

Visual long-term memory has a massive storage capacity for object details

Timothy F. Brady*, Talia Konkle, George A. Alvarez, and Aude Oliva*

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved August 1, 2008 (received for review April 8, 2008)

One of the major lessons of memory research has been that human memory is fallible, imprecise, and subject to interference. Thus, although observers can remember thousands of images, it is widely assumed that these memories lack detail. Contrary to this assumption, here we show that long-term memory is capable of storing a massive number of objects with details from the image. Participants viewed pictures of 2,500 objects over the course of 5.5 h. Afterward, they were shown pairs of images and indicated which of the two they had seen. The previously viewed item could be paired with either an object from a novel category, an object of the same basic-level category, or the same object in a different state or pose. Performance in each of these conditions was remarkably high (92%, 88%, and 87%, respectively), suggesting that participants successfully maintained detailed representations of thousands of images. These results have implications for cognitive models, in which capacity limitations impose a primary computational constraint (e.g., models of object recognition), and pose a challenge to neural models of memory storage and retrieval, which must be able to account for such a large and detailed storage capacity.

object recognition | gist | fidelity

We have all had the experience of watching a movie trailer and having the overwhelming feeling that we can see much more than we could possibly report later. This subjective experience is consistent with research on human memory, which suggests that as information passes from sensory memory to short-term memory and to long-term memory, the amount of perceptual detail stored decreases. For example, within a few hundred milliseconds of perceiving an image, sensory memory confers a truly photographic experience, enabling you to report any of the image details (1). Seconds later, short-term memory enables you to report only sparse details from the image (2). Days later, you might be able to report only the gist of what you had seen (3).

Whereas long-term memory is generally believed to lack detail, it is well established that long-term memory can store a massive number of items. Landmark studies in the 1970s demonstrated that after viewing 10,000 scenes for a few seconds each, people could determine which of two images had been seen with 83% accuracy (4). This level of performance indicates the existence of a large storage capacity for images.

However, remembering the gist of an image (e.g., “I saw a picture of a wedding not a beach”) requires the storage of much less information than remembering the gist and specific details (e.g., “I saw that specific wedding picture”). Thus, to truly estimate the information capacity of long-term memory, it is necessary to determine both the quantity of items that can be remembered and the fidelity (amount of detail), with which each item is remembered. This point highlights an important limitation of large-scale memory studies (4–6): the level of detail required to succeed at the memory tests was not systematically examined. In these studies, the stimuli were images taken from magazines, where the foil items used in the two-alternative forced-choice tests were random images drawn from the same set (4). Thus, the foil items were typically quite different from the

studied images, making it impossible to conclude whether the memories for each item in these previous experiments consisted of only the “gist” or category of the image, or whether they contained specific details about the images. Therefore, it remains unclear exactly how much visual information can be stored in human long-term memory.

There are reasons for thinking that the memories for each item in these large-scale experiments might have consisted of only the gist or category of the image. For example, a well known body of research has shown that human observers often fail to notice significant changes in visual scenes; for instance, if their conversation partner is switched to another person, or if large background objects suddenly disappear (7, 8). These “change blindness” studies suggest that the amount of information we remember about each item may be quite low (8). In addition, it has been elegantly demonstrated that the details of visual memories can easily be interfered with by experimenter suggestion, a matter of serious concern for eyewitness testimony, as well as another indication that visual memories might be very sparse (9). Taken together, these results have led many to infer that the representations used to remember the thousands of images from the experiments of Shepard (5) and Standing (4) were in fact quite sparse, with few or no details about the images except for their basic-level categories (8, 10–12).

However, recent work has also suggested that visual long-term memory representations can be more detailed than previously believed. Long-term memory for objects in scenes can contain more information than only the gist of the object (13–16). For instance, Hollingworth (13) showed that, when requiring memory for a hundred or more objects, observers remain significantly above chance at remembering which exemplar of an object they have seen (e.g., “did you see this power drill or that one?”) even after seeing up to 400 objects in between studying the object and being tested on it. This result suggests that memory is capable of storing fairly detailed visual representations of objects over long time periods (e.g., longer than working memory).

The current study was designed to estimate the information capacity of visual long-term memory by simultaneously pushing the system in terms of both the quantity and the fidelity of the representations that must be stored. First, we used isolated objects that were not embedded in scenes, to more systematically control the conceptual content of the stimulus set and prevent the role of contextual cues that may have contributed to memory performance in previous experiments. In addition, we used very subtle visual discriminations to probe the fidelity of the visual representations. Last, we had people remember several thou-

Author contributions: T.F.B., T.K., G.A.A., and A.O. designed research, performed research, analyzed data, and wrote the article.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

*To whom correspondence may be addressed. E-mail: tbrady@mit.edu or oliva@mit.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803390105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

