Journal of Experimental Psychology: General
2010, Vol. 139, No. 3, 558–578

0096-3445/10/$12.00 DOI: 10.1037/a0019165

Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects

Talia Konkle and Timothy F. Brady
Massachusetts Institute of Technology

George A. Alvarez
Harvard University

Aude Oliva
Massachusetts Institute of Technology

Humans have a massive capacity to store detailed information in visual long-term memory. The present studies explored the fidelity of these visual long-term memory representations and examined how conceptual and perceptual features of object categories support this capacity. Observers viewed 2,800 object images with a different number of exemplars presented from each category. At test, observers indicated which of 2 exemplars they had previously studied. Memory performance was high and remained quite high (82% accuracy) with 16 exemplars from a category in memory, demonstrating a large memory capacity for object exemplars. However, memory performance decreased as more exemplars were held in memory, implying systematic categorical interference. Object categories with conceptually distinctive exemplars showed less interference in memory as the number of exemplars increased. Interference in memory was not predicted by the perceptual distinctiveness of exemplars from an object category, though these perceptual measures predicted visual search rates for an object target among exemplars. These data provide evidence that observers' capacity to remember visual information in long-term memory depends more on conceptual structure than perceptual distinctiveness.

Keywords: perceptual distinctiveness, object recognition, categorization, memory capacity, memory fidelity

The human brain can store a large number of objects, events, words, and pictures, often after only a single exposure. In a landmark study of visual long-term memory, Standing (1973) showed people 10,000 images and found memory performance in a recognition memory task was quite high (83%). Together with other studies (Shepard, 1967; Standing, Conezio, & Haber, 1970), these results demonstrated the existence of a massive capacity long-term store for visual information. However, fundamental questions were left unexplored by these seminal studies: How detailed are the visual representations? How are thousands of memory representations organized and stored so that they can be successfully retrieved at a later point in time? In the current work we begin to address these questions by examining the contributions of conceptual and perceptual distinctiveness to long-term memory for visual objects.

The Fidelity of Long-Term Memory for Visual Information

To estimate the capacity of any storage system, one must evaluate not only how many items can be remembered (*quantity*) but also how much is remembered per item (*fidelity*). The work of Standing (1973) and others made it clear that observers can successfully remember thousands of pictures after only a single exposure. However, the representations of the images in these large-scale memory experiments were assumed to be sparse in detail: Because all the images were chosen to be visually and categorically distinct, observers could succeed at the two-alternative forced-choice memory task by using a gist-like representation—for example, one that captured only the basic category or meaning of the image (Chun, 2003; Simons & Levin, 1997; Wolfe, 1998). Only recently have studies systematically probed the level of detail or fidelity with which hundreds or thousands of image representations are maintained in long-term memory (Brady, Konkle, Alvarez, & Oliva, 2008; Hollingworth, 2004; Vogt & Magnussen, 2007). All of these studies demonstrated that observers could recognize a significant amount of detail about each object or image.

Talia Konkle, Timothy F. Brady, and Aude Oliva, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; George A. Alvarez, Department of Psychology, Harvard University.

This work was partly funded by a National Defense Science and Engineering Graduate Fellowship to Talia Konkle, a National Science Foundation Graduate Research Fellowship to Timothy F. Brady, National Research Service Award Fellowship F32-EY016982 to George A. Alvarez, and National Science Foundation Career Award IIS-0546262 to Aude Oliva. We thank M. Greene, B. Hidalgo-Sotelo, A. Hollingworth, K. Nakayama, M. Ross, A. Schachner, A. Torralba, and J. Wolfe for helpful conversation and comments on the article and E. Vul for help with the test–retest reliability simulations. The stimuli database may be found on our website at http://cvcl.mit.edu/MM

Correspondence concerning this article should be addressed to Talia Konkle, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: tkonkle@mit.edu, tfbrady@mit.edu, alvarez@wjh.harvard.edu, or oliva@mit.edu

In a study by Hollingworth (2004), for example, observers remained significantly above chance at remembering which exemplar of an object category they had seen after studying about 100 objects embedded in a scene context, even with up to 400 intervening objects in between study and test. This suggests that which particular object exemplar you saw—not just the category of the object—can successfully be stored for a hundred or more objects in visual long-term memory. In Brady et al. (2008), observers saw thousands of categorically unique objects, each presented once, over the course of a 5.5-hr study session. Observers' memory was then tested using a two-alternative forced-choice task that pitted a previously seen object against a different object image. Critically, the foil image could either be an object from a different, novel category (requiring observers to retain only basic-level category information to succeed, similar to Standing, 1973), an object from the same category (requiring observers to know which specific exemplar from that category they had seen), or the same object, but in a different configuration or pose (requiring observers to remember which state the object was in). Surprisingly, observers succeeded at even the more subtle discriminations, achieving 87% accuracy in both the state and exemplar conditions. Together with the previous results, this suggests that visual long-term memory can store a large number of items with substantial detail, at least under the experimental conditions that were tested in these studies.

Categorical Distinctiveness in Visual Long-Term Memory

The items to be remembered from Standing (1973) were gathered from various magazines, and thus presumably reflected everyday scenes and events. However, Standing also observed that long-term memory capacity for vivid pictures, images that depicted oddities (e.g., a dog holding a pipe, a crashed airplane), was even better: After seeing 1,000 vivid pictures, Standing calculated that observers remembered 992 of them. This result suggests that *distinctiveness* is a doorway to long-term storage and retrieval: Memories in a sparsely populated position within the space of representation are more likely to be retrieved than memories embedded in a dense space of competing memories (e.g., Eysenck, 1979; Nairne, 2006; Rawson & Van Overschelde, 2008; Schmidt, 1985; von Restorff, 1933).

Even when nonvivid images were used in the previous large-scale memory studies (Brady et al., 2008; Shepard, 1967; Standing, 1973; Standing et al., 1970), one salient aspect of these studies is that the thousands of images presented to observers were distinctive from each other in at least one important way: They were largely categorically unique. For instance, in Brady et al. (2008), the 2,500 objects presented for encoding were from different basic-level or entry categories (e.g., one bike helmet, one couch, one coffee mug, etc.). Thus it is possible that the capacity to retrieve enough information to succeed at subtle within-category memory tests (e.g., Brady et al., 2008) depends on the categorical distinctiveness of the items encoded into memory.

A study by Vogt and Magnussen (2007) brings interesting data to bear on this point. Vogt and Magnussen observed that people could tell apart an image of a new door from a previously studied door with about 85% accuracy after studying over 400 images of doors with surrounding background information. This suggests that even with 400 images from the same category, observers can successfully recognize details about each image. However, in an important follow-up manipulation, Vogt and Magnussen showed that observers were not remembering the doors, per se, but local visual details: When the images were edited to remove other objects and nondoor details from the scene (e.g., light fixtures, window signs, structures on the surrounding walls), performance dropped by 20%. In other words, when asked to remember only the doors with nondoor details removed, people were not as successful. Observers' memory performance instead seemed to depend on encoding distinctive details about the background of each of the doors.

Perceptual Versus Conceptual Distinctiveness

There are many reasons to suppose that categorical uniqueness is particularly important for supporting such high memory performance in visual long-term memory. For example, it is widely accepted that categories are the organizing structure of long-term knowledge in the verbal domain (e.g., by spreading activation models; J. R. Anderson, 1983; Collins & Loftus, 1975), with some privileged status for basic-level categories (Mervis & Rosch, 1981) or entry-level categories (Jolicoeur, Gluck, & Kosslyn, 1984). This suggests that a unique basic-level category may be a powerful retrieval cue, preventing interference from other perceptually similar memories. Classic studies also demonstrate the critical role of semantic organization in verbal memory; for example, details of a story that are consistent within an existing schema are more likely to be remembered than those which are not (e.g., R. C. Anderson & Pichert, 1978; Bransford & Johnson, 1972). Semantic information can also lead to systematic errors; for example, when recalling word lists, semantically related nonpresented words will be falsely remembered (Deese, 1959; Roediger & McDermott, 1995). These kinds of results provide further evidence that conceptual knowledge is an organizing principle for the storage and retrieval of information in memory.

However, our understanding of the role of conceptual knowledge for organizing and retrieving information from long-term memory has predominantly been derived from studies that examine memory for text or verbal stimuli. When considering long-term memory for visual information, perceptual features are likely to play a more central role in long-term memory. Object categories have a range of perceptual features: dominoes all have the same shape, whereas leaves have very different shapes; pumpkins have similar colors, whereas stamps have very different colors. Are these perceptual dimensions critical for organizing and retrieving memory for visual object information? One reason to believe this is the case is that the nature of the content to be remembered is perceptual (images): Attending to the perceptual features of words (the case or font of text, the gender of a speaker's voice) may not help memory for the words, but similar "surface-level" attention might be critical for visual long-term memory.

On the other hand, we might not expect a strong role for perceptual distinctiveness in memory because objects can be recognized with line drawings, in different lighting conditions, and from different viewpoints (e.g., Biederman, 1987), suggesting the underlying representations may not necessarily require perceptual details for accurate retrieval. Further, object categories have a range of subordinate category structure; for example, there are many kinds of cookies and only a few kinds of bowties. This subcategorical structure arises from knowledge and experience and does not necessarily reflect the perceptual features of the objects. As more items from a category are loaded into memory, how do perceptual and conceptual dimensions of variation matter for retaining detailed object representations in memory?

The Current Study

Here we present a series of experiments that explore the fidelity of our visual long-term memory and examine which stimulus dimensions support detail in visual long-term memory representations. Understanding which dimensions predict memory performance constrains models of encoding and retrieval and enables us to make inferences about the underlying organization of visual long-term memory. In Experiment 1, observers viewed thousands of objects with a variable number of exemplars from each category. If categories matter for visual long-term memory, then the more items per category are studied, the worse memory performance should be.

In Experiments 2 and 3 we obtained estimates of the conceptual and perceptual distinctiveness of the exemplars within an object category to examine the contributions of such distinctiveness to memory interference. We find that the conceptual distinctiveness within an object category predicts the difficulty of remembering multiple exemplars within that category. Surprisingly, the perceptual distinctiveness within a category does not predict interference in long-term memory. Finally, in Experiment 4 we show that our measures of perceptual distinctiveness predict visual search times using the same stimuli, demonstrating that the lack of correlation between perceptual distinctiveness and memory interference is not due to invalid ranking measures. Our results, showing a double dissociation between the perceptual and conceptual contributions to visual long-term memory and visual search, support the conclusion that conceptual distinctiveness plays a more prominent role in supporting detailed representations in visual long-term memory than perceptual distinctiveness.

Experiment 1: Interference in Visual Long-Term Memory

In nearly all previous large-scale studies investigating the capacity of visual long-term memory (Brady et al., 2008; Standing, 1973), the items presented were categorically unique—for example, a single hairdryer, a single coffee mug, and so on. Thus, it is possible that the capacity to retain detailed information about thousands of visual objects in long-term memory relies on having encoded categorically distinctive items.

In the present experiment we examined this question by showing observers thousands of objects, where we varied the number of exemplars seen from each category from one exemplar to 16 exemplars (see examples of object categories in Figure 1). For each of these conditions, we tested memory by presenting a new exemplar and a previously studied exemplar, and observers indicated which item they viewed during the study session. This

Figure 1. Sixteen categories of object exemplars, sampled from the set of 200 categories used for the recognition memory task. The full database of stimuli is available from our website.

allowed us to examine the role of categorical interference in memory: If observers can remember 16 exemplars from a category as well as they can remember one exemplar from a category, then memory performance across these conditions will be equal. This pattern of data would suggest that retaining detailed object representations does not depend on the categorical distinctiveness of the items in the set. Alternatively, if memory performance decreases systematically as more and more exemplars from a category are loaded into memory, then this pattern of data would suggest an important role for categorical distinctiveness in visual long-term memory.

Method

Participants. Eighteen adults (aged 20–35) gave informed consent and received a compensation of $100 for participating in the 6-hr experiment. All participants had self-reported normal or corrected-to-normal vision. All of the participants were tested simultaneously, using computer workstations that were closely matched for monitor size and viewing distance.

