However, most subjects performed above chance on the task for all categories (mean sum score=0.62, SD=0.49). Note that for each category, subjects performed very well on the ten exact same trials, suggesting that it was the trial difficulty in terms of object similarity, not the task itself, that was challenging (mean sum score on exact same trials =0.86-0.92 on each category).

To remove the influence of any individual bias in performance on the matching task, we used signal detection theory to compute d prime (d') (Green & Swets, 1966) for presentation in each visual field separately (Figure 22). We computed the correlations between left and right hemisphere d' (Table 15, Panel E). We did not observe a within versus between category difference in the correlation between hemispheres (within category mean r=0.30, between category mean r=0.23, two-sided Fisher's Z test: Z=0.62, p=0.54). Between categories there was not a greater relationship between matching performance in the left hemisphere (mean r=0.22; Table 15, Panel D) versus the right hemisphere (mean r=0.26; Panel F; two-sided Fisher's Z test: Z=-1.09, p=0.28).

Table 15. Correlations (r) between performance on each of the measures in Study 3: VET, SVET, and lateralized matching, which is divided by left hemisphere and right hemisphere trials. Each task is show for cars, planes, shoes and birds. Correlations within category but cross task (VET-SVET) or hemisphere (left-right) are outlined on the diagonals of Panel B and D. Values shown in bolded red are statistically significant ($r_{Crit}(117)$=.177, p<.05).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		A. VETs															
1	VET-Car	-															
2	VET-Plane	**0.28**	-														
3	VET-Shoe	0.13	**0.26**	-													
4	VET-Bird	0.16	**0.45**	**0.28**	-												
		B. SVET and VET				C. SVETs											
5	SVET-Car	**0.51**	0.09	**-0.25**	-0.14	-											
6	SVET-Plane	**0.28**	**0.32**	-0.05	0.11	0.32	-										
7	SVET-Shoe	0.15	-0.05	**0.35**	-0.13	-0.06	-0.12	-									
8	SVET-Bird	0.05	0.16	**0.21**	0.15	-0.05	0.15	-0.03	-								
										D. Left Hemisphere							
9	Left Car	0.18	0.00	-0.10	0.00	**0.19**	0.09	0.02	-0.01	-							
10	Left Plane	**0.22**	**0.26**	-0.03	0.15	**0.21**	**0.26**	-0.09	0.12	**0.36**	-						
11	Left Shoe	0.07	-0.01	**0.21**	0.03	-0.01	-0.03	**0.26**	0.04	0.08	0.17	-					
12	Left Bird	0.14	**0.24**	0.08	**0.19**	-0.09	0.07	-0.12	0.00	0.15	**0.32**	**0.27**	-				
										E. Cross Hemisphere				F. Right Hemisphere			
13	Right Car	**0.29**	0.05	0.04	-0.06	**0.26**	**0.27**	-0.04	0.04	**0.28**	**0.37**	**0.21**	0.15	-			
14	Right Plane	0.15	**0.20**	0.03	**0.24**	-0.01	0.10	-0.12	0.07	**0.19**	**0.33**	**0.23**	**0.30**	**0.28**	-		
15	Right Shoe	0.09	0.17	0.16	0.11	0.01	0.06	**0.19**	0.07	0.14	**0.32**	**0.34**	**0.23**	**0.31**	**0.37**	-	
16	Right Bird	-0.10	0.01	0.14	**0.18**	-0.10	0.00	-0.02	0.05	0.02	**0.26**	**0.38**	**0.27**	0.14	**0.26**	**0.19**	-

Contribution of SVET to VET performance in each hemisphere. To determine if SVET performance moderated the hemispheric contribution to VET performance, we performed a multi-step regression analysis.

First, for each category, we tested a model that predicted VET performance using SVET scores, the two lateralized matching scores, and the interaction between lateralized matching and SVET (all predictors entered simultaneously - Table 16). Most interestingly, the SVET showed a significant interaction with lateralized matching to predict VET scores: SVET*left hemisphere matching for cars and planes and SVET*right hemisphere for shoes and nearly (p=0.059) for birds. This reveals that VET performance, a task for which the images can be fixated, was most similar to matching in one hemisphere, as a function of semantic knowledge.

To further investigate these interactions, we computed how VET scores residualized for SVET (therefore a relatively pure index of visual performance when tested foveally) were predicted by matching performance in each hemisphere for subjects with low vs. high SVET scores (according to a median split). In these datasets, the SVET contributed to VET significantly for cars, planes, and shoes (car: R^2=26.0%, $p\leq$0.0001; plane: R^2=10.4%, p=0.0003; shoe: R^2=12.2%, $p\leq$0.0001; bird: R^2=2.1%, p=0.12). Therefore, these represent the different magnitudes of variance accounted for by regressing out the SVET from VET scores (twice as much for cars as for planes or shoes, and very little for birds).

Using these VET residuals, we then proceeded to test the hypothesis that McGugin et al. (2012a) advanced to explain their right-lateralized expertise effects for planes, but bilateral expertise effects for cars: that semantic knowledge would play a role in extending expertise effects to the left hemisphere. This test will assess whether visual performance in subjects with different levels of semantic knowledge may be the result of unequal contributions from the two hemispheres.

Table 16 Multiple regressions for each category predicting VET with SVET, right hemisphere matching, left hemisphere matching, and the interaction of SVET with left and right hemisphere matching.

CAR

Model and predictor	β	*SE*	*t*	*p*
VET Theta (R^2 adjusted = 29.7%)				
Intercept	-0.235	0.117	-2.010	0.046
SVET Select Theta	0.322	0.093	3.460	0.001
Right Hem Matching Dpr	0.628	0.275	2.290	0.024
Left Hem Matching Dpr	0.181	0.304	0.595	0.553
SVET Select*Right Hem Matching	0.388	0.230	1.690	0.095
SVET Select*Left Hem Matching	**-0.494**	**0.197**	**-2.500**	**0.014**

PLANE

Model and predictor	β	*SE*	*t*	*p*
VET Theta (R^2 adjusted = 15.3%)				
Intercept	-0.496	0.220	-2.260	0.026
SVET Select Theta	0.041	0.202	0.203	0.839
Right Hem Matching Dpr	0.141	0.297	0.475	0.636
Left Hem Matching Dpr	0.405	0.208	1.950	0.054
SVET Select*Right Hem Matching	-0.220	0.200	-1.100	0.272
SVET Select*Left Hem Matching	**0.449**	**0.199**	**2.260**	**0.026**

SHOE

Model and predictor	β	*SE*	*t*	*p*
VET Theta (R^2 adjusted = 16.7%)				
Intercept	-0.520	0.268	-1.940	0.055
SVET Select Theta	0.908	0.227	4.000	0.000
Right Hem Matching Dpr	0.518	0.338	1.530	0.128
Left Hem Matching Dpr	0.357	0.287	1.240	0.217
SVET Select*Right Hem Matching	**-0.696**	**0.266**	**-2.620**	**0.010**
SVET Select*Left Hem Matching	-0.093	0.185	-0.502	0.617

BIRD

Model and predictor	β	*SE*	*t*	*p*
VET Theta (R^2 adjusted = 6.2%)				
Intercept	-0.826	0.274	-3.010	0.003
SVET Select Theta	0.298	0.252	1.180	0.239
Right Hem Matching Dpr	0.248	0.203	1.220	0.224
Left Hem Matching Dpr	0.398	0.236	1.690	0.094
SVET Select*Right Hem Matching	-0.390	0.204	-1.910	0.059
SVET Select*Left Hem Matching	0.234	0.237	0.988	0.325

Table 17 shows the results of these regressions. Italics indicate hemispheres for which we previously observed a significant contribution of SVET*hemisphere in VET. We will focus our discussion on these results, as our previous analyses led us to believe that SVET scores could predict a meaningful difference in visual performance in those hemispheres. We will not compare low/high SVET effects for the non-italicized results as we did not previously obtain a significant interaction of SVET*hemisphere. For instance, while for cars high SVET subjects show a contribution of the right hemisphere matching while low SVET subjects do not, this interaction was not quite significant (see Table 16).

Table 17 Multiple regressions for each category predicting VET(SVET) with left and right hemisphere matching for subjects with low SVET scores (left) and high SVET scores (right). Italics indicate hemispheres for which we previously observed a significant contribution of SVET*hemisphere in VET.