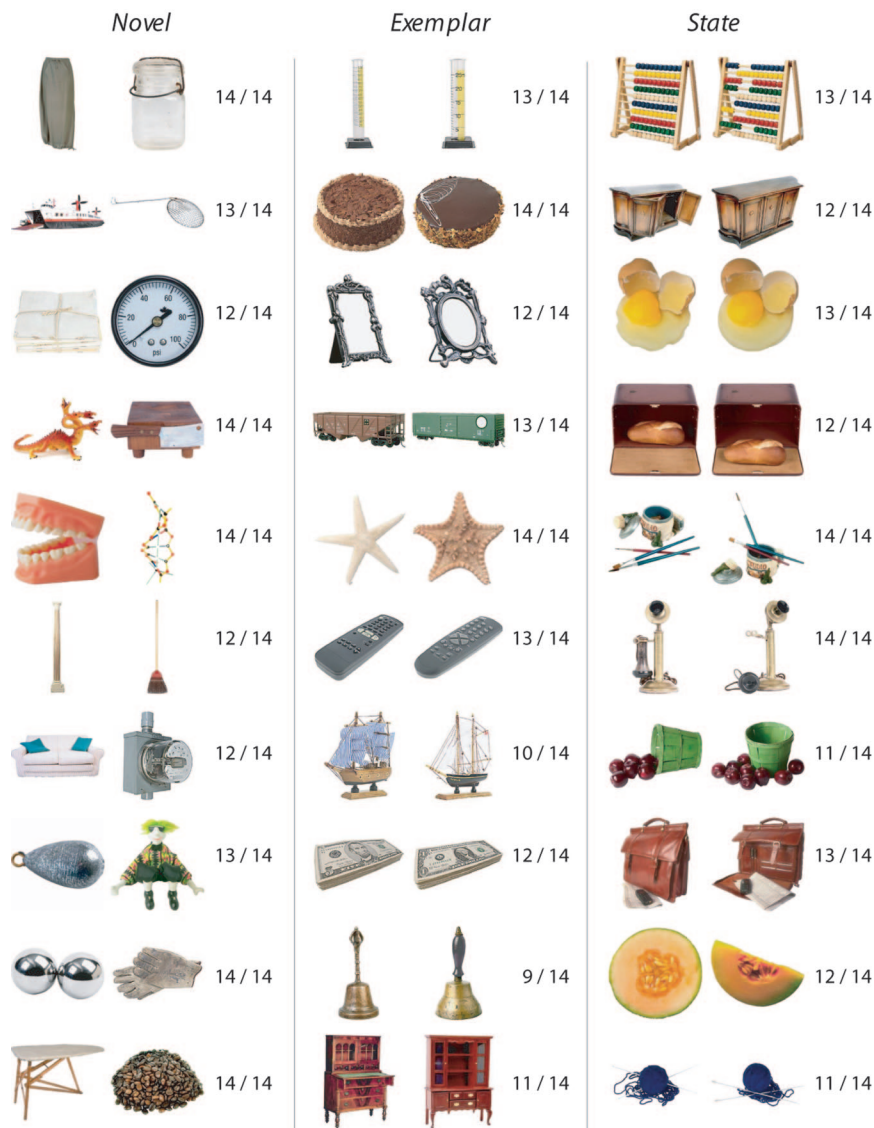


Fig. 1. Example test pairs presented during the two-alternative forced-choice task for all three conditions (novel, exemplar, and state). The number of observers reporting the correct item is shown for each of the depicted pairs. The experimental stimuli are available from the authors.

sand objects. Combined, these manipulations enable us to estimate a new bound on the capacity of memory to store visual information.

Results

Observers were presented with pictures of 2,500 real world objects for 3 s each. The experiment instructions and displays were designed to optimize the encoding of object information into memory. First, observers were informed that they should try to remember all of the details of the items (17). Second, objects from mostly distinct basic-level categories were chosen to minimize conceptual interference (18). Last, memory was tested with a two-alternative forced-choice test, in which a studied item was paired with a foil and the task was to choose the studied item, allowing for recognition memory rather than recall memory (as in 4).

We varied the similarity of the studied item and the foil item in three ways (Fig. 1). In the novel condition, the old item was paired with a new item that was categorically distinct from all of the previously studied objects. In this case, remembering the category of the object, even without remembering the visual

details of the object, would be sufficient to choose the appropriate item. In the exemplar condition, the old item was paired with a physically similar new object from the same basic-level category. In this condition, remembering only the basic-level category of the object would result in chance performance. Last, in the state condition, the old item was paired with a new item that was exactly the same object, but appeared in a different state or pose. In this condition, memory for the category of the object, or even for the identity of the object, would be insufficient to select the old item from the pair. Thus, memory for specific details from the image is required to select the appropriate object in both the exemplar and state conditions. Critically, observers did not know during the study session which items of the 2,500 would be tested afterward, nor what they would be tested against. Thus, any strategic encoding of a specific detail that would distinguish between the item and the foil was not possible. To perform well on average in both the exemplar and the state conditions, observers would have to encode many specific details from each object.