General procedure. The general procedure is illustrated in Figure 2. In the study phase, observers were presented with 2,800 color images of real-world objects. Each image (subtending 7.5° × 7.5° visual angle) was presented for 3 s, followed by an 800-ms fixation cross. Observers were instructed that they would be presented with a stream of thousands of objects and that the task was to "remember them all." While studying the items, observers additionally performed a *repeat detection task* to maintain focus. For this task they were told to press a button to indicate if the current item had been presented previously in the study stream. The study session was broken up into 10 blocks of 20 min each in which 280 images were shown. Between blocks, participants were given a 5-min break and were not allowed to discuss any of the images they had seen. Halfway through the study session, a 20-min break was given.

At the end of the study session, observers were given a 10-min break and then participated in the test phase, in which they completed a *recognition memory task*. Two items were presented on the screen—one previously seen old item and one new foil item—and observers reported which item they had seen before in a two-alternative forced-choice task. Observers completed 240 forced-choice tasks: 200 exemplar-level tests and 40 in a baseline novel category condition. They completed these trials at their own pace and were asked to emphasize accuracy, not speed, in making their responses. Before starting the experiment, we explained to observers exactly how their memory for the items in the study phase would be tested and presented several examples of the kinds of two-alternative forced-choice tasks to expect. None of these example object images or categories were used in the subsequent experiment.

Recognition memory task. To probe the role of categorical interference on recognition memory, we systematically varied how many exemplars were presented in the study phase for each category. For any given object category, observers studied either 1, 2, 4, 8, or 16 exemplars during the study phase. For each observer, 40

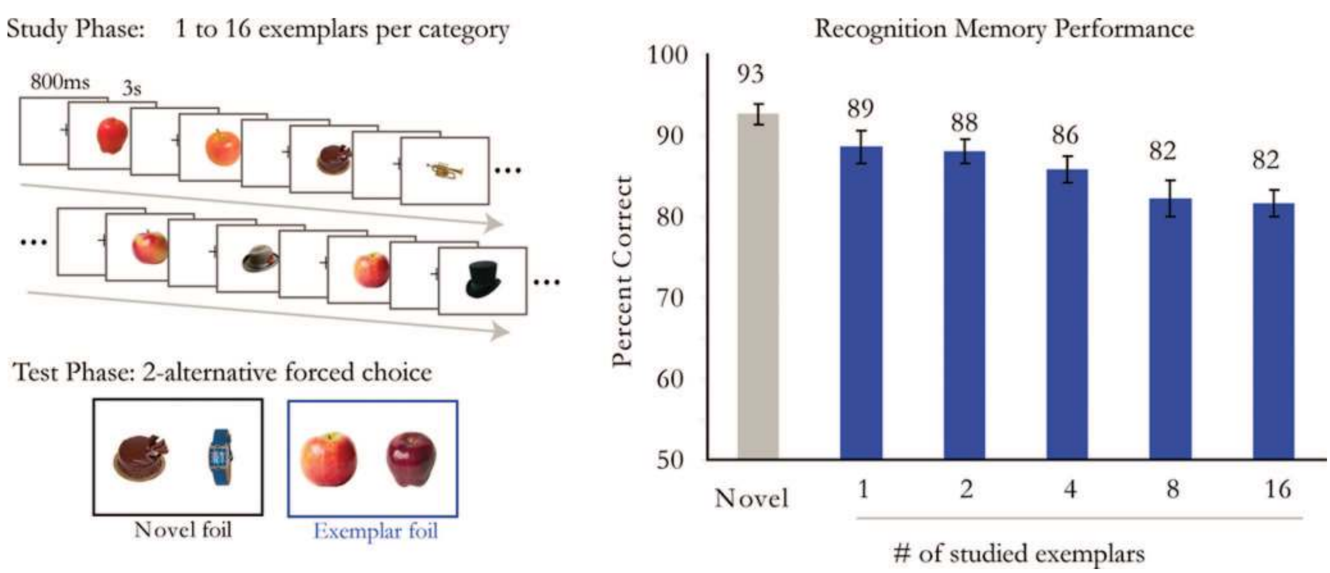

Figure 2. Left panel: Methods. During the study phase, 2,800 images were presented, one at a time, for 3 s each, with an 800-ms fixation between images. The number of exemplars presented from a given category was varied from one to 16. At test, two images were presented on the screen, and observers had to indicate which object they had seen during the study phase. In the novel condition, the foil image was a categorically distinct item, and in the categorical interference conditions, the foil item was a new exemplar from the same category. Right panel: Recognition memory performance. When observers viewed a singleton item from a category and were tested against an item from a distinct category (novel condition), memory performance was 93%. When a singleton item from a category was tested against an exemplar foil (one studied-exemplar condition), performance was 89%. Recognition memory performance decreased approximately 2% for each doubling of exemplars in memory. Error bars reflect ±1 *SEM*.

object categories were assigned to each interference condition, such that across observers each category was used in each condition. This means that for all 200 object categories used in the exemplar-level recognition memory task, the category was presented equally often with 1, 2, 4, 8, or 16 exemplars. All the images from these 200-object categories were distributed uniformly throughout the study phase.

The exemplar-level tests contained one previously seen old item and one new foil item from the same category. Because a variable number of exemplars from a category could be presented during the study phase, we always tested memory for the first item presented from each category. This helped ensure that all tested items were at “equal ground” at the time of encoding. Further, because the remaining object exemplars from that category were presented after the tested item, we are examining the role of retroactive interference: Any decrease in memory performance as the number of studied exemplars per category increases is due to interference from subsequently presented items. Participants were not aware of this manipulation or of the number of exemplars per category.

Importantly, the stimuli used as the test and the foil image in the two-alternative forced-choice task were the same for each observer. To choose these, for each category a pair of two exemplars was drawn randomly from the 17 available exemplar images. One of these exemplars was presented during the study stream, and the other was only viewed as the foil at the memory task. Which image served as the foil or the studied item was counterbalanced across observers.

Because all of the memory tasks were the same for all observers, the only varying factors across observers were (a) which item of the test–foil pair was presented during the study phase and (b) how many other exemplars were presented within that category. Because all categories were seen in all conditions, across observers the stimulus set is completely counterbalanced. This means than any effects on performance at the recognition tests are due to interference from studied exemplars, and cannot be driven by test–foil pairings that may be easy or more difficult, or any overall ease or difficulty of particular object categories.

We also included a baseline memory task condition, consisting of a studied item and a foil of a different basic-level category (a *novel* condition, as in Standing, 1973, and Brady et al., 2008). Forty objects with distinct categories were presented in the study session. At test these were paired with 40 other objects, from categories distinct from all other object categories seen during study phase. As before, the test pair items were always the same, and which novel item was presented in the study stream was counterbalanced across observers. To choose these, the 80 items from 80 distinct categories were randomly assigned to make 40 test pairs.

Repeat detection task. To maintain attention and to probe memory online, participants performed a repeat detection task during the study phase. Observers were told that an item could be repeated from any point in the entire study session, and they responded to exact repeat items by pressing the spacebar. They were given feedback only when they responded, with the fixation cross turning red if they incorrectly pressed the space bar (false alarm) or green if they correctly detected a repeat (hit), and were given no feedback for misses or correct rejections. Of the 2,800 images shown in the study session, 240 were repeated images (e.g., on average about one in every 12 items was a repeat image, or one every ~40 s). Which particular image repeated was randomized across observers. All object categories used for the repeat detection task were distinct from the object categories used for the recognition memory task.

Unbeknownst to the participants, the repeat images were systematically inserted to test both the impact of categorical interference and the impact of the number of intervening items. To probe the role of categorical interference in repeat detection, repeated items could come from categories in which 1, 2, 4, 8, or 16 exemplars had been previously viewed. The repeated image was always the last item of that category to be shown. Thus, we can measure repeat detection performance as a function of number of preceding within-category items, examining the role of proactive interference. There were 40 repeat trials for each condition of preceding exemplars (16, 8, 4, 2, 1) with an additional 40 trials for the one-exemplar condition, yielding a total of 240 repeats.

Additionally, we manipulated how many intervening items occurred between the repetition of an image, from 1, 15, 63, 255, to 1,023 intervening items. Repeat items were inserted into the stream uniformly, with the constraint that all of the lengths of *n*-backs (2-back, 16-back, 64-back, 256-back, and 1,024-back) had to occur equally in the first half of the experiment and the second half. This ensured that fatigue would not differentially affect images that were repeated from further back in the stream. Of the 240 repeat trials, 48 happened after one item, 48 happened after 15 items, and so forth. This manipulation was crossed with the categorical interference manipulation. Thus, for example, each of the 48 repeats with one intervening item was preceded by a variable number of exemplars from the same category, from one to 16. Specifically, there were eight repeats for each categorical interference condition $\times$ *n*-back length, with eight additional repeats for all *n*-back conditions with one exemplar.

Stimuli. The total image database contained 4,760 different images of single objects, from 520 different object categories. A set of 200 object categories, each with 17 exemplars (3,400 images), was used for the recognition memory task. This allowed a full counterbalance of these items across observers, with up to 16 presented in the study phase, and one more reserved as the foil item in the test phase. Additionally, 80 distinct categories with one exemplar each were used for the novel condition in the recognition memory task (80 images). The remaining 240 object categories were used in the repeat detection task, in which we did not perform a full counterbalance across stimuli. Therefore the same categories were seen with 16, 8, 4, 2, or one exemplar(s) by each observer. There were 40 categories with 16 exemplars, 40 categories with eight exemplars, 40 categories with four exemplars, 40 categories with two exemplars, and 80 categories with one exemplar (1,280 images).

Stimuli were gathered using both a commercially available database (Hemera Photo-Objects, Vol. I & II) and Internet searches using Google Image Search. The 200 categories used in the memory recognition task were the first 200 categories for which we could obtain 17 exemplars. For each category, care was taken to try to span the variety of exemplars, colors, and shapes that existed in the category, but all exemplars within a category were chosen to have a similar viewpoint. Figure 1 illustrates some example sets. The full database of stimuli is available on our website.

Data analysis. All analyses of variance (ANOVAs) reported are 1 × 5 repeated measures unless otherwise indicated, and effect sizes are reported as eta-squared (η^2), which can be interpreted as the proportion of variance accounted for by the independent variable (iv): $\eta^2 = SS_{iv}/(SS_{iv} + SS_{error})$. For d' calculations, perfect hit rates were adjusted to .995, and zero false alarm rates were adjusted to .005.

Results

Recognition memory task. The results of the forced-choice memory task are shown in Figure 2. When a singleton item in the study stream was tested against a novel category at test (novel condition), memory performance was 92.6% correct (*SEM* = 1.3%), consistent with previous findings of visual long-term memory requiring only memory for the item's category (Brady et al., 2008). When a singleton item was tested against an exemplar foil (one-exemplar condition), memory performance was 88.6% (*SEM* = 2.0%). There was a difference that approached significance in performance between novel and one-exemplar conditions: In other words, when a single item from a distinct category was studied, memory performance for such an item is slightly higher when tested against a foil from a different basic-level category than when tested against an exemplar-level foil, $t(17) = 2.03$, $p = .06$. These data replicate the results of Brady et al., 2008, and also demonstrate that the ability to remember a categorically unique item (at both the novel and exemplar level) is not affected by the presence of many exemplars of other object categories within the study stream.

Next, we examined the impact of subsequently presented items from the same category. As the number of exemplars presented in the study stream increases, memory at the subsequent two-alternative forced-choice task decreases, $F(4, 68) = 6.46$, $p < .001$, $\eta^2 = .28$. We calculated the slope of memory performance as a function of the $\log_2$ number of exemplars for each observer and use this as a measure of interference attributable to the number of exemplars from a category in memory (average interference = –2.0%, *SEM* = 0.4%). Surprisingly, the degree of interference from multiple exemplars was minimal, averaging only a 2% drop in memory performance with each doubling of the number of the exemplars in memory: Even with 16 exemplars from a category in mind, observers could still distinguish one of those from a new 17th exemplar with 82% accuracy. Thus, overall, the memory performance in this task was quite high. However, this interference slope was significantly different than zero, demonstrating that there was minimal, but reliable, interference, $t(17) = 4.6$, $p < .0001$.