LOW SVET					
	Model and predictor	β	*SE*	*t*	*p*
CAR	VETresid(SVET) (R^2 adj = 9.5%)				
	Intercept	-0.065	0.139	-0.466	0.643
	Right Hem Matching Dpr	0.405	0.370	1.090	0.278
	Left Hem Matching Dpr	*0.951*	*0.352*	*2.700*	*0.009*
PLANE	VETresid(SVET) (R^2 adj = 2.0%)				
	Intercept	-0.239	0.258	-0.927	0.358
	Right Hem Matching Dpr	0.544	0.309	1.760	0.083
	Left Hem Matching Dpr	*-0.202*	*0.291*	*-0.695*	*0.490*
SHOE	VETresid(SVET) (R^2 adj = 4.2%)				
	Intercept	-0.744	0.331	-2.250	0.028
	Right Hem Matching Dpr	*0.687*	*0.444*	*1.550*	*0.127*
	Left Hem Matching Dpr	0.324	0.345	0.940	0.351
BIRD	VETresid(SVET) (R^2 adj = 14.2%)				
	Intercept	-0.855	0.280	-3.050	0.004
	Right Hem Matching Dpr	0.585	0.209	2.800	0.007
	Left Hem Matching Dpr	0.302	0.243	1.240	0.219

HIGH SVET					
	Model and predictor	β	*SE*	*t*	*p*
CAR	VETresid(SVET) (R^2 adj = 8.5%)				
	Intercept	-0.305	0.180	-1.690	0.096
	Right Hem Matching Dpr	1.024	0.380	2.700	0.009
	Left Hem Matching Dpr	*-0.755*	*0.450*	*-1.680*	*0.098*
PLANE	VETresid(SVET) (R^2 adj = 8.1%)				
	Intercept	-0.582	0.335	-1.740	0.087
	Right Hem Matching Dpr	-0.036	0.454	-0.079	0.937
	Left Hem Matching Dpr	*0.721*	*0.290*	*2.480*	*0.016*
SHOE	VETresid(SVET) (R^2 adj = -3.3%)				
	Intercept	0.025	0.436	0.056	0.955
	Right Hem Matching Dpr	*-0.055*	*0.482*	*-0.113*	*0.910*
	Left Hem Matching Dpr	0.169	0.466	0.363	0.718
BIRD	VETresid(SVET) (R^2 adj = -0.8%)				
	Intercept	-0.395	0.466	-0.848	0.400
	Right Hem Matching Dpr	-0.036	0.352	-0.101	0.920
	Left Hem Matching Dpr	0.483	0.397	1.220	0.229

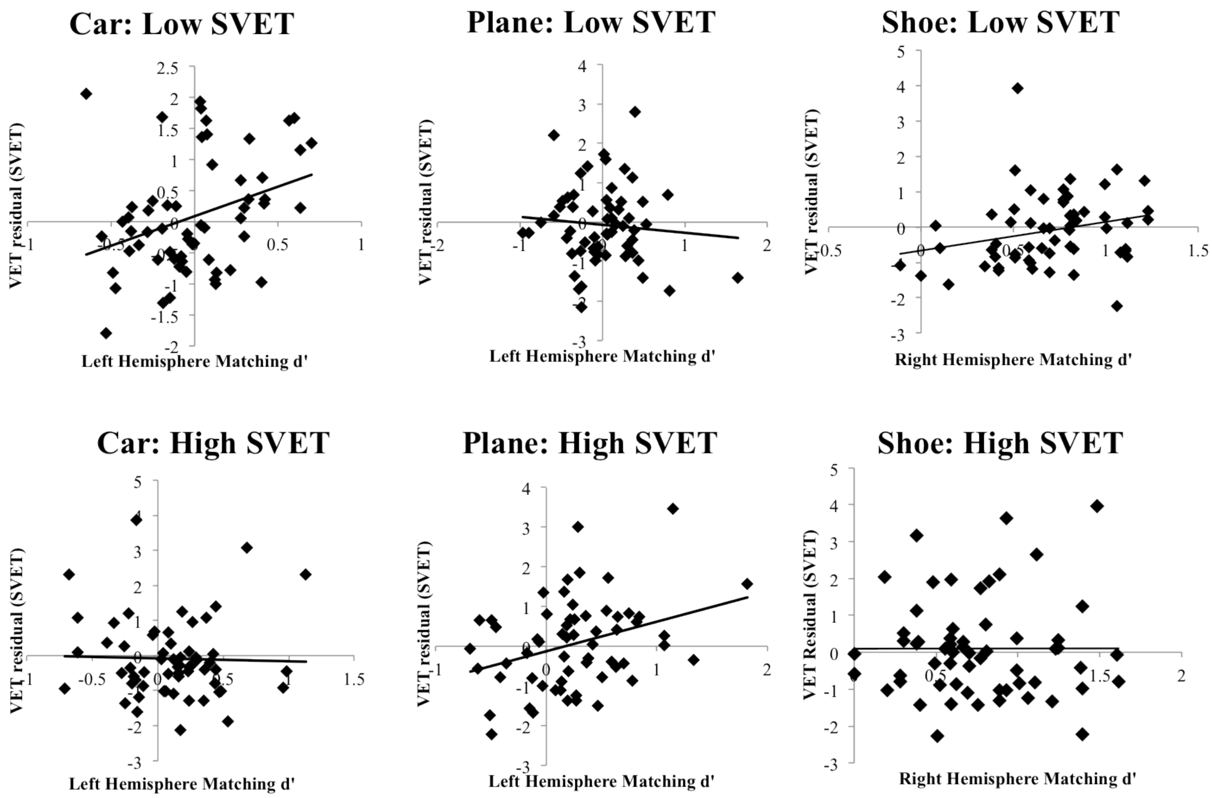
For cars, we observed that left hemisphere matching predicted VET residuals for subjects with low SVET scores (partial correlation: $r(115)=0.34$, β=0.95 $p=0.009$) but not for those with high SVET scores (partial correlation: $r(115)=-0.03$, β=-0.76, $p=0.10$) (Figure 23). In contrast, for planes, we found the opposite relationship. We observed that left hemisphere matching was strongly related to VET residuals for subjects with high SVET scores (partial correlation: $r(115)=0.31$, β=0.72, $p=0.02$) but not for those with low SVET scores (partial correlation:

$r(115)=-0.09$, $\beta=-0.20$, $p=0.49$). For shoes, we did not find a statistically significant contribution of right hemisphere to VET residuals in either low or high SVET subjects, however, qualitatively, the right hemisphere matching was more strongly related to VET residuals for subjects with low SVET scores (partial correlation: $r(115)=0.25$, $\beta=0.69$, $p=0.13$) than high SVET scores (partial correlation: $r(115)=0.01$, $\beta=-0.06$, $p=0.91$).

The findings for planes fit our prediction based on the results of McGugin and colleagues (2012a), with greater semantic knowledge associated with greater left hemisphere contribution to visual processing. Had we tested only planes, the story might have been simple. But the opposite pattern was found for cars. An explanation of the car results in left hemisphere is less apparent, and the results could differ from those for planes either because cars is a category for which more subjects possess semantic knowledge (compared to planes) and/or because SVET itself accounted for more of the overall variance in VET performance.

Despite showing a significant interaction in the initial model, in this analysis we did not observe a significant right hemisphere contribution to VET(SVET) for shoes for either low or high SVET subjects, although the qualitative pattern suggests a stronger contribution for low SVET subjects. Since we had no prediction about how the right hemisphere contribution to VET performance would be mediated by SVET scores, we note that the effect for shoes is generally similar to that for birds. While the interaction between SVET and hemisphere was only marginally significant in the original model for birds, when SVET scores were split we observed a positive contribution of right hemisphere matching to VET(SVET) in low SVET subjects only. This may indicate a different strategy for those with limited semantic bird knowledge, one that relies more on processing in the right hemisphere than for those with greater semantic bird knowledge who can apply knowledge beyond visual information to the task.

Figure 23. Scatterplots showing the partial correlations of VET residual (SVET) with left hemisphere matching d' for cars and planes and right hemisphere matching d' for shoes, split by median SVET scores: low SVET scores (top row) and high SVET scores (bottom row).