Performance was remarkably high in all three of the test conditions in the two-alternative forced choice (Fig. 2). As

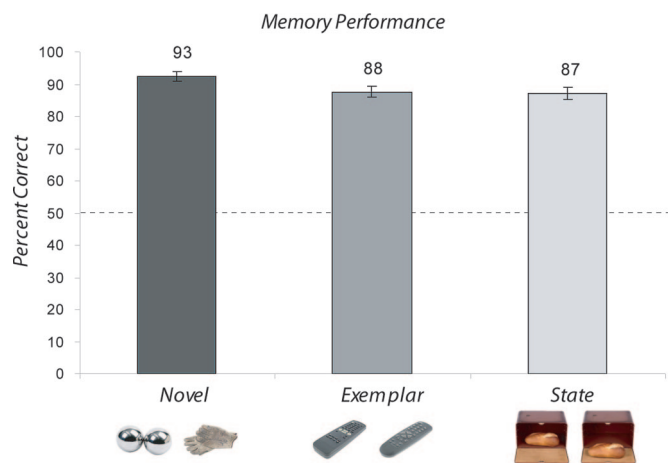


Fig. 2. Memory performance for each of the three test conditions (novel, exemplar, and state) is shown above. Error bars represent SEM. The dashed line indicates chance performance.

anticipated based on previous research (4), performance was high in the novel condition, with participants correctly reporting the old item on 92.5% (SEM, 1.6%) of the trials. Surprisingly, performance was also exceptionally high in both conditions that required memory for details from the images: on average, participants responded correctly on 87.6% (SEM, 1.8%) of exemplar trials, and 87.2% (SEM, 1.8%) of the state trials. A one-way repeated measures ANOVA revealed a significant effect of condition, $F(2,26) = 11.3, P < 0.001$. Planned pairwise t tests show that performance in the novel condition was significantly more accurate than the state and exemplar conditions [novel vs. exemplar: $t(13) = 3.4, P < 0.01$; novel vs. state: $t(13) = 4.3, P < 0.01$; and exemplar vs. state, n.s. $P > 0.10$]. However, reaction time data were slowest in the state condition, intermediate in the exemplar condition, and fastest in the novel conditions [M = 2.58 s, 2.42 s, 2.28 s, respectively; novel vs. exemplar: $t(13) = 1.81, P = 0.09$; novel vs. state: $t(13) = 4.05, P = 0.001$; and exemplar vs. state: $t(13) = 2.71, P = 0.02$], consistent with the idea that the novel, exemplar, and state conditions required increasing detail. Participant reports afterward indicated that they were usually explicitly aware of which item they had seen, as they expressed confidence in their performance and volunteered information about the details that enabled them to pick the correct items.

During the presentation of the 2,500 objects, participants monitored for any repeat images. Unbeknownst to the participants, these repeats occurred anywhere from 1 to 1,024 images previously in the sequence (on average, one in eight images was a repeat). This task insured that participants were actively attending to the stream of images as they were presented and provided an online estimate of memory storage capacity over the course of the entire study session.

Performance on the repeat-detection task also demonstrated remarkable memory. Participants rarely false-alarmed (1.3%; SEM, $\pm 1\%$), and were highly accurate in reporting actual repeats (96% overall; SEM, $\pm 1\%$). Accuracy was near ceiling for repeat images with up to 63 intervening items, and declined gradually for detecting repeat items with more intervening items (supporting information (SI) Text and Fig. S1). Even at the longest condition of 1,023 intervening items (i.e., items that were initially presented ≈ 2 h previously), the repeats were detected $\approx 80\%$ of the time. The correlation between the performance of the observers at the repeat-detection task and their performance in the forced choice was high ($r = 0.81$). The error rate as a function of number of intervening items fits well with a standard

power law of forgetting ($r^2 = 0.98$). The repeat-detection task also shows that this high capacity memory arises not only in two-alternative forced-choice tasks, but also in ongoing old/new recognition tasks, although the repeat-detection task did not probe for more detailed representations beyond the category level. Together with the memory test, these results indicate a massive capacity-memory system, in terms of both the quantity and fidelity of the visual information that can be remembered.

Discussion

We found that observers could successfully remember details about thousands of images after only a single viewing. What do these data say about the information capacity of visual long-term memory? It is known from previous research that people can remember large numbers of pictures (4–6) but it has often been assumed that they were storing only the gist of these images (8, 10–12). Whereas some evidence suggested observers are capable of remembering details about a few hundred objects over long time periods (13), to our knowledge, no experiment had previously demonstrated accurate memory at the exemplar or state level on such a large scale. The present results demonstrate visual memory is a massive store that is not exhausted by a set of 2,500 detailed representations of objects. Importantly, these data cannot reveal the format of these representations, and should not be taken to suggest that observers have a photographic-like memory (19). Further work is required to understand how the details from the images are encoded and stored in visual long-term memory.