Although observers were told to emphasize accuracy, not speed in their unpaced memory task trials, we also recorded reaction times (RTs). Correct responses had a mean RT that was 1.2 s faster than the mean RT for errors (correct RT = 2.5 s, incorrect RT = 3.7 s), $t(17) = 5.10$, $p < .001$. Additionally, when only a single exemplar of a category was studied, correct responses were faster when the foil item was from a novel category than when it was another exemplar from the same category (novel condition RT = 2.2 s, one-exemplar condition RT = 2.3 s), $t(17) = 2.06$, $p = .05$. RTs for correct responses increased monotonically as the number of exemplars in memory increased, from 2.3 s in the one-exemplar condition to 2.7 s in the 16-exemplar condition [RT = 2,300 ms + 106 ms*$\log_2$(exemplars)], $F(4, 68) = 9.45$, $p < .001$, $\eta^2 = .36$. Incorrect RTs were not different as a function of the number of studied exemplars, $F(4, 68) < 1$, *ns*, $\eta^2 = .03$.

Repeat detection performance. We first examined the effect of number of intervening items between the item and its repetition, collapsing across how many exemplars of the same category were previously studied. Sensitivity (d') for repeat detection was calculated for each condition, which takes into account hit rates corrected by the overall false alarm rate. As the number of intervening items increases, repeat detection performance decreases systematically, $F(4, 68) = 43.5$, $p < .001$, $\eta^2 = .72$. Performance on the repeat detection task during the study session is shown in Figure 3 (left panel). Note that in these data, we cannot distinguish between effects due to the amount of elapsed time and effects due to the number of total items intervening.

The corresponding false alarm rate and hit rates for these data are as follows: 95% hits for one intervening item, 94% hits for 15 intervening items, 87% hits for 63 intervening items, 79% hits for 255 intervening items, and 63% hits for 1,023 intervening items, with an overall false alarm rate of 3.9% (*SEM* = 0.5). With one

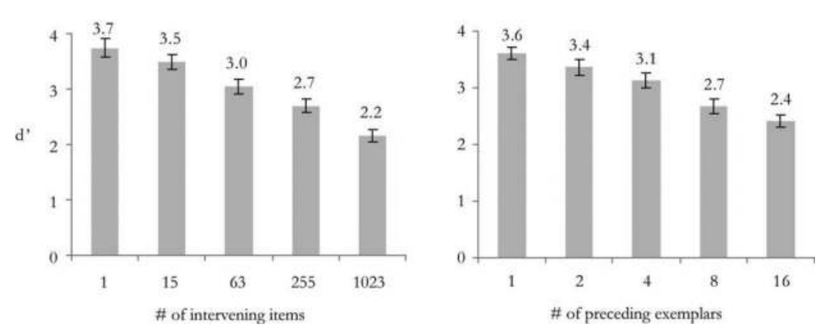

Figure 3. Repeat detection performance. Left panel: Sensitivity for detecting repeated images in the stream, as a function of how many intervening items appeared between the first presentation of the image and the repeat presentation. Note that these data could be due to the amount of elapsed time between repetitions or the number of intervening items. Right panel: Sensitivity for detecting repeated images in the stream, as a function of the number of prior exemplars presented before the first presentation of the to-be-repeated image. All error bars reflect ±1 *SEM*.

intervening item, the hit rate was 95%, suggesting that people were maintaining focus throughout the entire 5.5-hr study session. Further, even for repeats with 1,023 intervening items, in which the item was initially presented over 2 hr before it was repeated, repeat detection performance was still high (63% hit rate, $d' = 2.2$). Finally, people rarely had false alarms, with on average 100 false alarm responses out of 2,560 possible correct rejections.

We next examined repeat detection performance as a function of the number of preceding items of the same category, collapsing across the number of intervening items between initial presentation and repeat presentation. There was a significant effect of the number of preceding within-category items on d': As the number of preceding items within the category increases, the repeat detection sensitivity decreases, $F(4, 68) = 69.0$, $p < .001$, $\eta^2 = .80$. Measures of sensitivity are shown in Figure 3 (right panel). Again, overall the repeat detection performance was high (all d' scores > 2.0). Hit rates were 84% for one preceding exemplar, 88% for two preceding exemplars, 85% for four preceding exemplars, 81% for eight preceding exemplars, and 79% for 16 preceding exemplars; this decrease is reliable, $F(4, 17) = 12.2$, $p < .001$, $\eta^2 = .42$. The false alarm rate was calculated such that only false alarms to the items of the same category were counted; they were 0.2%, 2.5%, 2.7%, 4.4%, and 6.4%, respectively, for the 1-, 2-, 4-, 8-, and 16-preceding exemplar conditions; this increase is also reliable, $F(4, 17) = 9.4$, $p < .001$, $\eta^2 = .36$. Finally, observers' overall performance in the recognition memory task was positively correlated with their overall performance in the repeat detection task ($r^2 = .61$, $p < .001$).

Discussion

Recognition memory task. Overall, memory performance at test was high: Even after spending 5 hr viewing 2,800 objects—including 16 different backpacks, 16 different binoculars, 16 different apples, and so on—observers could discriminate between old and new exemplars 82% of the time. These data demonstrate that observers are remembering more than a gist-like representation in these large-scale memory studies, and they are consistent with the results of Brady et al. (2008) showing high-fidelity memory representations (see also Koutstaal & Schacter, 1997).

These data also suggest that categorical organization has a strong effect on the maintenance of detailed visual long-term memory representations. To place the effect of categorical interference in perspective, compare the present results to the landmark findings of Standing (1973; Standing et al., 1970). Standing et al. (1970) presented observers with 2,500 items and found 91% correct recognition performance for items foiled against a novel item, quite similar to the present study (2,800 items, 93% correct). Standing (1973) then increased the number of items to remember dramatically: After studying 10,000 items, observers' performance dropped to 83% correct. This suggests that the addition of 7,500 items in Standing's studies caused performance to drop from 91% to 83%. It is interesting that, in the present experiment, the addition of only 16 items of the same category caused a drop in performance for our exemplar-level foils from 89% to 82%. This drop is similar in magnitude to the addition of 7,500 new categorically distinct items in Standing's studies. Of course, in both cases, it is important to point out that memory performance at a level of 83% correct after 10,000 items in the case of Standing or 16 exemplars and 2,500 total unique items is still quite high.

The present results suggest that category information is an organizing principle in visual memory, showing strong support for the idea that categorical uniqueness can prevent items from interfering with each other in visual long-term memory.

Repeat detection task. The primary purpose of the repeat detection task was to require a sustained task so participants maintained focus throughout the study phase. However, the accuracy and RT data also allow us to probe online memory performance with a low-prevalence old–new recognition task (rather than a two-alternative forced-choice task in the memory task). Additionally, the old–new discrimination performance in the present study required exemplar-level precision in memory (whereas in Brady et al., 2008, only categorical precision was needed to succeed at the repeat detection task). Overall, performance was high, with observers at 63% hits ($d' = 2.2$) when the item repeated over 2 hr previous, and at 79% hits ($d' = 2.4$) when the item was preceded by 15 other exemplars from the same category before it was repeated.

These data suggest a strong effect of categorical organization on visual memory: As the number of preceding exemplars increased from one to 16, d' dropped. This drop was comparable to the drop that occurred when the total number of intervening items increased from one to 1,023. Thus, remembering 16 exemplars from the same category caused about as much interference as remembering over 1,000 objects of different categories. These data reinforce the idea that categorical organization plays an important role in the storage and retrieval of information in visual long-term memory.

Although the purpose of study was not to distinguish between different models of item-recognition memory, there are systematic effects in the hit rate and false alarm rates that constrain models of memory retrieval (e.g., global-matching retrieval models; McClelland & Chappel, 1998; Shiffrin & Steyvers, 1997). For example, we observed that false alarm rates increased as more exemplars were already in memory, presumably because any given new item is more likely to falsely match an old item (e.g., Koutstaal & Schachter, 1997). In other words, when presented with a new image of a book, an observer may be unsure if she has seen this particular book before, but because she has seen many books, she may be more likely to believe this was one of them. This would lead to more false alarms with more exemplars studied. However, this same logic also predicts observers should get more hits, because repeated items will also benefit from having many other categorically similar items studied (e.g., see Shiffrin, Huber, & Marienelli, 1995). In our data, however, we observed that the hit rate decreased slightly with more exemplars in memory.

Why should hit rates decrease here—for example, why should one be less accurate in saying "yes that's a book I saw already" if one has seen that book and 16 others? One possibility is that the presence of many book representations in memory degrades the trace of a new book via proactive interference, leading to a worse match when that book is repeated. An alternate possibility is that there is no trace degradation but that the decrease in hits results from noise during the retrieval process. For example, if the current book leads to the retrieval of the proper stored representation as well as the representation of other similar books, then the trace activated during retrieval would be a poor match to the current item's representation, leading to fewer hits. Although the current data do not distinguish between these hypotheses, they do constrain existing global-matching models.

The memory task and the repeat detection task let us probe, respectively, the impact of retroactive and proactive interference. The items tested in the two-alternative forced-choice memory task were always the first of their category to be presented in the study stream, with additional exemplars presented afterwards. Thus the decrease in recognition memory performance with increasing exemplars from a category was caused by subsequently presented items, or retroactive interference. Conversely, in the repeat detection task performed during the study stream, the repeated item was always the last exemplar presented from an object category. Thus the decrease in repeat detection as a function of increasing exemplars was caused by interference from preceding items, or proactive interference. Although the present data do not allow us to examine the relative impact of these two kinds of interference, they do demonstrate that visual memory representations are subject to interference by both preceding items and following items that are categorically similar.

Experiment 2: Distinctiveness Rankings and Visual Memory

Experiment 1 suggests that the more *categorically distinct* the thousands of items to be remembered are, the better the memory performance observed in the forced-choice task. However, increasing the number of exemplars presented for each category effects both how conceptually distinct an item is in the set of to-be-remembered items and also how perceptually distinct an item is, assuming that on average two items from the same category are more perceptually similar than two items from different categories. In Experiments 2 and 3, we aimed to understand what is contributing to within-category interference in memory by examining both the conceptual and perceptual distinctiveness of items within a category.

By *perceptual distinctiveness*, we are specifically referring to dimensions of perceptual variation that can be assessed for any object category (e.g., color, shape), without any existing knowledge about what an object is. Beyond these basic dimensions, perceptual features are likely to be category specific (e.g., the curvature of the mug relative to the curvature of the handle) and to be acquired through visual experience (e.g., Schyns, Goldstone, & Thibaut, 1998), and thus begin to blur the line between perceptual and conceptual features. We chose to focus on shape and color dimensions for several reasons. First, evidence suggests that there are neural mechanisms in the visual cortex that are dedicated to processing these object properties (e.g., for color, see, Bartels & Zeki, 2000; Brouwer & Heeger, 2009; Conway, Moeller, & Tsao, 2007; Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998; for shape, see, Malach et al., 1995; Grill-Spector, Kushnir, Edelman, Itzchak, & Malach, 1998; Kourtzi & Kanwisher, 2000, 2001). Second, influential theories of cognitive development have identified shape and color as core object properties which young children use for object individuation (Tremoulet, Leslie, & Hall, 2000; Wilcox, 1999) and object classification (Carey, 1985; Keil, 1989; Keil, Carter Smith, Simons, & Levin, 1998). Finally, shape and color are basic feature dimensions with consequences for many perceptual tasks such as visual search (e.g., Duncan & Humphreys, 1989).