Conclusions

In Study 3, we used the SVET as an independent measure to ask how levels of semantic knowledge for a category might affect visual processing of those objects in the brain. Our hypothesis based on previous findings (Curby et al., 2004; McGugin et al., 2012a) was that subjects with high semantic knowledge might recruit both hemispheres more equally in a visual task, while those with less semantic knowledge would use mostly the right hemisphere, therefore suggesting that left hemisphere recruitment may result from automatic activation of semantic knowledge during a visual task.

We used SVET scores together with a peripheral visual matching task to target left and right hemisphere processing to predict the performance on the VET, a foveal visual task. We found a significant interaction between semantic knowledge in the left hemisphere for cars and planes, and in the right hemisphere for shoes and birds. The results for planes suggest this interaction is different for low versus high SVET subjects, which match our predictions as well as those by McGugin and colleagues (2012a) in fMRI. To better interpret the results for cars, shoes, and birds, future studies that examine the effect of category experience in more detail and test a wider range of categories are needed to broaden our understanding of when these semantic and hemispheric interactions occur and in what direction.

Perhaps even more importantly for the goals of this dissertation, Study 3 demonstrates that the SVET is a valuable measurement tool that provides additional information beyond measures of visual processing for investigating the effects of semantic knowledge and object experience on visual and cognitive processes. This study provides an example of how individual differences work can contribute to our understanding of high-level visual processes, allowing us to ask questions that cannot be answered by group analyses. We hope to continue using the SVET as a measure in future work and will make it available to other psychologists who may wish to use it as well.

Chapter 5 – Conclusion

In this dissertation we have made significant progress in the measurement of semantic knowledge for an object category, and with novel measures, we learned about how semantic knowledge relates to visual performance, domain-general abilities, and category experience. Now instead of just measuring performance on a single bird recognition test, we could have Theo and Liz complete an entire battery of tests that would tell us about their visual and semantic knowledge with birds and other categories, including faces, and self-report measures of experience in these domains. If Liz performs well on VET-Bird but not SVET-bird and Theo does well on both VET-Bird and SVET-Bird, how might we understand their visual bird performance? We might look at visual performance on all VETs and see that Liz performs well on the other VETs and on the CFMT, suggesting that her visual bird performance is not the result of bird experience but reflects her high domain-general visual ability. If Liz also reports low experience with birds that would suggest her poor SVET-Bird performance is because of a lack of bird experience, especially if her scores are not low on all SVETs or for categories with which she reports more experience. If Theo reports above average bird experience and his VET-Bird and SVET-Bird performance are well correlated with each other but not with other categories, this would suggest that Theo's performance with birds is high at least in part because he has greater experience in this domain. Of course the exploration of individual differences is best done in the context of a much larger data set, such as the ones we gathered in this dissertation.

Our goal in this work was to create a valid and reliable measure of semantic knowledge that could be used for many different categories and which, together with other measures, would help us better understand object recognition performance.

We successfully created a standardized, non-visual measure of semantic knowledge, the SVET, for eight object categories. The SVET, which tests knowledge of object names and labels for a category, is concise and can be completed by subjects at all levels of category experience. The SVET is also adaptable to many categories, which is important because testing performance on many different categories is critical to interpreting performance as reflecting either domain-general or domain-specific influences. We found that each of the SVETs was reliable and offered good coverage of all levels of performance in both an online sample and a university sample as well as with experts tested on the SVET-Bird. We provided evidence for the validity of the VET in a number of ways: i) SVETs showed domain-specific correlations with their corresponding VETs, a result that combines both convergent and discriminant validity; ii) similarly, SVETs showed domain-specific correlations with reports of experience; iii) the SVET discriminated experts from novices in the bird domain; iv) SVETs were more related to domain-specific measures than to fluid intelligence or face recognition. The SVET thus provides a novel and valuable tool to measure semantic performance independent of visual performance.

Beyond measurement of semantic knowledge we were interested in SVET performance because it offered another measure that would be influenced by category experience. As such the SVET could eventually provide a way to estimate experience independent of self-report. One analysis of particular interest was the relationship between visual (VET) and semantic (SVET) performance for the same category. We hypothesized that the shared variance between these two different tasks should reflect primarily domain-specific experience. In other words, after we removed the influence of domain-general variables (age, sex, *Gf*, and non-category performance: an estimate of general visual ability for the VET and of general verbal ability for the SVET), we found that experience as measured by our self-report measure contributed independently and significantly to the VET-SVET relationship for six of the eight categories we tested. Indeed, in most cases, once regressing out experience there was very little shared variance left between VET and SVET, less then 3% (the exception was cars, where the correlation still accounted for 16% of the variance). Two categories did not demonstrate a significant independent contribution of experience: shoes and mushrooms. This could be because of a large sex effect that was strongly correlated with shoe experience, accounting for much of the variance for shoes, and because mushrooms demonstrated low variability in both experience and VET and SVET performance in our sample, leaving little individual variability to explain with a small range of experience. Whether the differences between categories found here replicate or depend on properties of our specific sample remains to be seen. Overall, these results generally supported our hypothesis that the main contributor to domain-specific overlap between visual and semantic performance is experience. These results also serve to validate our domain-specific self-report measures of experience.

In our analyses removing domain-general factors to see if the remaining shared variance between VET and SVET reflected only expertise, we found in some cases, such as cars, that a significant share of variance remained unexplained after removing the contribution of experience. Aside from the possibility that this could be due to error in measurement of experience, we explored whether some of this variance could be attributed to a strategy of using names to remember objects in the visual task. With a category for which even experts lack subordinate-level names, shoes, we found that the relationship between VET and SVET remained strong. And with a category for which there was great variability in the ability to name objects at the subordinate-level, birds, we found that while bird naming may account for some of the variance shared by VET and SVET, this variance was independent of self-reported experience with birds.

Furthermore, whereas in our university sample for birds we observed that removing experience rendered the VET-SVET partial correlation non-significant, for birders we found that the VET and SVET remained correlated after partialing out the role of experience. This result mirrors a situation that has been postulated for face recognition in the normal population: when experience is high, experience is expected to account for less of the variability in performance (Gauthier et al., in press).

In the introduction we made the prediction based on previous results (Davis et al., 2011; Hedley et al., 2011; Wilhelm et al., 2010; Wilmer et al., 2010) that two domain-general abilities, *v* and *Gf*, would not be related. The idea that domain-general abilities are related to a common underlying factor has been central to a great deal of work in the study of intelligence, as expressed in the classic idea of a general factor (*g*) (Chiappe & MacDonald, 2005; Garlick, 2002; Horn, 1968; Horn & Cattell, 1966; Humphreys, 1979). This is one reason to be excited about new developments in the study of individual differences in face recognition, and by extension now also object recognition: individual abilities that can be reliably measured and that are not strongly related to *g* have the potential to broaden the scope of predictions and understanding in human behavior.

To the extent that face recognition can be assumed to be a good estimate of *v*, we found that face recognition (CFMT) and *Gf* were correlated with each other, although the correlation accounts for a very small proportion of the variance (r=0.14, R^2=0.02), especially given that these are highly reliable measures.

We also found a correlation between each VET and *Gf*, a relationship that was statistically significant except for cars and shoes (mean for all categories, r=0.22). Interpreting such a correlation is somewhat difficult, because it could reflect how intelligence influenced learning of domain specific information, or it could reflect the correlation between *Gf* and domain-general visual ability (*v*) as expressed in each domain-specific test. However, we can also look at the correlation between *Gf* and an aggregate VET score (VET-All) for all eight categories. By aggregating across categories, we would expect to reduce domain-specific contributions, to the extent these contributions are not correlated between categories. In Study 2A, the correlation between VET-All and *Gf* was numerically higher (r=0.32) than the average correlation with each category (mean r=0.21)*,* suggesting a domain-general *v* could be related to *Gf.* Interestingly, VET-All was even more correlated with CFMT (r=0.41), again more than the average correlation between CFMT and each VET (mean r=0.27). *Gf* appears to share more variance with measures of object recognition than with our measure of face recognition, and this difference may be difficult to attribute to experience, which in itself does not relate to *Gf,* either when averaged across all categories (r=-0.06) or when computed separately for each category (mean r=-0.03).