The Information Capacity of Memory. Memory capacity cannot solely be characterized by the number of items stored: a proper capacity estimate takes into account the number of items remembered and multiplies this by the amount of information per item. In the present experiment we show that the information remembered per item is much higher than previously believed, as observers can correctly choose among visually similar foils. Therefore, any estimate of long-term memory capacity will be significantly increased by the present data. Ideally, we could quantify this increase, for example, by using information-theoretic bits, in terms of the actual visual code used to represent the objects. Unfortunately, one must know how the brain encodes visual information into memory to truly quantify capacity in this way.

However, Landauer (20) provided an alternate method for quantifying the capacity of memory by calculating the number of bits required to correctly make a decision about which items have been seen and which have not (21). Rather than assign images codes based on visual similarity, this model assigns each image a random code regardless of its visual appearance. Memory errors happen when two images are assigned the same code. In this model the optimal code length is computed from the total number of items to remember and the percentage correct achieved on a two-alternative forced-choice task (see SI Text). Importantly, this model does not take into account the content of the remembered items: the same code length would be obtained if people remembered 80% of 100 natural scenes or 80% of 100 colored letters. In other words, the bits in the model refer to content-independent memory addresses, and not estimated codes used by the visual system.

Given the 93% performance in the novel condition, the optimal code would require 13.8 bits per item, which is comparable with estimates of 10–14 bits needed for previous large-scale experiments (20). To expand on the model of Landauer, we assume a hierarchical model of memory where we first specify the category and the additional bits of information specify the exemplar and state of the item in that category (see SI Text). To match 88% performance in the exemplar conditions, 2.0 additional bits per item are required for each item. Similarly, 2.0

additional bits are required to achieve 87% correct in the state condition. Thus, we increase the estimated code length from 13.8 to 17.8 bits per item. This raises the lower bound on our estimate of the representational capacity of long-term memory by an order of magnitude, from $\approx 14,000$ ($2^{13.8}$) to $\approx 228,000$ ($2^{17.8}$) unique codes. This number does not tell us the true visual information capacity of the system. However, this model is a formal way of demonstrating how quickly any estimate of memory capacity grows if we increase the size of the representation of each individual object in memory.

Why examine the capacity of people to remember visual information? One reason is that the evolution of more complicated cognition and behavioral repertoires involved the gradual enlargement of the long-term memory capacities of the brain (22). In particular, there are reasons to believe the capacity of our memory systems to store perceptual information may be a critical factor in abstract reasoning (23, 24). It has been argued, for example, that abstract conceptual knowledge that appears amodal and abstracted from actual experience (25) may in fact be grounded in perceptual knowledge (e.g., perceptual symbol systems; see ref. 26). Under this view, abstract conceptual properties are created on the fly by mental simulations on perceptual knowledge. This view suggests an adaptive significance for the ability to encode a large amount of information in memory: storing large amounts of perceptual information allows abstraction based on all available information, rather than requiring a decision about what information might be necessary at some later point in time (24, 27).

Organization of Memory. All 2,500 items in our study stream were categorically distinct and thus had different high level, conceptual representations. Long-term memory is often seen as organized by conceptual similarity (e.g., in spreading activation models; see refs. 28 and 29). Thus, the conceptual distinctiveness of the objects may have reduced interference between them and helped support the remarkable memory performance we observed (18). In addition, recent work has suggested that the representation of the perceptual features of an object may often differ depending on the category the object is drawn from (30). Taken together, these ideas suggest an important role for categories and concepts in the storage of the visual details of objects, an important area of future research.

Another possible distinction in the organization of memory is between memory for objects, memory for collections of objects, and memory for scenes. Whereas some work has shown that it is possible to remember details from scenes drawn from the same category (15), future work is required to examine massive and detailed memory for complex scenes.

Familiarity versus Recollection. The literature on long-term memory frequently distinguishes between two types of recognition memory: familiarity, the sense that you have seen something before; and recollection, specific knowledge of where you have seen it (31). However, there remains controversy over the extent to which these types of memory can be dissociated, and the extent to which forced-choice judgments tap into familiarity more than recollection or vice versa (32). In addition, whereas some have argued that perceptual information is often more associated with familiarity and conceptual information is associated more with recollection (31), this view also remains disputed (32, 33). Thus, it is unclear the relative extent to which choices of the observers in the current two-alternative forced-choice tests were based on familiarity versus recollection. Given the perceptual nature of the details required to select the studied item, it is likely that familiarity has a major role, and that recollection aids recognition performance on the subset of trials in which observers were explicitly aware of the details that were most helpful to their decision (a potentially large subset of trials,

based on self-reports). Importantly, however, whether observers were depending on recollection or familiarity alone, the stored representation still requires enough detail to distinguish it from the foil at test. Our main conclusion is the same whether the memory is subserved by familiarity or recollection: observers encoded and retained many specific details about each object.