To address the role of conceptual distinctiveness, we specifically targeted the subordinate category structure, that is, the number of kinds of a given object category, such as kinds of cars. Given the importance of basic-level categories from Experiment 1, we predicted that further subordinate category structure might provide additional support for high-fidelity visual object representations. That is, objects in memory will interfere with each other less if they are conceptually distinctive at the subcategory level. It is important to note that conceptual differences between two kinds have to, at some point, be realized in perceptual features or else the two images will be the same. However, for some object categories, there are many different kinds that are similar in color and shape (e.g., cell phones, TVs), and other categories may be highly distinctive in color and shape but be of all of the same kind (e.g., bottles, buckets). Thus, to the extent that shape, color, and kinds are uncorrelated in our 200-category set, we can probe for the independent contributions of perceptual and conceptual distinctiveness on alleviating interference in memory.

Experiment 2 used the following procedure. First, we obtained quantitative estimates of each object category's variation along three focused dimensions—color, three-dimensional shape, and subordinate category structure (i.e., the variety of kinds). Next, we calculated an interference slope for each category reflecting how memory performance in Experiment 1 was impacted by increasing the number of exemplars. Finally, we examined if there was any correlation between the distinctiveness measures and the memory interference scores.

Method

Stimuli. The database consisted of the 200 object categories with 17 exemplars from each category used in Experiment 1.

Participants. Eighteen adults (aged 20–35) gave informed consent and participated in one to three of the ranking experiments. They received $10/hr of participation. Twelve observers ranked each dimension. Seven participants completed one ranking experiment, four completed two ranking experiments, and seven completed all three ranking experiments. Because the aim was to obtain estimates for each category and each dimension, we allowed observers to rate more than one dimension if they had time, with the order of rating tasks randomized within and across observers. Further, as these ranks were gathered to predict memory interference in Experiment 1, no observers in Experiment 2 were the same as in Experiment 1.

Procedure. Participants were given instructions about the dimension along which they were going to rank the stimuli. For each trial, all 17 exemplars of a category were presented on a 30-in. computer monitor. Each object was displayed to fit within approximately $8° \times 8°$ of visual angle. Observers were instructed to assign a rating to characterize the set of items in the object category on a scale from 1 (*very similar*) to 5 (*very distinctive*) by clicking a numbered button on the screen. Participants completed 200 trials, one for each category, and were given as much time as they needed to respond. The order of trials was randomized across observers and ranking dimensions.

For color rankings, participants rated how similar or varied the colors were between all the exemplars of a category. Example categories at the low and high end of the scale were given for reference. Here a category with very similar colored exemplars was jack-o-lanterns, and one with very different colored exemplars was flags. For shape rankings, participants rated how similar or varied the three-dimensional shapes of the objects were, with example categories of golf balls for similar shapes and leaves for different shapes. For kind rankings, participants rated how few or many kinds of that object there were. Here, example categories were bowties and soda: There are few kinds of bowties but many kinds of soda pop. The example categories used in the instructions to the participants were chosen on the basis of pilot ranking completed by the four authors. Participants were instructed to focus solely on the dimension of interest and to ignore the others.

Data analysis. The reliability of the distinctiveness judgments was assessed using two different measures. First, the mean correlation (MC) reflects the average of all observer pairwise correlations. Second, the effective rater reliability (*R*) takes into account the MC across observers as well as the number of observers. *R* can be interpreted as the percent of the underlying true ranks that has been recovered for a given set of raters (Rosenthal & Rosnow, 1991). We also calculated the intraclass correlation (ICC) estimates on the rank data, which showed similar results as the MC estimates. ICC estimates are reported in Appendix A, Table A1.

Results

Data from two observers were excluded because their average correlation with other observers' data was greater than 2 *SD*s from the set of observers' average correlations. Postexperiment surveys suggest that these two observers might have used the 1–5 scale backward. Therefore, the analyses presented below were performed on 12 observers for the kind ranks and 11 observers for the shape and color ranks.

For all 200 categories, the mean rating was calculated for the color, shape, and kind tasks. These mean ratings were used as a measure of distinctiveness for each category along each dimension. For example, the set of 17 coffee mug images were highly distinctive in color (4.8), moderately distinctive in shape (2.9), and not very distinctive in the number of kinds represented (1.3). Figure 4 shows an example category with low and high distinctiveness for kinds, shapes, and colors. In our image database, there are few kinds of mugs but many kinds of cars; balls are very similar in three-dimensional shape, whereas different pieces of exercise equipment are very distinctive from each other in shape; keyboard keys are similar in color, while water guns are distinctive in color.

Across the 200 categories, the average pairwise correlation between observers was 0.64 for color ranks ($z = 9.00$, $p < .001$), 0.55 for shape ranks ($z = 7.69$, $p < .001$), and 0.25 for kind ranks ($z = 3.57$, $p < .001$). *R*, calculated based on the average pairwise correlation, accounting for the number of observers, was high for all three dimensions (shape: $R = .91$; color: $R = .95$; kind: $R = .80$). Comparing the dimensions, the conceptual distinctiveness (kinds) and the shape distinctiveness of that category were moderately correlated ($r = .307$, $p < .001$). Color and shape ratings showed a trend for a negative correlation ($r = -.136$, $p = .06$). Finally, color was not significantly correlated with kinds ($r = -.122$, $p = .08$).

Distinctiveness and interference. We next obtained a measure of interference for each category using the data from Experiment 1. That is, for each of the 200 categories, we calculated the change in memory performance for increasing numbers of studied exemplars (i.e., 1, 2, 4, 8, and 16). If a category has an interference slope of 0, this indicates that there was no impairment in memory as the number of exemplars to be remembered from that category increased. As the interference slope becomes more negative, this indicates worse memory performance with additional exemplars. The slope was calculated for each category as the change in percent correct as a function of $\log_2$(exemplars). Due to the across-

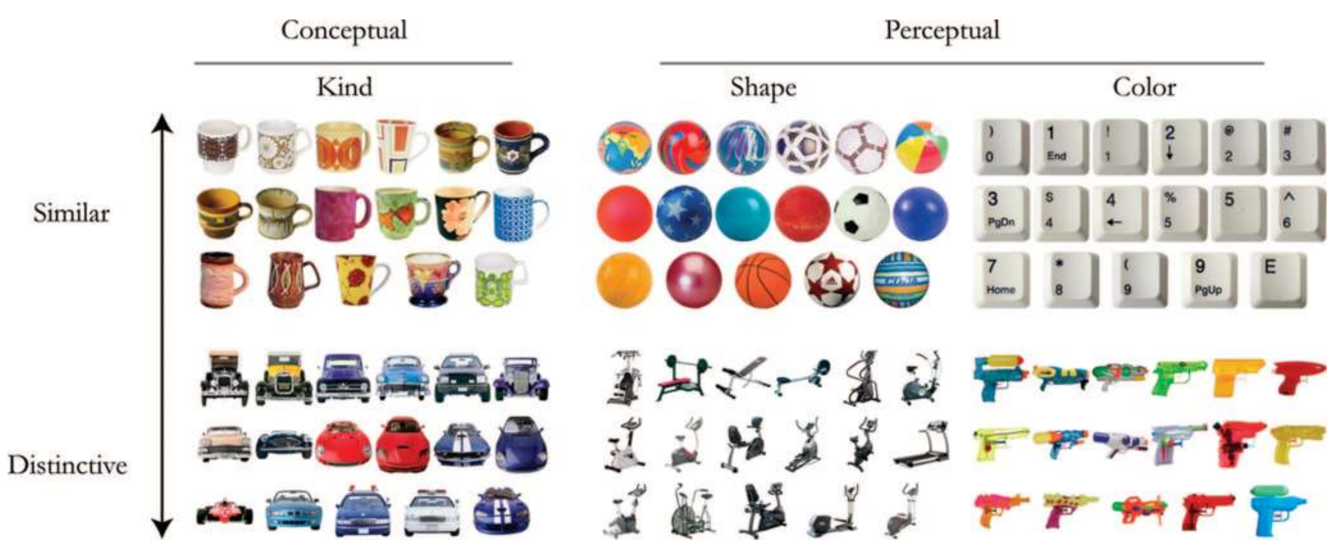

Figure 4. Distinctiveness rankings. Distinctiveness ratings were gathered for all 200 object categories along three dimensions: conceptual distinctiveness (are there few or many different kinds of this object?), shape distinctiveness (how similar or distinctive are the three-dimensional shapes of these objects?), and color distinctiveness (how similar or distinctive are these items in color?). Example categories at each end of the continuum (very similar and very distinctive) are shown for all three dimensions.

subjects design, the estimate of any single category's interference slope was noisy (see Appendix A for measures of reliability); however, the large number of object categories made it possible to examine the correlation between distinctiveness and the degree of interference in memory.

The correlations between interference in memory and the three distinctiveness ratings are shown in Figure 5. Categories composed of conceptually similar items showed more interference with increasing exemplars in memory; categories composed of conceptually distinct items were relatively spared from interference with increasing exemplars in memory. Overall, this correlation between conceptual distinctiveness of a category and the memory interference slope was significant ($r = .150$, $p = .03$). In contrast, there was no correlation between either of the perceptual distinctiveness measures and the interference slope (color: $r = -.020$, $p = .78$; shape: $r = .040$, $p = .58$). In other words, as more exemplars are loaded into memory, subsequent memory performance was not affected by whether these additional items were of similar or distinctive colors. Likewise, categories with items that were similar or distinctive in shape were equally likely to cause interference in memory with increasing exemplars.

For visualization purposes, the 200 categories were divided into three bins, based on the mean and standard deviation of the distinctiveness scores (<1 *SD*, within ±1 *SD*, and >1 *SD*). For statistical purposes, the correlations and calculations were performed without the binned ranks, using the average rank for each category and the interference slope. Although the correlations between conceptual ranks and memory interference were reliable, the magnitude was small. When we take into account the reliability of the measures of these factors, this correlation indicates that conceptual distinctiveness accounts for about 13% of the explainable variance in memory interference (see Appendix A, Table A1, r^2 adjusted column).

To assess the contributions of these three factors on memory interference, we also conducted a stepwise multiple linear regression analysis. This yielded a model with only the conceptual ranking included (partial correlation denoted with the abbreviation *pr*; conceptual included, $pr = .153$, $t[198] = 3.60$, $p = .03$; color excluded, $pr = .003$, $t < 1$, $p = .96$; shape excluded, $pr = -.010$, $t < 1$, $p = .89$).

Discussion

These data present a dissociation between the contribution of perceptual and conceptual distinctiveness in visual long-term memory. When categories have subordinate category structure—in other words, there are many kinds of objects—people can remember more of them, even if they all have similar shapes and colors (e.g., cell phones, televisions). Indeed, the most distinctive categories had an average interference slope of zero, suggesting that there was no cost to increasing the number of conceptually distinctive exemplars in memory. However, perceptual distinctiveness did not predict how well objects would be remembered as the number of exemplars increased. This finding is important given that we are examining memory for visual information, where exemplars are often different on perceptual dimensions (e.g., the shapes and arrangements of buttons on remote controls).

For terminological purposes, we refer to *distinctiveness in the number of kinds* as *conceptual distinctiveness* (rather than *kind-distinctiveness*). However note that the instructions were explicitly aimed at the number of kinds, and not about general knowledge or functions of the items per se. The conceptual distinctiveness of items depends on knowledge or experience with that category. For example, some raters might know much about cars or kinds of bread and give them high rankings, whereas other raters might perceive no obvious subordinate categorical structure to kinds of bread or cars and might give low rankings to these categories. For each category, we average the distinctiveness measures across raters and presume this reflects the *average subordinate category structure* for each category across the general population. In the memory interference measures, we also collapse over any individual differences with particular object categories when we calculate the interference score for each category. This, if anything, would make a correlation between conceptual distinctiveness and interference in memory more difficult to find. However, even in averaging people's expertise in the ranking and the memory measures, we still observe a significant correlation between conceptual distinctiveness and memory interference.