In contrast, the correlation between SVET and *Gf* was always weaker than between VET and *Gf* for every category. The correlation between SVET and *Gf* was statistically significant but small for six of the eight categories. SVET-All, the average of SVET performance averaged across all eight categories, was correlated with *Gf* more (r=0.20) than any of the single category SVETs (mean r=0.10).

While some theories of intelligence suggest that all domain-general cognitive abilities are related to *g*, others dispute this common relation for some domain-general abilities (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Friedman et al., 2006). Why is the VET more correlated with *Gf* then the SVET? It may be because it is a learning and memory task, while the SVET is in essence a measure of crystallized knowledge. Of course, the extent to which someone learns over the course of a short VET, or the kind of learning that occurs, may depend on how much they know about a category to start with. For instance, for faces, learning might be limited to the specific exemplars on the test, whereas for mushrooms in novices, subjects may learn that some features are more diagnostic of identity for the entire category. In future work it would be interesting to investigate what domain-general abilities are more related to acquiring knowledge tested in the SVET.

Future work might also explore why we observed stronger sex differences in the male-interest categories for the SVET than the VET, and if this is related to any domain-general abilities or differences in types experience (e.g. recognizing objects versus reading about them) with those categories. We would also be curious to investigate further what might explain the remaining domain-specific variability we observed between VET and SVET after removing age, sex, *Gf,* non-category task performance, and self-reported experience. Perhaps there is another aspect of category-specific experience not captured by our experience measure, or domain-general abilities that apply selectively to performance with some categories more than others.

This work investigating the contributions to object recognition performance represents a new direction in individual differences work in high-level vision. While some work has been done to understand relationships between face recognition and other performance measures (Dennett et al., 2011; Gauthier et al., in press; Wilhelm et al., 2010; Wilmer et al., 2010; 2012) investigating the contribution of experience in domains with greater variability in experience has not been done. Here we applied a framework in which domain-specific measures can be used to estimate domain-specific effects after partialing out domain-general variance. Similar methods have been used in cognitive domains (Hambrick, 2003; Hambrick et al., 2007; 2008; Stanovich & Cunningham, 1992) but have not been applied to questions in object recognition. In fact, the next step in this research program might follow in the footsteps of these cognitive studies to apply structural equation modeling methods to our measures, to estimate latent variables that correspond to both domain-specific and domain-general factors.

Our use of item response theory (IRT) methods to measure subject performance and refine our measures is atypical in cognitive psychology, but we found it useful both for test creation and data interpretation. We hope to continue and extend our use of IRT in future work on the SVET and other individual differences measures.

We hope that the SVET will be a useful tool that can be employed as an independent measure to understand other perceptual and cognitive phenomena. In Study 3 we demonstrated how SVET scores might be used to understand lateralization of visual processing in the brain, which we hypothesized based on previous work might recruit the left hemisphere more with greater semantic knowledge (Curby et al., 2004; McGugin et al., 2012a). We found that for planes, increased left hemisphere performance was related to high SVET scores as predicted, however interpretation for other categories was less clear and will require more investigation – but the results suggested that lateralization of visual processing interacts with a subject's level of semantic knowledge. The SVET thus appears to measure informative variability that could be helpful to investigate individual differences in high-level visual processes.

In conclusion, this dissertation presents the SVET as a reliable and valid measure of semantic knowledge and demonstrates its use for measurement of semantic knowledge for a range of categories and populations. It also illustrates the use of the SVET as a measure of category-specific performance that reflects category-specific experience to understand the abilities and experience that contribute to performance. In the future we will continue to refine and expand the SVET, both in new categories and extended versions for expert populations, and we will make it available to other psychologists who wish to use it in their research. We hope this work contributes to the exciting new research area of individual differences in object recognition and look forward to continuing this research using the SVET together with fellow psychologists.

REFERENCES

Barton, J. J. S., Hanif, H., & Ashraf, S. (2009). Relating visual to verbal semantic knowledge: the evaluation of object recognition in prosopagnosia. *Brain*, *132*(12), 3456–3466. doi:10.1093/brain/awp252

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., … Yovel, G. (2009). Diagnosing prosopagnosia: effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455. doi:10.1080/02643290903343149

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bub, D. N., Masson, M. E. J., & Lin, T. (2013). Features of planned hand actions influence identification of graspable objects. *Psychological Science*, *24*(7), 1269–1276. doi:10.1177/0956797612472909

Chiappe, D., & MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *Journal of General Psychology*, *132*(1), 5–40. doi:10.3200/GENP.132.1.5-40

Cho, S.-J., Wilmer, J., Herzmann, G., McGugin, R., Fiset, D., Van Gulick, A. E., … Gauthier, I. (submitted). Item Response Theory Analyses of the Cambridge Face Memory Test (CFMT).

Cole, D. A., Cho, S.-J., Martin, N. C., Youngstrom, E. A., March, J. S., Findling, R. L., … Maxwell, M. A. (2012). Are increased weight and appetite useful indicators of depression in children and adolescents? *Journal of Abnormal Psychology*, *121*(4), 838–851. doi:10.1037/a0028175

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, *8*(3), e57410. doi:10.1371/journal.pone.0057410.t001

Curby, K. M., Hayward, W. G., & Gauthier, I. (2004). Laterality effects in the recognition of depth-rotated novel objects. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(1), 100–111.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PLoS ONE*, *6*(12), e28800. doi:10.1371/journal.pone.0028800

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., … Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359–1364. doi:10.1126/science.1194140

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2011). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587–605. doi:10.3758/s13428-011-0160-2

Duchaine, Brad, & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *17*(2), 249–261. doi:10.1162/0898929053124857

Duchaine, Brad, & Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585.

doi:10.1016/j.neuropsychologia.2005.07.001

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Services.

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *8*(4), 341–349.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence-Erlbaum.

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category Learning Increases Discriminability of Relevant Object Dimensions in Visual Cortex. *Cerebral Cortex*. doi:10.1093/cercor/bhs067

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. doi:10.1111/j.1467-9280.2006.01681.x

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, *109*(1), 116–136. doi:10.1037//0033-295X.109.1.116

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, *20*(3), 507–523. doi:10.1080/02643290244000275

Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Van Gulick, A. E. (in press). Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of Vision*.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. doi:10.1038/72140

Germine, L. T., Duchaine, B., & Nakayama, K. (2011a). Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition*, *118*(2), 201–210. doi:10.1016/j.cognition.2010.11.002

Germine, L., Cashdollar, N., Düzel, E., & Duchaine, B. (2011b). A new selective developmental deficit: Impaired object recognition with normal face recognition. *Cortex*, *47*(5), 598–607. doi:10.1016/j.cortex.2010.04.009

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. doi:10.3758/s13423-012-0296-9

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology General*, *123*(2), 178–200.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hambrick, D. Z. (2003). Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition. *Memory & Cognition*, *31*(6), 902–917. doi:10.3758/BF03196444

Hambrick, D. Z., Meinz, E. J., & Oswald, F. L. (2007). Individual differences in current events knowledge: Contributions of ability, personality, and interests. *Memory & Cognition*, *35*(2), 304–316. doi: 10.3758/BF03193451

Hambrick, D. Z., Pink, J. E., Meinz, E. J., Pettibone, J. C., & Oswald, F. L. (2008). The roles of ability, personality, and interests in acquiring current events knowledge: A longitudinal study. *Intelligence*, *36*(3), 261–278. doi:10.1016/j.intell.2007.06.004

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research information support. *Journal of Biomedical Informatics*, *42*(2), 377–381.

Hedley, D., Brewer, N., & Young, R. (2011). Face recognition performance of individuals with Asperger syndrome on the Cambridge Face Memory Test. *Autism Research : Official Journal of the International Society for Autism Research*, *4*(6), 449–455. doi:10.1002/aur.214

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*(3), 242–259.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270.

Hughson, A. L., & Boakes, R. A. (2002). The knowing nose: The role of knowledge in wine expertise. *Food Quality and Preference*, *13*(7), 463–472.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence*, *3*(2), 105–120.