Constraints on Models of Object Recognition and Categorization.

Long-term memory capacity imposes a constraint on high-level cognitive functions and on neural models of such functions. For example, approaches to object recognition often vary by either relying on brute force online processing or a massive parallel memory (34, 35). The present data lend credence to object-recognition approaches that require massive storage of multiple object viewpoints and exemplars (36–39). Similarly, in the domain of categorization, a popular class of models, so-called exemplar models (40), have suggested that human categorization can best be modeled by positing storage of each exemplar that is viewed in a category. The present results demonstrate the feasibility of models requiring such large memory capacities.

In the domain of neural models, the present results imply that visual processing stages in the brain do not, by necessity, discard visual details. Current models of visual perception posit a hierarchy of processing stages that reach more and more abstract representations in higher-level cortical areas (35, 41). Thus, to maintain featural details, long-term memory representations of objects might be stored throughout the entire hierarchy of the visual processing stream, including early visual areas, possibly retrieved on demand by means of a feedback process (41, 42). Indeed, imagery processes, a form of representation retrieval, have been shown to activate both high-level visual cortical areas and primary visual cortex (43). In addition, functional MRI studies have indicated that a relatively mid-level visual area, the right fusiform gyrus, responds more when observers are encoding objects for which they will later remember the specific exemplar, compared with objects for which they will later remember only the gist (44). Understanding the neural substrates underlying this massive and detailed storage of visual information is an important goal for future research and will inform the study of visual object recognition and categorization.

Conclusion

The information capacity of human memory has an important role in cognitive and neural models of memory, recognition, and categorization, because models of these processes implicitly or explicitly make claims about the level of detail stored in memory. Detailed representations allow more computational flexibility because they enable processing at task-relevant levels of abstraction (24, 27), but these computational advantages tradeoff with the costs of additional storage. Therefore, establishing the bounds on the information capacity of human memory is critical to understanding the computational constraints on visual and cognitive tasks.

The upper bound on the size of visual long-term memory has not been reached, even with previous attempts to push the quantity of items (4), or the attempt of the present study to push both the quantity and fidelity. Here, we raise only the lower bound of what is possible, by showing that visual long-term memory representations can contain not only gist information but also details sufficient to discriminate between exemplars and states. We think that examining the fidelity of memory representations is an important addition to existing frameworks of visual long-term memory capacity. Whereas in everyday life we may often fail to encode the details of objects or scenes (7, 8, 17), our results suggest that under conditions where we attempt to encode such details, we are capable of succeeding.

Materials and Methods

Participants. Fourteen adults (aged 20–35) gave informed consent and participated in the experiment. All of the participants were tested simultaneously, by using computer workstations that were closely matched for monitor size and viewing distance.

Stimuli. Stimuli were gathered by using both a commercially available database (Hemera Photo-Objects, Vol. I and II) and internet searches by using Google Image Search. Overall, 2,600 categorically distinct images were gathered for the main database, plus 200 paired exemplar images and 200 paired state images drawn from categories not represented in the main database. The experimental stimuli are available from the authors. Once these images had been gathered, 200 were selected at random from the 2,600 objects to serve in the novel test condition. Thus, all participants were tested with the same 300 pairs of novel, exemplar, and state images. However, the item seen during the study session and the item used as the foil at test were randomized across participants.

Study Blocks. The experiment was broken up into 10 study blocks of ≈ 20 min each, followed by a 30 min of testing session. Between blocks participants were given a 5-min break, and were not allowed to discuss any of the images they had seen. During a block, ≈ 300 images were shown, with 2,896 images shown overall: 2,500 new and 396 repeated images. Each image (subtending 7.5 by 7.5° of visual angle) was presented for 3 s, followed by an 800-ms fixation cross.

Repeat-Detection Task. To maintain attention and to probe online memory capacity, participants performed a repeat-detection task during the 10 study blocks. Repeated images were inserted into the stream such that there were between 0 and 1,023 intervening items, and participants were told to respond by using the spacebar anytime that an image repeated throughout the entire study period. They were not informed of the structure of the repeat conditions. Participants were given feedback only when they responded, with the fixation cross turning red if they had incorrectly pressed the space bar (false alarm) or green if they had correctly detected a repeat (hit), and were given no feedback for misses or correct rejections.

Overall, 56 images were repeated immediately (1-back), 52 were repeated with 1 intervening item (2-back), 48 were repeated with 3 intervening items (4-back), 44 were repeated with 7 intervening items (8-back), and so forth, down to 16 repeated with 1,023 intervening items (1,024-back). Repeat items were inserted into the stream uniformly, with the constraint that all of the lengths of n-backs (1-back, 2-back, 4-back, and 1,024-back) had to occur equally in the first half of the experiment and the second half. This design ensured that fatigue would not differentially affect images that were repeated from further back in the stream. Due to the complexity of generating a properly counterbalanced set of repeats, all participants had repeated images appear at the same places within the stream. However, each participant saw a different order of the 2,500 objects, and the specific images repeated in the n-back conditions were also different across participants. Images that would later be tested in one of the three memory conditions were never repeated during the study period.