In this study, we did find a moderate correlation between shape ratings and conceptual ratings, consistent with a "shape" bias in category formation (Landau, Smith, & Jones, 1988). However, in

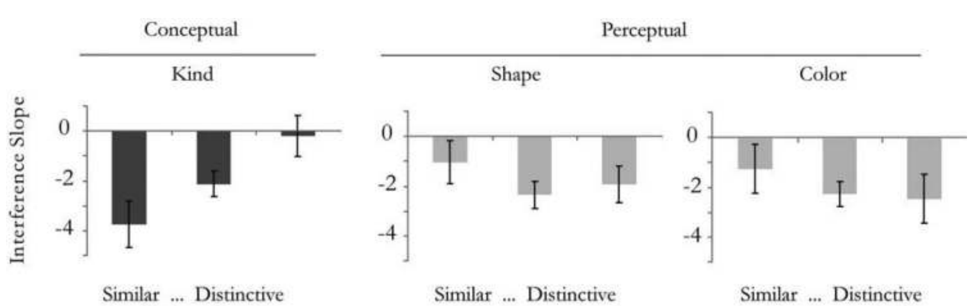

Figure 5. Memory interference and distinctiveness ratings. Object categories were divided into three bins based on the mean rank and standard deviation of the distinctiveness scores. The average interference slope from Experiment 1 is plotted against distinctiveness scores for conceptual, shape, and color ranking dimensions. Negative slope values indicate memory interference, and the steeper the slope the greater the interference. Conceptual distinctiveness correlated with the interference slope: The more conceptually distinct items from a category there were, the less interference there was as the number of exemplars in memory increased. Perceptual dimensions of color and shape did not correlate with the degree of interference in memory. Error bars reflect ±1 *SEM*.

general the three dimensions were not highly correlated, suggesting that conceptual variation for these stimuli is not directly proportional to the amount of perceptual variation along these dimensions. Importantly, it is variation along this conceptual dimension that predicts the degree to which items will interfere in memory, and not variation along perceptual dimensions.

One potential concern is that the perceptual rankings were along focused dimensions of shape and color, rather than capturing a more general perceptual distinctiveness over many perceptual features. This may account for why we did not observe a relationship between the perceptual distinctiveness measures and memory interference. To address this possibility, in Experiment 3 we obtained ratings on overall perceptual distinctiveness.

Experiment 3: Overall Perceptual Distinctiveness Rankings

In Experiment 3, observers made judgments about the overall perceptual distinctiveness of a set of exemplars. Observers were instructed to judge only overall visual appearance, including features such as size, color, shape, and texture, and so forth, and to ignore any knowledge they had of the depicted items, including their functions, the number of kinds, and so on. Additionally, a different set of observers completed the same task but with the images presented upside down. We expected this manipulation to help observers focus on the visual appearance and draw less on existing knowledge about what the objects were.

Method

Twenty-six adults (aged 20–35) gave informed consent and participated in the overall perceptual ranking experiment, with 13 observers presented with upright images and 13 observers presented with inverted images. One observer from each task also participated in Experiment 2, and no observers were the same as those in Experiment 1.

Observers were instructed to give the set of 17 objects a rating on a scale from 1 (*very similar in overall visual appearance*) to 5 (*very different in overall visual appearance*). Care was taken to remind observers to focus only on the visual or perceptual features such as size, color, shape, texture, and so forth and to ignore other aspects such as the number of kinds, the functions, or any other knowledge of the object category. For the inverted task, observers were told that, to help them focus on the visual appearance, the objects would be presented upside down. Example categories were jack-o-lanterns (very similar in overall visual appearance) and sippy cups (very different in overall visual appearance). All other methods were the same as in Experiment 2.

Results

Data from two observers were excluded because their average correlation with other observers' data was greater than 2 *SD*s from the set of observers' average correlations. Postexperiment surveys again suggest that these observers might have used the 1–5 scale backward. The analyses below used data from 12 observers for both upright and inverted versions of the overall perceptual distinctiveness task.

For all 200 categories, the mean overall perceptual distinctiveness rating was calculated for the upright and inverted groups. For example, birds were highly distinctive in overall perceptual judgments (4.0 upright, 4.1 inverted), whereas chessboards were not distinctive in overall perceptual judgments (1.8 upright, 2.0 inverted). Across the 200 categories, the average pairwise correlation between observers was .31 for upright images ($z = 4.42$, $p < .001$) and .24 for inverted images ($z = 3.4$, $p < .001$). Adjusting for the number of observers, both variants of the task showed a high effective rater reliability (upright: $R = .84$; inverted: $R = .79$). Further, upright and inverted overall perceptual distinctiveness judgments were significantly correlated with each other ($r = .71$, $p < .001$).

Next we examined the relationship between overall perceptual distinctiveness measures and the color, shape, and kind ranks from Experiment 2. When upright, overall perceptual judgments were highly correlated with shape ($r = .70$, $p < .001$) but were more correlated with the number of kinds than with color (kinds: $r = .35$, $p < .001$; color: $r = .22$, $p < .01$). When the same judgment was made by different observers on inverted images, overall perceptual distinctiveness was more evenly correlated with both shape and color (shape: $r = .51$, $p < .001$; color: $r = .44$, $p < .001$), and the correlation with number of kinds was numerically reduced, but still present ($r = .28$, $p < .001$).

Distinctiveness and interference. Next we examined the relationship between the overall perceptual distinctiveness ranks and memory interference. Neither of the overall perceptual distinctiveness ranks made on upright images or inverted images was significantly correlated with memory interference slopes (upright: $r = .10$, $p = .17$; inverted: $r = .01$, $p = .83$). Thus, consistent with the results of Experiment 2, categories composed of overall visually similar items showed no more interference in memory with increasing exemplars than did categories with overall visually distinctive items. The correlations between interference in memory and the two measures of overall perceptual distinctiveness are shown in Figure 6.

Although nonsignificant, there was a moderate correlation strength between overall perceptual distinctiveness on upright images and memory interference ($r = .10$, $p = .17$). A post hoc analysis suggests that this correlation is driven solely by the shared variance between upright overall perceptual distinctiveness and conceptual distinctiveness ratings of the number of kinds (see Appendix B).

Stepwise linear regression analysis using all of the ranking dimensions to predict memory interference also confirmed these findings, yielding a model with only conceptual ranking included, and all other predictors excluded (kinds included: $pr = .153$, $t[198] = 3.60$, $p = .03$; color excluded: $pr = .003$, $t < 1$, $p = .96$; shape excluded: $pr = -.01$, $t < 1$, $p = .89$; upright overall percept excluded: $pr = .047$, $t < 1$, $p = .51$; inverted overall percept excluded: $pr = -.028$, $t < 1$, $p = .69$).

Discussion

Measures of overall perceptual distinctiveness, whether on upright or inverted images, were only moderately correlated with the specific shape and color rankings, leaving open the possibility that these new perceptual ranking might correlate with memory interference slopes. However, neither upright nor inverted versions of

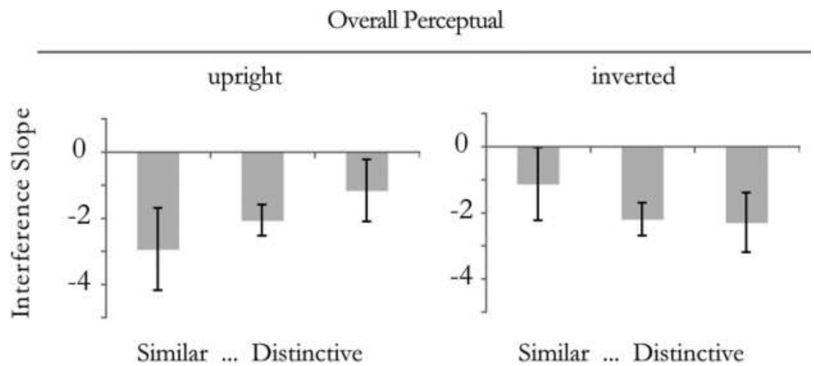

Figure 6. Memory interference and overall perceptual distinctiveness ratings. Object categories were divided into three bins based on the mean rank and standard deviation of the distinctiveness scores. The average interference slope from Experiment 1 is plotted against distinctiveness scores for overall perceptual ratings on images presented either upright or inverted. Error bars reflect ±1 *SEM*.

overall perceptual distinctiveness significantly correlated with memory interference.

We also observed that judgments of overall perceptual distinctiveness were not the same when images were upright versus inverted. One possible explanation for this difference is that when judging overall visual appearance, especially on upright objects, knowledge about different object categories might cause observers to adjust the amount of weight given to different feature dimensions. For example, given a category such as "apples" for which color variation is relevant for distinguishing which kind it is and shape variation is less important, observers might have intrinsically discounted shape variation and relied more on color variation when making the overall visual distinctiveness judgment. When inverted, observers may more easily discount what the objects are to focus on what they look like. Although it may not be possible to fully separate overall visual appearance from existing object knowledge, our attempts to do so with explicit instruction and inverted object images support the conclusion that perceptual distinctiveness does not account for the degree of within-category interference observed in Experiment 1.

In Experiments 2 and 3, we employed a correlational approach to explore the nature of the category interference effect observed in Experiment 1 (i.e., the finding that memory performance decreased as the number of within-category exemplars increased). This approach takes advantage of natural variation in object knowledge and the perceptual variability of the object categories tested. Future experiments can experimentally manipulate these factors by training observers to learn new categories (e.g., arbitrary categories for novel objects) which vary in their degree of perceptual distinctiveness. We predict that the degree of interference from additional exemplars within a category will depend on the degree of subcategorical knowledge, and not the degree of perceptual similarity within a category. However, these factors may be difficult to manipulate independently, as perceptual similarity judgments have been shown to depend on learned categories (Gauthier, James, Curby, & Tarr, 2003). Although future research should experimentally manipulate conceptual and perceptual distinctiveness, for the present purposes it is important to validate our perceptual distinctiveness measures, to make sure that the lack of correlation with memory interference is not due to limitations of the ranking procedures employed.

Experiment 4: Distinctiveness Rankings and Visual Search

The previous experiments support the hypothesis that conceptual distinctiveness, and not perceptual distinctiveness, alleviates memory interference. However, one concern with the lack of correlation between perceptual ranking on memory is that the perceptual rankings might not be valid measures of perceptual distinctiveness (e.g., because they are too noisy, the range is too restricted, or they simply do not provide a measure of perceptual distinctiveness that can predict performance on any other task). To address these concerns, we used a task that is known to depend on perceptual distinctiveness of items—visual search (e.g., Duncan & Humphreys, 1989). If the perceptual rankings correlate with search times, then we can conclude that the lack of correlation between perceptual distinctiveness and memory interference is not due to invalid measures of perceptual distinctiveness.

Method

Participants. Six adults (aged 20–35 years old) gave informed consent and received $40 for participating in the 4-hr search experiment. No observers in Experiment 4 participated in Experiment 1.

Stimuli. The image database was the same set used in Experiments 2 and 3, consisting of 200 object categories with 17 exemplars each. On each trial, all images were selected from one of the 200-object categories and presented in randomly selected locations from an invisible 7 × 5 grid (with each object subtending approximately 5° × 5° visual angle). At the center of the display one of the 17 exemplars was presented with a black outline frame, designating that item as the target item. Sixteen other items were placed randomly in the remainder of the display (see Figure 7). On target present trials, the target could appear at any location within the 7 × 5 grid.

Procedure. Observers performed a visual search task. On each trial, a random category from the set of 200 was selected, and a random exemplar from the set was chosen to be the target for that trial. The target item was presented at the center of the display highlighted with a black outline frame, and the remaining items were randomly placed on the rest of the display (see Figure 7). The

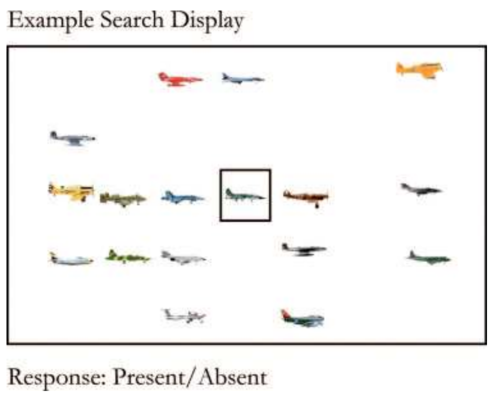

Figure 7. Visual search experiment methods. An example search display is shown. The target item was presented in a box at the center of the display. Observers responded as quickly as possible whether the target was present or absent in the surrounding display. Average search time was measured for each of the 200 object categories.

task was to determine as quickly and as accurately as possible whether the target item appeared anywhere else on the display. On half of the trials, the target was absent and all 16 nontarget exemplars were presented. On the other half of the trials, the target was present, and 15 different nontarget exemplars were shown (randomly chosen without replacement from the set of 16 nontargets). Observers completed four 1-hr sessions of 800 trials, yielding 16 trials per category.