James, K. H., James, T. W., Jobard, G., Wong, A. C.-N., & Gauthier, I. (2005). Letter processing in the visual system: Different activation patterns for single letters and strings. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(4), 452–466. doi:10.3758/CABN.5.4.452

James, T. W., & Gauthier, I. (2003). Auditory and Action Semantic Features Activate Sensory-

Specific Perceptual Brain Regions. *Current Biology*, *13*(20), 1792–1796. doi:10.1016/j.cub.2003.09.039

James, T. W., & Gauthier, I. (2004). Brain areas engaged during visual judgments by involuntary access to novel semantic information. *Vision Research*, *44*(5), 429–439. doi:10.1016/j.visres.2003.10.004

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognitive Psychology*, *16*(2), 243–275.

Kann, J. (1950). A Translation Of Broca's Original Article On The Location Of The Speech Center. *Journal of Speech and Hearing Disorders*, *15*, 16–20.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Marsolek, C. J. (1999). Dissociable Neural Subsystems Underlie Abstract and Specific Object Recognition. *Psychological Science*, *10*(2), 111–118.

McGugin, R. W., Gatenby, C., Gore, J., & Gauthier, I. (2012a). High-resolution imaging of expertise reveals reliable object selectivity in the FFA related to perceptual performance. *Proceedings of the National Academy of Scince*, *109*(42), 17063–17068. doi: 10.1073/pnas.1116333109

McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012b). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research*, *69*, 10–22. doi:10.1016/j.visres.2012.07.014

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. New York, NY: Psychological Corporation.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., …

Engle, R. W. (2012). No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study. *Journal of Experimental Psychology General*. doi:10.1037/a0029082

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological Science*, *22*(4), 464–471. doi:10.1177/0956797611401753

Rosch, E., Mervis, C. B., Gray, W., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Psychology*, *7*, 573–605.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: people with extraordinary face recognition ability. *Psychonomic Bulletin and Review*, *16*(2), 252–257. doi:10.3758/PBR.16.2.252

Schmidt, F. L. (2014). A General Theoretical Integrative Model of Individual Differences in Interests, Abilities, Personality Traits, and Academic and Occupational Achievement: A Commentary on Four Recent Articles. *Perspectives on Psychological Science*, *9*(2), 211–218. doi:10.1177/1745691613518074

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, *20*(1), 51–68.

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

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16*(2), 145–151. doi:10.1111/j.0956-7976.2005.00795.x

Thurstone, L. L. (1938). *Primary Mental Abilities*. Chicago, IL: University of Chicago Press.

Van Gulick, A. E., & Gauthier, I. (in press). The perceptual effects of learning object categories that predict perceptual goals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, *383*(6597), 254–256. doi:10.1038/383254a0

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience*, *5*(5), 491–499. doi:10.1038/nn839

Wagner, A. D., Desmond, J. E., Demb, J. B., Glover, G. H., & Gabrieli, J. D. E. (1997). Semantic Repetition Priming for Verbal and Pictorial Knowledge: A Functional MRI Study of Left Inferior Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *9*(6), 714–726. doi:10.1126/science.8316836

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces-One element of social cognition. *Journal of Personality and Social Psychology*, *99*(3), 530–548. doi:10.1037/a0019972

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: The example of face recognition. *Cognitive Neuropsychology*, *29*(5-6), 360–392. doi:10.1080/02643294.2012.753433

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., … Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*(11), 5238–5241. doi: 10.1073/pnas.0913053107

Wiseman, S., MacLeod, C. M., & Lootsteen, P. J. (1985). Picture recognition improves with subsequent verbal information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 588–595.

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009a). Conditions for facelike expertise with objects: becoming a Ziggerin expert--but which type? *Psychological Science*, *20*(9), 1108–1117. doi:10.1111/j.1467-9280.2009.02430.x

Wong, A. C.-N., Palmeri, T. J., Rogers, B. P., Gore, J. C., & Gauthier, I. (2009b). Beyond Shape: How You Learn about Objects Affects How They Are Represented in Visual Cortex. *PLoS ONE*, *4*(12), e8405. doi:10.1371/journal.pone.0008405.g004

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2012). The nature of experience determines object representations in the visual system. *Journal of Experimental Psychology General*, *141*(4), 682–698. doi:10.1037/a0027822

Woolley, A. W., Gerbasi, M. E., Chabris, C. F., Kosslyn, S. M., & Hackman, J. R. (2008). Bringing in the Experts: How Team Composition and Collaborative Planning Jointly Shape Analytic Effectiveness. *Small Group Research*, *39*(3), 352–371. doi:10.1177/1046496408317792

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, *15*(8), 1234–1242. doi:10.1093/cercor/bhi006

Zell, E., & Krizan, Z. (2014). Do People Have Insight Into Their Abilities? A Metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. doi:10.1177/1745691613518075

APPENDIX

Appendix A. Complete SVETs (1.0) for eight categories. Each SVET is 51 trials: 48 test trials and 3 catch trials. Each row shows the 3 names (1 real, 2 foil) presented on each trial. The real name is shown in orange. Trials are ordered (excluded catch trials) approximately from easiest to most difficult.

SVET-Car.

Trial	Name1	Name2	Name3
1	Infiniti Kobuk	Scion dT	Dodge Viper
2	BMW M56	Mercury Manitu	Ford Mustang
3	Pontiac G7	BMW Caspari	Ford Taurus
4	Lincoln Leaf	Nissan Sentra	Porsche Crossfire
5	Chrysler Osprey	Mitsubishi Prancer	Nissan Altima
6	Volvo Focus	Mercedes-Benz C300	Mercury Alero
7	Hyundai Altitude	Mitsubishi Eclipse	Kia Gala
8	Chevrolet Flash	Volkswagen El Peso	Buick Regal
9	Toyota Prius	Jaguar Lisbon	Scion xR
10	Volvo GS350	Chevrolet Lancer	Dodge Charger
11	Hyundai Yucatan	Buick LeSabre	Lincoln Jetta
12 *Catch*	Honda Civic	Palm Tree	Snickers Bar
13	Toyota Calisto	Chrysler PT Cruiser	Hyundai Corolla
14	Pontiac GTO	Aston Martin Matrix	Subaru Woodlands
15	Pontiac Sky	Cadillac DeVille	Volvo Z60
16	Volvo S60	Suzuki 911 Carrera	Volkswagen Juniper
17	Nissan Muse	Audi A6	Chevrolet LaCrosse
18	Hyundai Elantra	Saturn Neon	Kia Cloud
19	Dodge Festival	Audi Vita	Mazda Miata
20	Chevrolet Camaro	Cadillac Escort	Subaru Malibu
21	Jaguar XJ	Lamborghini Nuvola	Acura NRX
22	Audi Z4	Mazda Kizashi	Chevrolet Volt
23 *Catch*	Winter Storm	Ford Fiesta	Rose Garden
24	BMW 580d	Volkswagen GTI	Toyota Lucerne
25	Nissan Azera	BMW 550i	Kia Golf
26	Suzuki Prestige	Infiniti G37	Pontiac S550
27	Lamborghini Gallardo	Toyota Sonata	Lincoln Olympic
28	Oldsmobile Cavalier	Lexus Aventador	Volvo C70
29	Chrysler Concorde	Lexus CD350	Buick Inspiron
30	Mercury Grand Marquis	Suzuki Avenger	Honda Yaris
31	Buick Chesapeake	Subaru Impreza	BMW 490x
32	Mazda Blaze	Ford Fiber	Honda Fit
33	Saturn Fuze	Honda Soul	Toyota Avalon
34	Kia Forte	Mitsubishi STZ	Infiniti Dream
35	Dodge Grand Prix	Mitsubishi Ion	Subaru Legacy
36	Saab Eban	Lincoln MKZ	Lexus Sable
37	Scion G6	Mercury Galant	Cadillac XTS
38	Oldsmobile Primo	Porsche 538	Mercury Milan
39	Saab 3-9	Hyundai Genesis	Cadillac Revel
40	Bentley Continental GT	Mercedes-Benz Park Avenue	Chrysler Crusader
41	Lexus ES300	Ford Impala	Acura Optima
42	Acura QR320	Porsche Cayman	Subaru Camry
43 *Catch*	Denim Skirt	Yorkshire Terrier	Toyota Matrix
44	Chrysler Maxima	Scion tC	Lamborghini Magnum
45	Volkswagen Mulsanne	Buick Intrepid	Oldsmobile Aurora
46	Saab 9-5	Acura Sebring	Nissan xD
47	Aston Martin DB9	Honda Octave	Infiniti ILX
48	Lexus LF-CC	Bentley Beetle	Jaguar 9-3
49	Saturn Fusion	Acura TSX	Saab S80
50	Audi Allroad	Cadillac Amethyst	Bentley Baltic
51	Audi A9	Oldsmobile Rocoto	Honda Insight

SVET-Plane.