Forced-Choice Tests. Following a 10-min break after the study period, we probed the fidelity with which objects were remembered. Two items were presented on the screen, one previously seen old item, and one new foil item. Observers reported which item they had seen before in a two-alternative forced-choice task.

Participants were allowed to proceed at their own pace and were told to emphasize accuracy, not speed, in making their judgments. The 300 test trials were presented in a random order for each participant, with the three types of test trials (novel, exemplar, and state) interleaved. The images that would later be tested were distributed uniformly throughout the study period.

ACKNOWLEDGMENTS. We thank P. Cavanagh, M. Chun, M. Greene, A. Hollingworth, G. Kreiman, K. Nakayama, T. Poggio, M. Potter, R. Rensink, A. Schachner, T. Thompson, A. Torralba, and J. Wolfe for helpful conversation and comments on the manuscript. This work was partly funded by National Institutes of Health Training Grant T32-MH020007 (to T.F.B.), a National Defense Science and Engineering Graduate Fellowship (T.K.), National Research Service Award Fellowship F32-EY016982 (to G.A.A.), and National Science Foundation (NSF) Career Award IIS-0546262 and NSF Grant IIS-0705677 (to A.O.).

- Sperling G (1960) The information available in brief visual presentations. *Psychol Monogr* 74:1–29.
- Phillips WA (1974) On the distinction between sensory storage and short-term visual memory. *Percept Psychophys* 16:283–290.
- Brainerd CJ, Reyna VF (2005) *The Science of False Memory* (Oxford Univ Press, New York)
- Standing L (1973) Learning 10,000 pictures. *Q J Exp Psychol* 25:207–222.
- Shepard RN (1967) Recognition memory for words, sentences, and pictures. *J Verb Learn Verb Behav* 6:156–163.
- Standing L, Conezio J, Haber RN (1970) Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychon Sci* 19:73–74.
- Rensink RA, O'Regan JK, Clark JJ (1997) To see or not to see: The need for attention to perceive changes in scenes. *Psychol Sci* 8:368–373.
- Simons DJ, Levin DT (1997) Change blindness. *Trends Cogn Sci* 1:261–267.
- Loftus EF (2003) Our changeable memories: Legal and practical implications. *Nat Rev Neurosci* 4:231–234.
- Chun MM (2003) in *Cognitive Vision. Psychology of Learning and Motivation: Advances in Research and Theory*, eds Irwin D, Ross BH (Academic, San Diego, CA), Vol 42, pp 79–108.
- Wolfe JM (1998) Visual Memory: What do you know about what you saw? *Curr Biol* 8:R303–R304.
- O'Regan JK, Noë A (2001) A sensorimotor account of vision and visual consciousness. *Behav Brain Sci* 24:939–1011.
- Hollingworth A (2004) Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *J Exp Psychol Hum Percept Perform* 30:519–537.
- Tatler BW, Melcher D (2007) Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception* 36:1715–1729.
- Vogt S, Magnussen S (2007) Long-term memory for 400 pictures on a common theme. *Exp Psychol* 54:298–303.
- Castelhano M, Henderson J (2005) Incidental visual memory for objects in scenes. *Vis Cog* 12:1017–1040.
- Marmie WR, Healy AF (2004) Memory for common objects: Brief intentional study is sufficient to overcome poor recall of US coin features. *Appl Cognit Psychol* 18:445–453.
- Koutstaal W, Schacter DL (1997) Gist-based false recognition of pictures in older and younger adults. *J Mem Lang* 37:555–583.
- Intraub H, Hoffman JE (1992) Remembering scenes that were never seen: Reading and visual memory. *Am J Psychol* 105:101–114.
- Landauer TK (1986) How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognit Sci* 10:477–493.
- Dudai Y (1997) How big is human memory, or on being just useful enough. *Learn Mem* 3:341–365.
- Fagot J, Cook RG (2006) Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proc Natl Acad Sci USA* 103:17564–17567.
- Fuster JM (2003) More than working memory rides on long-term memory. *Behav Brain Sci* 26:737.
- Palmer TJ, Tarr MJ (2008) *Visual Memory*, eds Luck SJ, Hollingworth A (Oxford Univ Press, New York), 163–207.
- Fodor JA (1975). *The Language of Thought* (Harvard Univ Press, Cambridge, MA).
- Barsalou LW (1999) Perceptual symbol systems. *Behav Brain Sci* 22:577–660.
- Barsalou LW (1990) *Content and Process Specificity in the Effects of Prior Experiences*, eds Srull TK, Wyer RS, Jr (Erlbaum, Hillsdale, NJ), pp 61–88.
- Collins AM, Loftus EF (1975) A spreading activation theory of semantic processing. *Psychol Rev* 82:407–428.
- Anderson JR (1983) Retrieval of information from long-term memory. *Science* 220:25–30.
- Sigala N, Gabbiani F, Logothetis NK (2002) Visual categorization and object representation in monkeys and humans. *J Cog Neuro* 14:187–198.
- Mandler G (1980) Recognizing: The judgment of previous occurrence. *Psychol Rev* 87:252–271.
- Yonelinas AP (2002) The nature of recollection and familiarity: A review of 30 years of research. *J Mem Lang* 46:441–517.
- Jacoby LL (1991) A process dissociation framework: Separating automatic from intentional uses of memory. *J Mem Lang* 30:513–541.
- Tarr MJ, Bulthoff HH (1998) Image-based object recognition in man, monkey and machine. *Cognition* 67:1–20.
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.
- Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5:552–563.
- Tarr MJ, Williams P, Hayward WG, Gauthier, I. (1998) Three-dimensional object recognition is viewpoint dependent. *Nat Neurosci* 1:275–277.
- Bulthoff HH, Edelman S (1992) Psychological support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci USA* 89:60–64.
- Torralba A, Fergus R, Freeman W, 80 million tiny images: A large dataset for non-parametric object and scene recognition. *IEEE Trans PAMI*, in press.
- Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen* 115:39–61.
- Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8:457–464.
- Wheeler ME, Petersen SE, Buckner RL (2000) Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci USA* 97:11125–11129.
- Kosslyn SM, et al. (1999) The Role of Area 17 in Visual Imagery: Convergent Evidence from PET and rTMS. *Science* 284:167–170.
- Garoff RJ, Slotnick SD, Schacter DL (2005) The neural origins of specific and general memory: The role of fusiform cortex. *Neuropsychologia* 43:847–859.