Results

The overall error rate was low (<8%), and the following analyses were conducted on mean RT for correct responses collapsed across target presence (present vs. absent). RT was collapsed across observers and was calculated separately for each of the 200 categories, and it was then correlated with the conceptual and perceptual rankings obtained in Experiments 2 and 3.

There was no reliable correlation between conceptual distinctiveness and RT ($r = -.015$, $t < 1$, $p = .84$). However, both focused perceptual measures showed a significant correlation with RT: RT was faster when the within-category colors were more distinctive ($r = -.667$), $t(198) = 12.6$, $p < .001$, and when the within-category shapes were more distinctive ($r = -.146$), $t(198) = 2.08$, $p = .04$. Furthermore, both estimates of overall perceptual distinctiveness correlated with RT as well (upright: $r = -.520$, $p < .001$; inverted: $r = -.610$, $p < .001$). Taking into account the reliability of the ranks and search RTs, the amount of explainable variance accounted for by each dimension is, in order, inverted overall perceptual distinctiveness (64%), color (61%), upright overall perceptual distinctiveness (43%), and shape (4%; see Appendix A, Table A1, r^2 adjusted column). Figure 8 plots search RT as a function of the five distinctiveness rankings from Experiment 2 and 3, illustrating that perceptual, not conceptual, distinctiveness modulates search performance.

A stepwise multiple linear regression analysis using only the focused distinctiveness measures along color, shape, and conceptual dimensions yielded a model with both the color and shape included (partial correlation denoted with the abbreviation *pr*; color included: $pr = -.705$, $t[197] = 13.94$, $p < .001$; shape included: $pr = -.335$, $t[197] = 4.99$, $p < .001$; kinds excluded: $pr = -.036$, $t < 1$, $p = .62$). A stepwise linear regression analysis including all five distinctiveness measures includes both upright and inverted overall perceptual ranks and color and excludes shape and kinds (color included: $pr = -.601$, $t[196] = 10.52$, $p < .001$; upright overall percept included: $pr = -.301$, $t[196] = 4.41$, $p <$

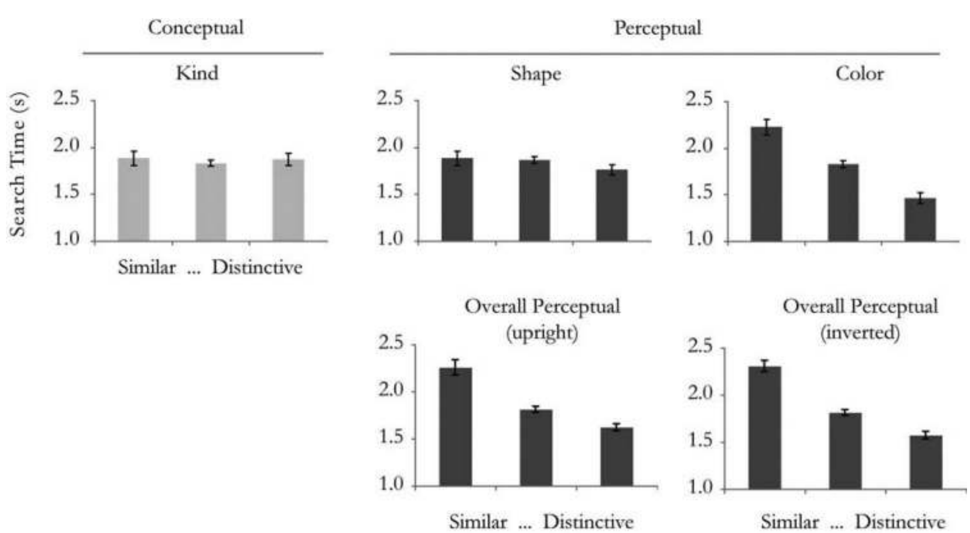

Figure 8. Search time and distinctiveness ratings. Object categories were divided into three bins based on the mean rank and standard deviation of the distinctiveness scores. The average search time from Experiment 4 is plotted for each bin, for conceptual, shape, and color ranking dimensions (Experiment 2), and overall perceptual rankings for upright and inverted images (Experiment 3). All perceptual measures correlated with search time: The more distinctive items were in color, shape, or overall appearance, the easier it was to find the search target amidst an array of exemplar distracters. Conceptual distinctiveness did not correlate with search time. Error bars reflect ±1 *SEM*.

.001; inverted overall percept included: $pr = .178$, $t[196] = 2.53$, $p = .01$; shape excluded: $pr = .13$; $t[196] = 1.86$, $p = .07$; kinds excluded: $pr = .12$; $t[196] = 1.7$, $p = .09$).

Finally, visual search RT does not predict memory interference ($r = .04$, $p = .54$). This provides an additional measure of perceptual distinctiveness that does not predict memory interference.

Discussion

Visual search rates for a target among exemplar foils was well predicted by how distinctive the set of exemplars within a category were in both color and shape, as well as in overall perceptual appearance. These results demonstrate that the within-category perceptual rankings are valid, and they confirm well-established findings that search is affected by similarity between targets and distracters and similarity between distracters and other distracters (Duncan & Humphreys, 1989). We did not observe a correlation between conceptual distinctiveness and search rates, even when taking into account the contributions of the other two factors. This is somewhat surprising, given that the visual search task has a slight memory demand: After fixating the target to find it among the field of distracters, a memory trace must be maintained so that inspected items can be compared to the target (see Shore & Klein, 2001, for a review of the role of memory in search). However, by design, we attempted to reduce the memory demands in the visual search task by having the target preview always present at the center of the display. Together, these correlations present a second dissociation between perceptual and conceptual contributions, here within the context of an online visual processing task.

We also found that the visual search RTs did not predict memory interference slopes from Experiment 1. Examining these two tasks of search and memory, there are some striking similarities: For search, observers are finding a target among distracters *in view*, whereas for recognition memory, observers "find" the same target from among the same "distracters" *in memory*. The lack of correlation between these two tasks implies that the factors that influence search times are not the same as the ones that influence visual long-term memory. Of course, in some sense this must be true; for example, in a search array visual acuity constrains the features that can be seen from items in the periphery and thus how attention is guided through the search array, whereas similar visual acuity constraints are not present in memory. However, one might expect at least a partial overlap between categories with easy targets to find *online* (search) and *offline* (memory), and we did not find any support for this prediction in the present data.

Finally, these data give rise to a double dissociation with the long-term memory experiment. Conceptual distinctiveness supports memory more than perceptual distinctiveness, whereas perceptual and not conceptual distinctiveness predicts search time. These data further confirm the results of Experiment 1, wherein memory for object details, seemingly "perceptual" information, is retrieved more easily and/or stored more efficiently when the item is conceptually distinct from other items in memory. Furthermore, the degree of perceptual overlap within a category does not predict what categories will be easier or harder to remember, but it does predict how difficult it will be to search for one among many. These results support, at a minimum, the conclusion that conceptual distinctiveness plays a more prominent role than perceptual distinctiveness in supporting detailed representations in visual long-term memory.

General Discussion

Observers are capable of remembering thousands of visual representations in long-term memory (e.g., Shepard, 1967; Standing, 1973), including detail beyond just the "gist" of the image (e.g., Brady et al., 2008; Hollingworth, 2004; Vogt & Magnussen, 2007). The aim of the current experiments was to explore what kind of information supports this detailed visual long-term memory. Our initial hypothesis was that knowledge about object category might provide a "conceptual hook" on which to index a single detailed representation but that multiple exemplars might exhaust the benefits of a conceptual hook quickly.

To examine this possibility, in Experiment 1 we presented observers with thousands of objects with a variable number of exemplars from each object category. We found that memory performance was quite high—even with 16 exemplars per category, observers were still well above chance at recognizing which of two exemplars they had previously seen (82%). This result indicates that 16 exemplars do not reach the bounds of exemplar-level memory capacity. However, we also found that memory performance was systematically lower as the number of stored exemplars increased, demonstrating that object categories play a role in maintaining such detailed memory representations.

To explore the nature of this categorical interference, our approach was to predict which object categories suffered from more interference or less interference in memory. In Experiments 2 and 3, for each category we estimated how distinctive the exemplars were in visual appearance (perceptual distinctiveness) and number of kinds (conceptual distinctiveness). We found that there was less interference for categories in which the exemplars were conceptually distinctive. On the other hand, the degree of perceptual distinctiveness did not predict the amount of interference: Perceptually distinctive exemplars lead to impaired memory performance just as much as perceptually similar exemplars. This was true for all four measures of perceptual distinctiveness we measured: shape, color, overall visual similarity, and inverted overall visual similarity. Experiment 4 verified that these perceptual measures are valid and predict performance on a visual search task, thus providing a double dissociation for the role of conceptual and perceptual distinctiveness in visual memory and visual search.

The present experiments expand our understanding of visual long-term memory in two ways. First, the fidelity of visual long-term memory representations appears to be higher than previously demonstrated, as observers succeeded at exemplar-level memory tasks after viewing thousands of objects with multiple exemplars per object category. Second, there was minor but reliable interference due to additional exemplars in memory, where the degree of memory interference can be predicted by how conceptually distinctive the exemplars are from each other, but not by their perceptual distinctiveness. The correlation between interference and conceptual distinctiveness ranks suggests that conceptual structure at the basic level and subordinate level supports the massive capacity of visual long-term memory.

Conceptual Knowledge Supports Detailed Visual Memory

Previous research has demonstrated that concepts can help support visual long-term memory representations. For example, Wiseman and Neisser (1974) presented observers with two-tone ambiguous face images (Mooney faces) and asked them to judge whether there was a face present. Although all of the images contained faces, observers remembered the images they had seen *as faces* better than images that did not make contact with this organizing concept. Similarly, memory for ambiguous shapes is improved when studied with an accompanying semantic label (Koutstaal et al., 2003, Experiment 1), and memory for real-world objects is better than memory for perceptually rich but nonmeaningful objects (Koutstaal et al., 2003, Experiment 2). Broadly, these data support the idea of a conceptual hook, wherein existing knowledge about object categories supports long-term storage. The current results expand upon this idea, suggesting that subordinate category structure can provide multiple hooks, which support long-term memory for item-specific details.

Although conceptual knowledge supports detailed long-term memory, it is still necessary to attend to the details in order to encode them into memory. For instance, focusing on the conceptual aspects of a stimulus can actually lead to impaired memory for visual details or other item-specific perceptual information (Intraub & Nicklos, 1985; G. R. Loftus & Kallman, 1979; Lupyan, 2008; Marks, 1991). For example, Marks (1991) found that observers had better memory for pictorial details when the study task required them to judge the physical features of the pictures (an item-specific task) compared to when they judged how well the pictures fit into scenes (a semantic task). This is consistent with the idea of transfer-appropriate processing (e.g., Morris, Bransford, & Franks, 1977; Roediger & Blaxton, 1987; Roediger, Weldon, & Challis, 1989), where memory performance depends on the match between how information was studied and how it is tested. These results suggest that attention to the details is necessary to remember them, and the current results add that item-specific details cannot be stored in memory without the support of preexisting conceptual knowledge.