Trial	Name1	Name2	Name3
1	F-16	Utah	Vapor
2	737	Serpens	Sheffield
3	Hellcat	W-66	LP-8
4	949	B-52	Flyingfish
5	C-130	Su-800	Black Skiff
6	P-209	Libra	B-2
7 *Catch*	747	Lean Cuisine	Facebook
8	Me 200	ES-69	F-105
9	HLB	J-49	DC-3
10	Starlight	Bouncer	717
11	A400	Cygnus	777
12	87A	CS100	1020
13	96Y	Hawk	Roehr
14	8030	T-38	Bylon
15	NDA	DC-10	Bowman Cx36
16	Lester	Rotterdam	Liberator
17	877	A320	Ruby
18	MD-80	Q70	1010
19	C-17	MD-20	Sl-60
20	N7	X-1	RYY
21	A-10	D789	BV 10
22	Falcon 900	Courante	T-017
23	A380	B-6 Sprinter	Robson
24	8900	A2 Lobo	Spitfire
25 *Catch*	Barnes and Noble	A319	Cool Whip
26	A-49	F/A-18	898
27	L-300	P-51	Tempo
28	CS300	Td 500	R-180
29	Gopher	Panther	MD-11
30	Lagrange	A340	797
31	Y 88	Juno	Citation Jet
32	DA20	Locus	51-md
33	Predator	T700	Ocelot
34	6690	Cub	r590
35	Protector	Dakota	78K
36	Yuri	CRJ 5007	E175
37 *Catch*	Reese's Cup	Walgreens	F2
38	KZ-66	AirPrince	L-1011
39	Yak-130	XX-30	MK-477
40	432	King Air	LF-105
41	Cherokee	Z-7	Arizona
42	Missouri	J25	Raven
43	AM 99	LJ 431	Su-47
44	67V	Booker B-8	Otter
45	Mosquito	Western Lair	A480
46	393	VB-40	Starship
47	Z1	Ju 88	RT-9
48	BT10	Bf 109	Nova 8
49	Ural	DC-300	Camel
50	F-25	Dash 8	19-10
51	Me 262	F-41	Nightranger

SVET-Transformer.

Trial	Name 1	Name 2	Name 3
1	Uppercut	Outlook	Megatron
2	Courage	Starscream	Top Notch
3	Fivepin	Razorclaw	Riot
4	Lavaman	Chromoburn	Quickstrike
5	Lightning Rod	Thunderclash	Firecraft
6	Highboxer	Tigatron	Terraclash
7	Bumblebee	Astromega	Receptor
8	Torrent	Ironhide	Delta Minor
9	Lordov	Amphius	Scorch
10	Quatraquake	Boomerjet	Sideswipe
11	Mallet	Dom	Soundwave
12	Ratchet	Volcano	Dasher
13	Wind Dagger	Camrod	Smokescreen
14	Dustrage	Roll Archer	Orion
15	Fox	Ricochet	Pitfall
16 *Catch*	Dunkin Donuts	Cheddar Cheese	Shipwreck
17	Fireflight	Triblast	Dune Snare
18	Bluebreak	Moor Knight	Grapple
19	Spearonus	Inferno	Fuse
20	Roadbuster	Combust	Loggerhead
21	Vulture	Blitzwing	Crash
22	Skyhammer	Starshooter	Neoblot
23	Moonrider	Windcharger	Converse
24	Terp	Prowl	Carbonspin
25 *Catch*	Frosted Flakes	Vanquish	Oatmeal Raisin
26	Flytrap	Tungsten	Reflector
27	Obsidian	Double Dare	Excelsion
28	Sunstreaker	Septawave	Proton
29	Cliffjumper	Nailclaw	Tanji
30	Grimmel	Giltwheel	Mirage
31	Jetstorm	Megaglide	Springshot
32	Speedswoop	Razorbyte	Talon
33	Thundercracker	Buzzcraft	Vilius
34	Flashrun	Solopred	Jazz
35	Hoverburst	Mort	Blurr
36	Enemy	Sonic Thunder	Bounce
37	Raincharge	Crosscut	Hustler
38	Zeus	Hurricane	Airlock
39	Long Haul	Quickjet	Junction
40	Breacher	Dawn Bird	Tracker
41	Wolfspur	Night Boomerang	Chase
42	Barracuda	Hubcap	Koben
43 *Catch*	Shrapnel	Diet Coke	J. Crew
44	Barricade	Ironwheel	Skidbit
45	Space Terror	Omicron Prime	Air Raid
46	Waveracer	Hound	Sotter
47	Canis Major	Cheetor	Sharpstrike
48	Hornet	Kickback	Crossfire
49	Punch	Deepwave	Roadflux
50	Arcee	Grimbolt	Victorion
51	Suntracker	Steelhead	Pointblank

SVET-Dinosaur.

Trial	Name1	Name2	Name3
1	Tyrannosaurus Rex	Asperdatylus	Telemosaurus
2	Phoboraptor	Triceratops	Ditlosaurus
3	Brachiosaurus	Paramaxilosaurus	Fabrilukosaurus
4	Lopholurius	Pirongocoelus	Velociraptor
5	Canthusius	Meranoleptes	Plateosaurus
6	Tarbonyx	Dragosaurus	Ceratosaurus
7	Pentaceratops	Eudontidectes	Microtarius
8	Pachycephalosaurus	Namibiasaurus	Reginasaurus
9	Plesiosaurus	Timorspondylus	Lanaptasaurus
10	Tonivius	Amygdalodon	Amerivenator
11	Geldanosaurus	Panoplosaurus	Scuriosaurus
12	Nodocaudosaurus	Spikosaurus	Segisaurus
13 *Catch*	Barosaurus	Betty Crocker	Nike
14	Dyptiodon	Protoceratops	Maxiosaurus
15	Tetrachelodon	Coleoptera	Megalosaurus
16	Celeritasaurus	Apatosaurus	Delphysis
17	Bactronychus	Dilophosaurus	Latimosaurus
18	Dromopedosaurus	Diplodocus	Gymnodontosaurus
19	Lestipidius	Parasaurolophus	Dneipidosaurus
20	Herbiodon	Archaeopteryx	Appellasaurus
21	Montanasaurus	Erhinodon	Stegoceras
22	Stuthioceratops	Centaurisaurus	Iguanodon
23	Ramseysaurus	Caenagnathus	Dirulius
24 *Catch*	KitchenAid	Titanosaurus	Microsoft
25	Spinosaurus	Roxithromius	Andromelosaurus
26	Saurolophus	Allocephale	Ceralopus
27	Compsognathus	Amorispinax	Artemidorus
28	Tetramorphodon	Oviraptor	Draconychus
29	Deinonychus	Salvatosaurus	Rugosaurus
30	Voloceratops	Hepatolodon	Ankylosaurus
31	Hadrosaurus	Letoraptor	Plateothersaurus
32	Angusticeratops	Yukonsaurus	Gallimimus
33	Vulcanodon	Poissalodon	Okavangosaurus
34	Lesothosaurus	Allobrachiosaurus	Voltaeodon
35	Paraprantadon	Telmatosaurus	Barocheirus
36	Segnoceratops	Zulosaurus	Deinocheirus
37	Corythosaurus	Styrenosaurus	Homodagnius
38	Mauryonyx	Rostrosaurus	Achillobator
39	Parbosaurus	Polybutisaurus	Lambeosaurus
40	Brassicasaurus	Deinosternus	Lapparentosaurus
41	Heptalogodon	Zephyrosaurus	Procerimimus
42	Prontosaurus	Orthithomimus	Hydrapentasaurus
43	Decacornutosaurus	Euovatosaurus	Stygimoloch
44 *Catch*	Cadillac	Crock-pot	Conchoraptor
45	Ornitholopolus	Niposcephales	Mircovenator
46	Seismosaurus	Tyrannoraptor	Pallosaurus
47	Microceratus	Dryptoplatyornis	Rhynchodon
48	Corposaurus	Monocyclosaurus	Mussaurus
49	Ceraphalangiamimus	Othnielia	Plurasaurus
50	Indostedosaurus	Skorpiovenator	Bagalosaurus
51	Maiasaura	Megacapitosaurus	Pachypedolus

SVET-Shoe.