Accuracy detecting repeat items during the study session

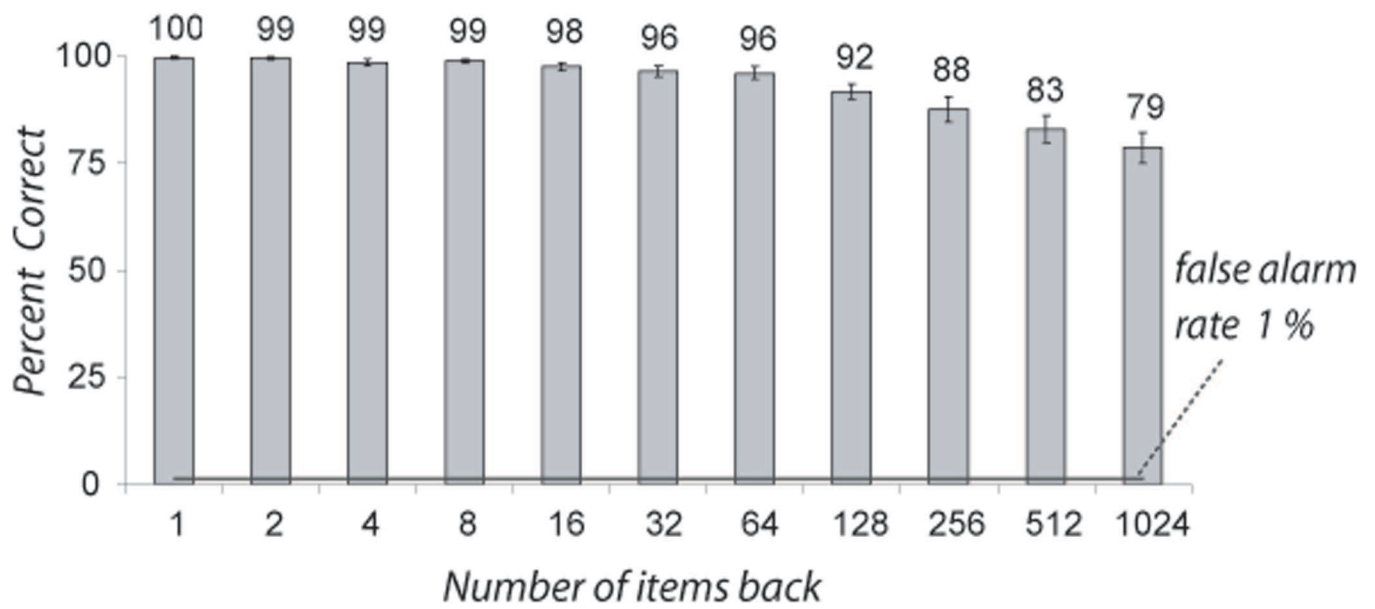


Fig. 51. Performance on detecting repeat images during the 5.5 h study session. Images were repeated with a different number of intervening items, from 0 to 1,023, by powers of 2.