Role of Concepts at Encoding and Retrieval

How might conceptual knowledge support detailed memory representations? One possibility is that prior knowledge of object categories can enable compressive encoding, reducing the information load of a presented stimulus. For example, compressive encoding has been demonstrated in working memory. In the verbal domain remembering familiar letter strings (e.g., FBI–CBS–NCAA–PBS) is much easier than unfamiliar strings (FB–ICB–SNA–AAP–BS; Bower & Springston, 1970). Similarly, in visual working memory, learned knowledge about color pairs enables more colors to be remembered (Brady, Konkle, & Alvarez, 2009; see also Feigenson & Halberda, 2008). Just as these working-memory representations rely on existing knowledge, long-term memory representations might rely on similar mechanisms. For example, using category-specific knowledge, it could take less space in memory to encode that an apple is round than that a bread loaf is round, because observers already know that apples are round and bread loaves can come in many shapes. On this account, compressive encoding allows observers to maintain more features from objects for which they have more prior knowledge. When presented with an item again at test, the richer representation stored in long-term memory would be more likely to match, improving recognition accuracy for objects for which we have more prior knowledge.

Alternatively, prior knowledge might support high memory performance by directing attention and encoding resources to only the details that are likely to distinguish between exemplars (e.g., Eysenck, 1979; Nosofsky, 1986; see also Goldstone, 1998). Indeed, the notion that diagnostic features support successful retrieval is fundamental to distinctiveness models of memory (Nairne, 2002; see also Hunt & Worthen, 2006; Schmidt, 1991), as well as models of categorization (Nosofsky, 1984, 1986; see also J. R. Anderson, 1991; Tenenbaum & Griffiths, 2001). In these models, the effectiveness of some cue or feature at retrieval depends not only on how likely it is to match an item stored in long-term memory but also on how unlikely it is to match other items. Thus, observers in our experiment may have used prior knowledge to encode only diagnostic features that would enable them to distinguish the item from exemplars within the category and from items across categories. At retrieval, these diagnostic features may enable observers to more accurately recognize which item was studied. Although the compressive encoding account predicts that observers are maintaining richer representations for some items and not others, this directed encoding account predicts that observers are maintaining a sparse representation of each item, where more informative and diagnostic details are stored for items for which we have more prior knowledge.

A final possibility is that prior knowledge creates new features for items people know more about, leading to an expanded representation. Studies of category learning provide support for this hypothesis, demonstrating that category learning can go beyond a strategic reweighting of relevant stimulus dimensions and instead involve the creation of new features to represent objects (Goldstone, Lippa, & Shiffrin, 2001; Hock, Webb, & Cavedo, 1987; Schyns & Murphy, 1994; Schyns & Rodet, 1997). For example, Schyns and Rodet (1997) showed that exposing observers to one kind of categorization task affects the features they use on a later categorization task, revealing that features are flexibly created and depend on an observer's past history with an object (i.e., their preexisting knowledge). Importantly, Schyns argued that new features for representing visual stimuli are created *as a consequence* of the categorization task and terms them *functional features* (e.g., Schyns & Murphy, 1994; see also Archambault, O'Donnell, & Schyns, 1999). On this account, observers in our experiments use preexisting knowledge about object categories to encode each item, where the richer the category and subordinate category information, the more functional features can be encoded. This enhanced coding model would allow observers to store more information from these stimuli. On this expanded encoding account, observers have a richer representation of some items, providing a higher probability of a match between a target item and this representation during retrieval. Both enhanced and compressive encoding (also called chunking or unitization) argue that observers have a richer representation for conceptually distinctive items but differ in that compression relies on hierarchically combining smaller parts into a larger whole, whereas newly created features might be large and not composed of smaller discrete parts (for discussion, see Goldstone, 1998; Schyns et al., 1998).

Although preexisting knowledge may allow for efficient encoding, it is also possible that conceptual distinctiveness is solely (or additionally) important at retrieval. In the strong version of this account, visual long-term memory representations are encoded with "equal" fidelity, and conceptual distinctiveness supports success at the memory task by providing an effective retrieval cue. This is a variant of the cue-overload hypothesis, which states that a memory retrieval cue becomes less effective as the number of studied items associated with that cue increases (M. J. Watkins, 1979; O. C. Watkins & Watkins, 1975) and that forgetting arises due to the ineffectiveness of retrieval cues, as opposed to memory trace degradation and retroactive interference (Wixted, 2004). To account for our results on this view, one would have to assume that memory is accessed on the basis of categorical retrieval cues, even when making exemplar-level comparisons. Consequently, increasing the number of remembered items sharing a single category cue will decrease the likelihood of accurate retrieval and recognition. When we have more knowledge about a category, subcategorical retrieval cues can be used, reducing the degree of cue overload.

Our results demonstrate that conceptual distinctiveness supports successful memory performance for visual memory tasks requiring exemplar-level detail. Future studies are required to understand whether this capacity is supported by compressive, directed, or enhanced encoding of the items, or if categorical interference occurs only at retrieval, and to what extent there is interference and degradation of stored memory traces due to conceptually similar items.

Content and Organization of Long-Term Memory

In the present experiments, we explored which features support the ability to maintain detailed long-term memory representations by looking for effects of similarity along different dimensions (conceptual and perceptual). Previous research has employed a similar approach to examine the content and organization of both short-term and long-term verbal memory representations. For instance, Baddeley (1966b) found that verbal short-term memory retention is much worse for a list of acoustically similar words than a control list, whereas there is a relatively small effect of semantic similarity on short-term memory for words. Long-term memory appears to show the opposite pattern: Lists of semantically similar words are remembered worse than a control list, whereas there is no effect of acoustic similarity on long-term memory for words (Baddeley, 1966a). On the basis of these patterns of acoustic and semantic interference, Baddeley argued that verbal long-term memory representations are largely conceptual in nature.

However, contrary to this hypothesis of purely conceptual encoding in verbal long-term memory, retrieval from long-term memory is primed not only by conceptually related items (E. F. Loftus, 1973) but also by phonetically similar items (E. F. Loftus, Senders, & Turkletaub, 1974). Moreover, it appears that words can be encoded acoustically in long-term memory when subjects are not required to perform a distracting short-term memory task just after learning the word list (Baddeley, 1966a). Combined, these results suggest that verbal long-term memory is organized in terms of both conceptual and phonetic information, favoring a multimodal view of memory in which each item is encoded and represented along multiple dimensions (E. F. Loftus et al., 1974). These findings from verbal long-term memory suggest our own results on visual long-term memory should be interpreted with caution.

Although we did not observe perceptual interference effects in the current study, observers were able to make exemplar-level discriminations which require memory for item-specific details. It is thus possible that visual long-term memory is organized primarily by categorical structure and that this conceptual representation provides a hook into an entire multimodal memory trace, enabling the storage and retrieval of both the conceptual content and item-specific perceptual details.

Relationship Between Visual and Verbal Long-Term Memory

It is widely found that stimuli presented in picture form lead to better memory performance than when those stimuli are presented in word form—the picture superiority effect (T. O. Nelson, Metzler, & Reed, 1974; Paivio, 1971; Paivio & Csapo, 1973; Weldon & Roediger, 1987). Put another way, memory capacity for pictures is larger than memory capacity for words (Shepard, 1967; Standing, 1973). A classic account for this is Paivio's dual-coding hypothesis (Paivio, 1971, 1986, 1991), in which pictures are likely to be implicitly named (e.g., Grill-Spector & Kanwisher, 2005) and thus benefit from both visual and verbal codes, whereas words are less likely to spontaneously be imagined pictorially and thus do not benefit from dual coding. However, pictures show better memory than words even when they are not being named (Madigan, 1983), and dual-coding predictions on memory performance when observers study pictures and are tested with words, and visa versa, are not borne out in experimental results (e.g., Mintzer & Snodgrass, 1999; Stenburg, Radeborg, & Hedman, 1995).

The sensory–semantic model (D. L. Nelson, Reed, & Walling, 1976) provided an alternate account for the picture superiority effect and suggested that pictures have stronger sensory codes than words do, leading to more distinct memory traces. Consistent with this idea, when semantically related pictures are studied, there is less interference in recognition memory than when corresponding semantically related words are studied (Dodson & Schacter, 2001; Israel & Schacter, 1997; Schacter, Israel, & Racine, 1999; Smith & Hunt, 1998; see also Shiffrin et al., 1995). On this account, pictures show less interference in recognition memory because they rely more on perceptual rather than conceptual features (cf. Hunt & McDaniel, 1993). However, this interpretation has largely been ruled out by subsequent studies (e.g., Schacter, Cendan, Dodson, & Clifford, 2001; Stenburg, 2006). For example, if the memory task is to identify whether an item from a previously studied category, requiring memory for conceptual information, memory is no worse when pictures were studied compared to when words were studied (Schacter et al., 2001). These data do not support the claim that pictures have less conceptual information than words (see also Potter & Faulconer, 1977; Stenburg, 2006).

One intuitive explanation for the difference in memory performance for pictorial and verbal stimuli is that when a picture is presented, much more preexisting knowledge can be brought to bear on the stimulus (e.g., not only that a car is pictured, but what kind of car it is, when it was likely made, etc.). This subordinate and associated knowledge is not present in a verbal stimulus that is a single word at the basic-level category (e.g., "car"). In this sense, pictures may simply give rise to more distinct conceptual traces than words, accounting for superior memory performance for pictures during standard recognition memory tasks.

Caveats

Given the fidelity of memory representations required to succeed at the current memory tasks, one might be tempted to believe that everyone has near photographic memory. However, we want to emphasize that this is certainly not the case for several reasons. First, we tested memory for subtle but meaningful differences (e.g., those along which exemplars differ from each other). Had we presented two objects that differed on an arbitrary dimension, say by a 1° rotation, it is likely that observers would have very poor memory performance. Second, the old–new repetition detection task provides a way to estimate how memory performance changes with elapsed time, given a continuous stream of input much like typical visual experience. Power law fits (see Wixted & Carpenter, 2007) suggest that while d' for the old–new task would still be above 1.0 after a day, it would likely fall below 1.0 after a month, and to below 0.6 after a year. Finally, memory for visual information is fundamentally a constructive process, which has been well documented in cases of eyewitness testimony (E. F. Loftus, 2003; see also Bower, Karlin, & Dueck, 1975; Carmichael, Hogan, & Walter, 1932). These studies demonstrate that the way a visual memory is queried influences the accuracy with which details are recalled (e.g., how fast were the cars going when they *smashed* vs. *bumped* into each other; E. F. Loftus & Palmer, 1974). Importantly, the idea that memory is reconstructed on the basis of experience and knowledge fits well with the central claim of this article that visual memory is supported by conceptual knowledge. A great challenge for visual memory research is to understand the relationship between knowledge of object categories and the visual features of underlying visual object representations (Palmeri & Tarr, 2008).

Conclusions

Visual long-term memory has a massive capacity to store information, both in the number of items that can be remembered and the amount of information that can be remembered about each item. Detailed memory representations seem to depend on conceptual knowledge, by which we mean that an item is more likely to interfere with another item if it has similar category or subordinate category information. Importantly, conceptual knowledge enables observers to maintain high-fidelity representations, and not simply gist-like representations that are fully abstracted away from perceptual detail. Here we show that existing knowledge about object categories and subordinate categorical structure supports the maintenance of detailed visual long-term memory. Visual memory capacities, often studied with controlled stimuli with which we have no prior experience, are part of an integrated conceptual system; as such visual memory capacity is intrinsically tied to what we know about what we are seeing.

References

Anderson, J. R. (1983, April 1). Retrieval of information from long-term memory. *Science, 220,* 25–30.

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

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17,* 1–12.

Archambault, A., O'Donnell, C., & Schyns, P. G. (1999). Blind to object changes: When *learning* the same object at different levels of categorization modifies its perception. *Psychological Science, 10,* 249–255.

Baddeley, A. D. (1966a). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18,* 302–309.

Baddeley, A. D. (1966b). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology, 18,* 362–365.

Bartels, A., & Zeki, S. (2000). The architecture of the colour centre in the human visual brain: New results and a review. *European Journal of Neuroscience, 12,* 172–193.

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports, 19,* 3–11.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition, 3,* 216–220.

Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology, 83,* 421–430.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual short-term memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138,* 487–502.