Trial	Name1	Name2	Name3
1	Cristallo	Gucci	Fazzolari
2	Angelo Frega	Anong	Prada
3	Comoros	Christopher Phan	Anne Klein
4	Nine West	Rebecca Fox	Aloft
5	Kenneth Cole	Londa	Steve Hart
6	Carolyn Palmer	Clover	Dolce Vita
7	Birdie Hamel	Thaksin	Michael Kors
8 *Catch*	Cuisinart	Honda	Vigotti
9	Semillon	Jimmy Choo	Madison Long
10	Guillaume Deschamps	Oscar de la Renta	Eze
11	Le Chat Chic	Christian Louboutin	Lindsey Speegle
12	Tai Ladd	Elliott Pierce	Miz Mooz
13	Dahlia	Versa	Yves Saint Laurent
14	Parade	Betsey Johnson	Dowell
15	Ruby	Soustel	Etienne Aigner
16	Phillip Weinkopf	Lotte	Kate Spade
17	Paul Xu	Enna	Manolo Blahnik
18	Marcus Rivera	Aldo	Cimarron
19	Isaac Mizrahi	Six Swans	Lily James
20	Alexandre Birman	Portici	Larkin
21	Arzog	Brian Atwood	M. Rose
22	Anika Taylor	James Colver	Balenciaga
23	Nissa Takou	Olivia Skelt	Pedro Garcia
24	Piper	Joseph Blount	Miu Miu
25	Zetta	Kalden White	Franco Sarto
26	Rebecca Minkoff	Daquin	Paolo Trella
27	Vasquez	Taryn Rose	Darby Hill
28 *Catch*	Hyundai	J. Renee	Pepperidge Farm
29	Giuseppe Zanotti	Francisco Soto	Sara and Sophie
30	Cole Haan	Operetti	Melissa Perry
31	Ava Amini	Ivanka Trump	Serra
32	Enzo Angiolini	Nicole Hall	Victor Russo
33	Azzuri	Via Spiga	Maison du Roi
34	Steve Madden	Isabelle Laurent	Five Degrees
35	Kevin Dunn	Badgley Mischka	Cecille
36	Lola Wong	Sam Edelman	DBA
37	Pollini	Arresi	P. Van Vliet
38	Marcelino	Stuart Weitzman	R. Campbell
39	Nina	Adele Hirsch	Molinelli
40	Sigerson Morrison	Pebble and Stream	Michael Williams
41	Alston Brett	Tiger Pearl	Seychelles
42 *Catch*	Alfani	John Deere	Duracell
43	Alice + Olivia	Belle Amie	Vega
44	Laurel	Charlotte Olympia	J.R. Santuk
45	Cote Vert	Vince Camuto	Sergio Nicoletti
46	Elizabeth and James	Joshua Gold	Claudia Escotto
47	Graham Wood	Gravelle	Chinese Laundry
48	Joan & David	Antonio Zaccaro	Harry and Hampton
49	Emilio Fenzi	Kelsi Dagger	Poz Poz
50	Donald J Pliner	Bella Domani	Eve Hatton
51	Revelle	Jean-Pierre Arnaud	Corso Como

SVET-Bird

Trial	Name1	Name2	Name3
1	Ochre Gabbro	Kassam Thrasher	Mountain Bluebird
2	Masked Golong	Blue Jay	Canyon Kingfisher
3	Great Mulmul	Streak-tailed Dogbird	Northern Raven
4	Purple-breasted Shrew	Olive Mohee	Barn Swallow
5	Savannah Sparrow	Tufted Gemthroat	Green Huckaloo
6 *Catch*	JCPenney	White-eyed Vireo	Tea Kettle
7	Gray-capped Woodear	Allegheny Bog Swallow	American Goldfinch
8	American Robin	Bluegrass Chickadee	Eastern Scrub Nutcracker
9	Wilmer's Cuckoo	Black-billed Cuckoo	Long-billed Rogalin
10	Holtcissel	Mountain Chickadee	Missouri Starwing
11	Blue Downbill	Cliff Swallow	Antique Sage
12	Winter Lazio	Pine Plover	Horned Lark
13	Bay Pipin	White-winged Parakeet	Liriope
14	Hooded Warbler	Cape Cod Myna	Western Kobuk
15	Northern Mockingbird	Silver-crowned Oriole	American Goldenwing
16	Long-tailed Steckle	Chesapeake Broadwing	Red-winged Blackbird
17	Rock Wren	Scoria	Oregon Groswing
18	Yellow-banded Vireo	Belted Kingfisher	Gold-collared Shortspur
19 *Catch*	American Tree Sparrow	Microwave Oven	Hyundai
20	Northern Cardinal	White-eyed Dotter	White-ringed Magpie
21	Fox Sparrow	Cascade Sparrow	Mouse Geum
22	Lapland Longspur	Orchard Spot-breast	Weigela
23	Scarlet Tanager	Blue-stripe Binbeak	Tri-colored Wheatear
24	Baltimore Oriole	Cloaked Queenbird	Broadbent's Flycatcher
25	Black-headed Peehatch	Brown-winged Digger	California Towhee
26	Hesperus Kinglark	Painted Ozark	Cassin's Kingbird
27	Thistle Grosbeak	Sage Thrasher	Wood Pennytail
28	Orange Shrub Vireo	Alder Flycatcher	Waxhaw
29	Jefferson's Bunting	Loggerhead Shrike	California Alewife
30	Warbling Vireo	Bush Moppet	Siouxland Jay
31	Brownpoll	Northern Gibbon	Gray Catbird
32	Coastal Abelia	Brown-spotted Foxtail	Tufted Titmouse
33	American Pipit	Scruffy Fletcher	Dark-horned Thrasher
34	Lark Tango	Phainopepla	Knight's Solitaire
35	Rose-throated Congaree	Brogan's Jay	Bicknell's Thrush
36	River Pointwing	Bohemian Waxwing	Dusky Nimpkin
37	Spot-breasted Pixie	McCown's Longspur	Pale-eyed Baylin
38	Kieffer Tanager	Bronze-headed Truit	Western Wood-Pewee
39 *Catch*	Reebok	Ziploc	Bullock's Oriole
40	Evening Grosbeak	Dakota Raven	Antietam
41	Yellow-eyed Junco	Kipp's Grackle	Red-throated Severne
42	Pinyon Jay	Vermilion-tipped Finch	Eastern Ruffe
43	Whiskered Thrush	Kirkland Waterthrush	Black-capped Gnatcatcher
44	Green-notched Starling	Indigo Bunting	Yellow-eyed Tatterfly
45	Missipi Kinglet	Kate's Warbler	Budgerigar
46	Violet-green Cowbird	Bobolink	Tortoise Crossbill
47	Pine Siskin	Crimson Wrenrobin	Blue Swinger
48	Emerald Mockingbird	Valerian	Eastern Phoebe
49	Grey Mountain Pinchot	Great Kiskadee	Cobbler's Oriole
50	Shiny Ridgehawk	Veery	Honeyed Manokin
51	Red-rumped Rusbin	American Treetit	Ovenbird

SVET-Leaf.