Supporting Information

Brady et al. 10.1073/pnas.0803390105

SI Text

Modeling Methods. Our model of memory capacity in bits per item is based on work by Landauer (1), who provided a method for estimating the number of bits required to correctly make a decision about which items have been seen and which have not. This simple model assigns each picture a random code (b bits long), rather than assign them based on visual similarity. If a foil item is assigned the same code as any old item, the old item cannot be distinguished from the foil. Because this is a two-alternative forced-choice task, an error will occur on one half of these cases. Given this model we can compute the length of code (in bits) necessary to achieve any given accuracy level for a given number of items.

For example, to achieve 88% correct with 1,000 items in memory, you must have a code at least 11.9 bits long (3,821 possible codes). The chance of a new item being assigned a code that overlaps an old item would then be 23% ($[1 - 1/3,821]^{1,000}$), and on half of trials observers would still answer correctly, resulting in 12% errors, or 88% correct performance. In general, the number of bits is related to memory performance in this model by the following equation: $b = -\log_2[1 - (2p - 1)^{1/n}]$, where p is the percentage correct, and n is the number of items in memory.

If this model is capturing a systematic property of memory, similar estimates for the number of bits should be obtained by using different numbers of pictures in memory (because performance should increase correspondingly). Supporting this, Landauer (1) found that similar calculations result in estimates of 10.0 bits with 20 pictures in memory and 99% correct judgments, 10.2 bits with 400 pictures in memory and 86% correct judgments, etc. By using this model, we find observers must have 13.8 bits of information per item based on our novel condition (92% correct with 2,500 pictures in memory). This suggests that the maximum capacity of memory, if all items were coded by using the optimal set of features (decision-level bits), would be $2^{13.8}$ (14,000) unique items.

To model the exemplar and state results, we make the assumption that memory is organized hierarchically, such that the bits for the category appear before the bits for the exemplar per state, resulting in greater similarity in the codes (and greater chance of confusion) for items within a category than items in different categories. In the exemplar condition, observers are holding one item of the category in memory and this results in $\approx 87.5\%$ correct. If observers have two bits of exemplar-level information about each studied item, then the probability of the a foil exemplar overlapping with a studied item would be 25% ($1/2^2$), resulting in 12.5% error trials, thus 87.5% correct performance, matching our empirical results. The same logic holds in the state condition which also has $\approx 87.5\%$ performance. Thus, our overall estimate of memory capacity is 17.8 bits per item, where 13.8 bits are required to code the category of the object, and the additional four bits per item are required to code which exemplar (two bits) and state (two bits) the object is to

successfully distinguish it from the foils. This suggests that maximum number of unique items that can be put in memory (assuming an optimal feature set) is $2^{17.8}$, or 228,000.

This model, in which the exemplar bits are separate from the category bits, is more conservative than giving unique codes without regard to category, since it accounts for the idea that we are more likely to confuse two teacups than a teacup and a tractor. However, our calculations do assume that the exemplar and state conditions draw on different bits, such that the information used to perform well in the exemplar tests is not the same as the information used for the state tests. This is compatible with memory representations in which it is possible to know that you saw an open door rather than a closed door (state condition) without knowing exactly which door it was (exemplar condition). However, even if up to half of the information was shared between the exemplar and state condition, we would still obtain an estimate of 16.8 bits of information, or 114,000 unique codes, still an order of magnitude over previous estimates.

A previous study by Hollingworth (2) also examined object representations on the order of hundreds of images. Participants were shown scenes with many embedded objects and were subsequently tested with exemplar-level foil items. To quantify the capacity of memory estimated in this experiment, we used the same hierarchical model. This study did not include a novel test condition to estimate the category bits, so we assumed a generous 99% performance, giving 14.27 bits. Performance in the exemplar level tests was 65%, or an additional 0.5 bits. Thus we estimate the memory capacity demonstrated in Hollingworth (2) between 14–15 bits, approximately equal to the estimates arrived at by Landauer (1). Interestingly, this study demonstrates that memory can store hundreds of objects with exemplar-level fidelity, and even this does not guarantee an increased estimate of memory capacity.

The hierarchical decision-level model is based on the number of items studied, and importantly, on the number of questions asked about each item. Previous large-scale memory studies only tested against a novel foil (one question). Here, we ask about novel, exemplar, and state comparisons (three questions, thus three sets of bits). Our estimate of memory capacity could be increased even further if more questions were asked about what kind of information observers have about remembered items. However, there is an important caution to this modeling approach: the questions asked about each item are probably not independent. For example, we could have included a fourth kind of test asking about the orientation of the presented object, and added more bits to the estimated code of each item. However, information about the state is probably also informative about orientation. Thus, asking 1,000 different questions to show an enormous memory capacity estimate is not sufficient because such questions will likely have overlapping information when considered under the true coding model of the visual system. On a positive note, if one could ask the right set of completely independent questions, one might be approximating the visual coding scheme.

1. Landauer TK (1986) How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognit Sci* 10:477–493.
2. Hollingworth A (2004) Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *J Exp Psychol Hum Percept Perform* 30:519–537.