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

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of Neuroscience, 29,* 13992–14003.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carmichael, L. C., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology, 15,* 73–86.

Chun, M. M. (2003). Scene perception and memory. In D. Irwin & B. Ross (Eds.), *Psychology of learning and motivation: Advances in research and theory: Cognitive vision* (Vol. 42, pp. 79–108). San Diego, CA: Academic Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized color modules in macaque extrastriate cortex. *Neuron, 56,* 560–573.

Deese, J. (1959). On the prediction of occurrence of particular verbal instructions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychological Bulletin & Review, 8,* 155–161.

Duncan, G., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Eysenck, M. W. (1979). Depth, elaboration, and distinctiveness. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 89–118). Hillsdale, NJ: Erlbaum.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences, USA, 105,* 9926–9930.

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

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition, 78,* 27–43.

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

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron, 21,* 191–202.

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience, 1,* 235–241.

Hock, H. S., Webb, E., & Cavedo, L. C. (1987). Perceptual learning in visual category acquisition. *Memory & Cognition, 15,* 544–556.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 519–537.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421–445.

Hunt, R. R., & Worthen, J. (2006). *Distinctiveness and memory.* New York, NY: Oxford University Press.

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 284–298.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4,* 577–581.

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

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Keil, F. C., Carter Smith, W., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition, 65,* 103–135.

Kourtzi, Z., & Kanwisher, N. G. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience, 20,* 3310–3318.

Kourtzi, Z., & Kanwisher, N. (2001, April 12). Representation of perceived object shape by the human lateral occipital complex. *Science, 293,* 1506–1509.

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

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

Landau, B., Smith, L. B., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299–321.

Loftus, E. F. (1973). Activation of semantic memory. *American Journal of Psychology, 86,* 331–337.

Loftus, E. F. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews Neuroscience, 4,* 231–234.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.

Loftus, E. F., Senders, J. W., & Turkletaub, S. (1974). The retrieval of phonetically similar and dissimilar category members. *American Journal of Psychology, 87,* 57–63.

Loftus, G. R., & Kallman, H. J. (1979). Encoding and use of detail information in picture recognition. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 197–211.

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

Madigan, S. (1983). Picture memory. In J. C. Yuille (Ed.), *Imagery, memory, and cognition: Essays in honor of Allan Paivio* (pp. 65–89). Hillsdale, NJ: Erlbaum.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., . . . Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, USA, 92,* 8135–8139.

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

McClelland, J. L., & Chappel, M. (1998). Familiarity breeds differentiation: A subjective likelihood approach to the effects of experience in recognition memory. *Psychological Review, 10,* 724–760.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32,* 89–115.

Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *American Journal of Psychology, 112,* 113–146.

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

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10,* 389–395.

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory* (pp. 27–46). New York, NY: Oxford University Press.

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). The pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 523–538.

Nelson, T. O., Metzler, J., & Reed, D. L. (1974). Role of details in long-term recognition of pictures and verbal descriptions. *Journal of Experimental Psychology, 102,* 184–186.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.

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

Paivio, A. (1971). *Imagery and verbal processes.* New York, NY: Holt, Rinehart, & Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, England: Oxford University Press.

Paivio, A. (1991). *Images in mind: The evolution of a theory.* New York, NY: Harvester Wheatsheaf.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology, 5,* 176–206.

Palmeri, T. J., & Tarr, M. (2008). Visual object perception and long-term memory. In S. Luck & A. Hollingworth (Eds.), *Visual memory* (pp. 163–207). Oxford, England: Oxford University Press.

Potter, M. C., & Faulconer, B. S. (1977, February 6). Time to understand pictures and words. *Nature, 253,* 437–438.

Rawson, K. A., & Van Overschelde, J. P. (2008). How does knowledge promote memory? The distinctiveness theory of skilled memory. *Journal of Memory and Language, 58,* 646–668.

Roediger, H. L., & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference* (pp. 349–379). Hillsdale, NJ: Erlbaum.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories:

Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioural research: Method and data analysis* (2nd ed.). New York, NY: McGraw-Hill.

Schacter, D. L., Cendan, D. L., Dodson, C. S., & Clifford, E. (2001). Retrieval conditions and false recognition: Testing the distinctiveness heuristic. *Psychonomic Bulletin & Review, 8,* 827–833.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1–24.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19,* 523–542.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 565–578.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–17.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. E. Edin (Ed.), *The psychology of learning and motivation* (Vol. 31, pp. 301–349). San Diego, CA: Academic Press.

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

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163.

Shiffrin, R., Huber, D. E., & Marienelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Shore, D. L., & Klein, R. M. (2001). On the manifestations of memory in visual search. *Spatial Vision, 14,* 59–75.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261–267.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5,* 710–715.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25,* 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2,500 visual stimuli. *Psychonomic Science, 19,* 73–74.

Stenburg, G. (2006). Conceptual and perceptual factors in the pictures superiority effect. *European Journal of Cognitive Psychology, 18,* 813–847.

Stenburg, G., Radeborg, K., & Hedman, L. R. (1995). The picture superiority effect in a cross-modality recognition task. *Memory & Cognition, 23,* 425–441.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640.

Tremoulet, P. D., Leslie, A. M., & Hall, G. D. (2000). Infant individuation and identification of objects. *Cognitive Development, 15,* 499–522.

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology, 54,* 298–303.

von Restorff, H. (1933). Uber die wirkung von bereichsbildungen im spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung, 18,* 299–342.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4,* 274–290.

Watkins, M. J. (1979). Engrams as cuegrams and forgetting as cue overload: A cueing approach to the structure of memory. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 347–372). New York, NY: Academic Press.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 442–453.

Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 15,* 269–280.

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition, 72,* 125–166.

Wiseman, S., & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *American Journal of Psychology, 87,* 675–681.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55,* 235–269.

Wixted, J., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science, 18,* 133–134.

Wolfe, J. M. (1998). Visual memory: What do you know about what you saw? *Current Biology, 8,* R303–R304.

Appendix A

Test–Retest Reliability Estimates for Memory Interference Slopes

The interference slopes for each category were calculated using performance from several observers (e.g., no single observer saw a category with 1, 2, 4, 8, and 16 exemplars); thus neither the average pairwise observer correlation nor the effective rater reliability R could be calculated from these data. Instead, to obtain a reliability estimate for these measures, we simulated test–retest reliability estimates. In general, the test–retest reliability of a Measure X is var(X)/var(X_{hat}), that is, the true variance in the underlying signal divided by the observed variance of the measured signal.

To obtain the reliability measures, reported in Table A1, we estimated the reliability of the slope measure using a simulation procedure. Here, the general approach was to draw two samples from the estimated slope distributions for each category, correlate these, repeat for 1,000 iterations, and then compute the average of the 1,000 correlations to estimate test–retest reliability. The specific simulation procedure was as follows. For each category, a linear model was used to estimate the slope in memory performance as a function of the number of studied exemplars (X_{hat}), and the standard error of that slope estimate (X_{se}). Next, the variance of X_{hat} was rescaled to be equal to var(X_{hat}) − mean(X^2_{se}). This step is necessary because the original slope distribution, X_{hat}, already reflects noise from X_{se}. Next, we drew two samples from the rescaled slope distributions ($X_{\text{hat-rescaled}}$ and X_{se}), generating two possible sets of interference measures for all 200 categories. We then computed the correlation between these two vectors and repeated this sampling procedure for 1,000 iterations. Finally, we calculated the average correlation of the 1,000 simulated correlations, which reflects our estimate of the test–retest reliability, R_{sim} (see Table A1). The simulation code is available on our website.

This estimate of reliability was also calculated for the distinctiveness ranks, where X_{hat} was the average rating across observers and X_{se} was the standard error of the mean across observers. The simulated test–retest reliability (R_{sim}) and effective rater reliability (R) are similar in magnitude and are reported in Table A1.

Calculating Percent of Explainable Variance

After correlating the distinctiveness measures and interference slopes, we subsequently computed the percent of the maximum variance that can be accounted for (r^2_{adjusted}), given the reliability of the measures. Assuming a perfect correlation between two measures, the maximum correlation that can be observed between two measures is limited by the reliability of both the measures: $R_{\text{obs}} = R_{\text{true}} \sqrt{} (R_1 * R_2)$, where R_1 and R_2 are the reliability estimates of the two measures and R_{obs} is the highest possible correlation that can be observed given that the true correlation between R_1 and R_2 (R_{true}) is 1 (see Vul, Harris, Winkielman, & Pashler, 2009). All estimates of R_{adjusted} use the test–retest reliability estimates from the simulation; thus both the distinctiveness and memory interference slopes have a reliability measure calculated from the same procedure.

Table A1
Reliability Measures and Correlations

Variable	Reliability measure				Distinctiveness rank correlation: Interference slope			Distinctiveness rank correlation: Search RT		
	MC	*ICC*	*R*	R_{sim}	original r^2	max r^2_{obs}	r^2_{adjusted}	original r^2	max r^2_{obs}	r^2_{adjusted}
Kinds	.25	.18	.80	.74	.023*	.178	.127	.000	.577	.000
Shape	.55	.50	.93	.93	.002	.223	.007	.026*	.725	.035
Color	.64	.62	.95	.95	.000	.228	.002	.449*	.741	.606
Overall percept (upright)	.30	.23	.84	.81	.010	.194	.051	.270*	.632	.428
Overall percept (inverted)	.24	.17	.79	.75	.000	.180	.001	.372*	.585	.636
Interference slope				.24	—	—	—	—	—	—
Search RT	.75	.35	.95	.78	—	—	—	—	—	—

Note. Four measures of reliability are shown for each of the rating experiments as well as the memory interference slopes and the search reaction times (RT). *MC* is the mean pairwise correlation across subjects; *ICC* is the intraclass correlation, which is an alternate measure of rater reliability (Bartko, 1966); *R* is the effective rater reliability (Rosenthal & Rosnow, 1991), which takes into account *MC* and the number of subjects and can be thought of as the percent signal recovered from the true ranks; and R_{sim} is the average test–retest correlation simulated from sampling each measurement, adjusted for the standard error in the measurement. We correlated the five distinctiveness dimensions with interference slope and search RT, with the percentage of variance accounted for shown in the original r^2 column. The maximum possible r^2 value that could be observed, given the reliability of the two measures and assuming a perfect underlying correlation, is shown in the max r^2_{obs} column. The proportion of explainable variance accounted for, given this maximum r^2 value, is shown in the r^2_{adjusted} column.
* $p < .05$.

(*Appendices continue*)

Appendix B

Post Hoc Analysis of Overall Perceptual Rankings and Conceptual Ranks

The correlation between overall perceptual distinctiveness measures (made on upright images) and memory interference slopes showed a moderate correlation strength, though it was not statistically significant ($r = .10$, $p < .17$). We also observed that upright perceptual ratings were significantly correlated with conceptual ratings ($r = .35$, $p < .001$). In this post hoc analysis, we examined the possibility that the correlation between memory interference and upright perceptual ratings was driven by shared variance with the conceptual ratings.

To do so, we divided the set of 200 categories into two subsets. The first subset contained the categories that had similar ranks for upright perceptual distinctiveness and conceptual distinctiveness (kinds). Of the 200 categories, 47 categories had average ranks within ± 0.5 for these two dimensions. Examining only these categories, there was a significant correlation between memory interference and these 47 categories' conceptual ranks ($r = .34$, $p = .02$) and upright overall perceptual ranks ($r = .31$, $p = .04$).

The second subset contained the remaining 153 categories, which had different ranks for the overall perceptual and conceptual dimensions. Importantly, the conceptual ranks of these 153 categories still showed a correlation with memory slope ($r = .16$, $p = .05$), whereas the overall perceptual ranks did not ($r = .01$, $p = .88$). This analysis suggests that the moderate correlation observed between upright overall perceptual ranks and memory interference can be attributed to its correlation with the conceptual dimension of the number of kinds.

Received March 25, 2009
Revision received January 8, 2010
Accepted January 11, 2010 ■