Trial	Name 1	Name 2	Name 3
1	Weeping Willow	Sweetnut	Dandelion Ash
2	River Birch	Winternut	Bronze Mountain Elm
3	Red Mountainwood	Venuswood	American Sycamore
4	Bur Oak	Green Hazel	Flowering Placket
5	Monte Cassino Oak	White Ash	Purple Watertree
6	Goldenbark Burr	Tennessee Grapin	Apricot
7	Prairie Redbark	Cat-eye Hickory	Sugar Maple
8	Japanese Maple	Orange Planterwood	Roundleaf Alder
9	Black Walnut	Yirgacheffe	Calumet Sycamore
10	Silver Firth	White Bruck	American Mountain-ash
11	Capaya	Boxelder Maple	Saranac Tupelo
12 *Catch*	Scarlet Oak	Cheese Nips	Adidas
13	Northern Winslow	Rock Elm	Wallich's Cherry
14	Pignut Hickory	Anthurium	Pittberry
15	Alstroemeria	Quaking Aspen	Yellow Oolong
16	Lily Elm	Pendleton Oak	American Beech
17	Flowering Dogwood	Yellow Cottonwood	Bristleleaf Catalpa
18	Mouse Oak	Norwegian Silkbark	Pecan
19	Yellow Poplar	California Bargo	Feather Willow
20	Rooibos	Redbud	Moon Plum
21	Victorian Poplar	Tibouren	Black Cherry
22	Cherrybark Oak	Brisco Birch	Broadleaf Dago
23	Black Brandywine	Crimson Walnut	Oregon White Oak
24 *Catch*	Springer Spaniel	Bigleaf Maple	American Airlines
25	Bigtooth Aspen	Martin's Locust	Western Tolvo
26	Dancing Ash	Cone Maple	Southern Magnolia
27	Southern Kamut	Post Oak	Red River Vosch
28	Christmas Maple	Red Alder	Pewter Oak
29	Montana Green Oak	Paper Birch	Coppernut
30	Slippery Elm	American Moffett	Meridan Whitewood
31	Sweetgum	Peruvian Hickory	Sepia
32	Mississippi Alder	Ebony Spleenwood	Blue Ash
33 *Catch*	Cadillac	Siamese Cat	Black Cottonwood
34	Coffee Gum	Horse Chestnut	Black Linwood
35	Mowamba	Black Tupelo	Sourroot
36	Delta Maidenhair	Jubilee Magnolia	Live Oak
37	Spanish Maple	Black Muscat	Eastern Cottonwood
38	Frosted Beech	Sassafras	Japanese Painted Birch
39	White Kava	Overcup Oak	American Finwood
40	Sweeney's Oak	Notched-bark Cottonwood	Honey Locust
41	Regal Poplin	Red Loden	Littleleaf Linden
42	Shellbark Hickory	Hudson Willow	Terrywood
43	Trembling Elm	Hackberry	Shiny Gum
44	Spine Oak	Sickle-leaf Willow	Ginkgo
45	Butternut	Colonial Bricktree	Honey Boxwood
46	Baldcypress	Ringed Dogwood	Littleleaf Tappan
47	Silver Aster	Valley Walnut	Tulip Poplar
48	Kentucky Coffeetree	Dixiewood	Mottlewood
49	Henwood	Water Tupelo	Jade Birch
50	Chervil	Netleaf Hackberry	Sierra Hickory
51	Yorkshire Aspen	Goldenrain	Sun Cypress

SVET-Mushroom.

Trial	Name 1	Name 2	Name 3
1	Portabello	Witches Brew	Pignoli
2	Porcini	Cabbage	Fiddle
3	Fluted Russica	Cannelle	Shiitake
4	White Truffle	Milky Scaber	Sugar Siullus
5 *Catch*	Dishwasher	Cauliflower	Taco Bell
6	River Vervain	Red-capped Scaber	Fern
7	Wood Ear	Molasses	Steely Wood
8	Vinegar	Black Saddle	Barrel
9	Cardoon	Bunny Ear	Green-spored Parasol
10	Black Perigord Truffle	Cat's Paw	Bachilucium
11	Fan	Russell's Redfoot	Wine-cap Stropharia
12	Zeller's Bolete	Camel	Death in the Afternoon
13	Black Tollius	Bleeding Plovit	Black Trumpet
14	Sea String	Mountain Puff	Matsutake
15	Burgundy Top	Morel	Hiziki
16	Midoni	Crimini	Scarlet Tulip
17	Painted Bark	Brown Shandy	Death Cap
18	Udupi	Cinnamon Cap	Mauricus
19	Chanterelle	Globe	Mozuku
20	Button	Snowcap	Beaver Tooth
21	Cipolini	Teddy Bear	Reishi
22	Horn-toothed Bolete	Shaggy Parasol	Habutai
23	Pig's Ear	Mouse of the Woods	Raven Claw
24 *Catch*	Macy's	Velveeta	King Bolete
25	Amber Stalk	Tavel	Enoki
26	Cognac	Cloud Ear	Sousaire
27	Portalo	Gombe	King Trumpet
28	Urikandji	Canopy	Velvet Foot
29	Oyster	Potelle	Ten Penny
30	Shiso	Palm	Bleeding Milkcap
31	Hen of the Woods	Egoji	Conch
32	Tarutake	Candy Cap	Alsace Brown
33	Sun-dotted	Bear's Head	Tri-colored Culotte
34	Courgette	Golden Needle	Chandelier
35	Parkeo	Straw	Giblet
36	Smoke	Salmon	Honey
37	Green Cap	Ballast	Old Man of the Woods
38	Petaluma	Paddy Straw	Bogie
39	Fawn	Starburst	Midnight Korme
40	Clam	Shaggy Mane	Royal Gilded
41 *Catch*	Field	Hershey's	Google
42	King's Head	Ivory Plume	Yellowfoot
43	Blue Foot	Gnome's Hat	Bobbin
44	Cassava	Angel Wings	Patapan
45	French Tardis	Hedgehog	Summer Cobalt
46	Crab Brittlegill	Elephant Trunk	Glass Cap
47	Fontanelle	Birch Bolete	Spring Fiori
48	Willow Ash	Diving Bell	Sweet Tooth
49	Fairy-ring	Ruffle Cap	Pag Lace
50	Satin Top	Harutake	Slippery Jack
51	Horse	Dotted Pin	Jester

Appendix B. Extended bird-specific experience questions for birders used in Study 2C. Note that for the order of the responses for question 6 are reversed from the others and so responses were adjusted before analysis.

1. At what age did you first develop an interest in birds?
__(Age)

2. At what age did you first start birding relatively seriously (e.g., spending time learning bird identifications, going on planned bird walks, joining local Audubon or ornithological societies, etc.)
__(Age)

3. How often do you go birding (specifically set aside time for bird watching at home or elsewhere)?
_ Less than once a year
_ 1-3 times per year
_ 4-6 times per year
_ 7-12 times per year (every 1-2 months)
_ 13-24 times per year (1-2 times per month)
_ 25-48 times per year (every 1-2 weeks)
_ 49 or more times per year (several times each week)

4. How often do you travel outside of your region (more than 1 hour travel time from your home), at least in part, for specific bird watching opportunities?
_ Almost never
_ 1 time per year
_ 2-3 times per year
_ 4-6 times per year
_ 7 or more times per year

5. How often have you planned a vacation with a primary intent of birding, on average?
_I am a professional who regularly identifies birds (e.g., ornithological research, photographer, tour leader, educator, wildlife resource manager)
_ More than once a year
_ Once a year
_ Every other year
_ Once every few years
_ Rarely or never

6. Do you keep a log (journal, online list, etc.) of birds that you see?
_ Never
_ Sometimes
_ Almost always

7. About how many different types of birds (specific species or subspecies) have you observed in person while birding during your lifetime?
__(Number)

8. How would you rate your own bird expertise for birds where you live?
_ I am a novice. Nearly all other birders I meet are more skilled than I am.
_ I am a beginner. Most birders I meet are more skilled than I am, but I occasionally meet other beginners like me when out birding.
_ I have intermediate birding skills. While there are many birders more skilled than I am, I can identify many birds that beginners cannot.
_ I have advanced birding skills. While I am not the most expert birder that I know in my area, I often identify birds quicker and more accurately than others.
_ I have expert birding skills. While not a professional, I often lead birding trips for my local birding societies, organize local bird counts, etc.
_ I have expert birding skills. While I have met some people who are more expert than I am, I have done things like lead birding tour groups professionally, conduct ornithological research, educate about bird identification and bird conservation, or work in wildlife management.
_ I have expert birding skills. I am recognized by my peers in my state, nationally, or internationally as someone other experts would turn to because of my expertise.

9. How many birding periodicals (magazines, newsletters, journals) do you subscribe to?
__(Number)

10. How many local, national, or international birding organizations do you belong to (groups involved in planning or tracking bird sightings, science of birds, bird identification, formal groups of bird enthusiasts, etc.)
__(Number)

11. How often do you attend birding events, conferences, or meetings with other bird enthusiasts?
_Almost never
_1-3 times per year
_ 4-6 times per year
_ 7-12 times per year (every 1-2 months)
_13-24 times per year (1-2 times per month)
_ 25-48 times a year (every 1-2 weeks)
_ 49 or more times per year (several times each week)