Psychon Bull Rev (2011) 18:469–475
DOI 10.3758/s13423-011-0075-z

How many pixels make a memory? Picture memory for small pictures

Jeremy M. Wolfe · Yoana I. Kuzmova

Published online: 5 March 2011

Abstract Torralba (*Visual Neuroscience, 26*, 123–131, 2009) showed that, if the resolution of images of scenes were reduced to the information present in very small "thumbnail images," those scenes could still be recognized. The objects in those degraded scenes could be identified, even though it would be impossible to identify them if they were removed from the scene context. Can tiny and/or degraded scenes be remembered, or are they like brief presentations, identified but not remembered. We report that memory for tiny and degraded scenes parallels the recognizability of those scenes. You can remember a scene to approximately the degree to which you can classify it. Interestingly, there is a striking asymmetry in memory when scenes are not the same size on their initial appearance and subsequent test. Memory for a large, full-resolution stimulus can be tested with a small, degraded stimulus. However, memory for a small stimulus is not retrieved when it is tested with a large stimulus.

In our online world, we consume a large number of thumbnail images. Those images contain less information than larger images of the same scene. Torralba (2009) wanted to measure the limit of how small an image could be while still conveying information about a scene. He took 256 × 256 pixel scenes, down-sampled and low-pass filtered them to create images from 4 × 4 pixels up to 256 × 256. He then up-sampled the smaller images to a 256 × 256 size, creating stimuli of the same physical size but of different pixel *resolution*. Torralba found that people could successfully categorize images at rates of 70%–80% correct when those images were a mere 32 × 32 pixels in resolution. The participants were 50%–60% correct at 16 × 16 pixels, and they were still above chance (in a 12-alternative forced choice task) at 8 × 8 pixels. Moreover, they could identify multiple objects in these scenes, even though the representations of those objects were far too degraded to allow them to be identified outside of the scene context. Thus, the representation of a scene can be based on a small amount of information and does not require the identification of the objects in the scene prior to the categorization of that scene.

Scenes (and photorealistic images in general) also support a massive memory. Research over the past four decades has shown that a few seconds of exposure to a picture of a scene or an object is adequate to create a durable memory for that picture (Konkle, Brady, Alvarez, & Oliva, 2010a, 2010b; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989; Shepard, 1967; Standing, Conezio, & Haber, 1970). Literally thousands of images can be coded into memory and retained for extended periods at high levels of accuracy. Most recently, it has been shown that this ability is not limited to thousands of highly distinctive images. Observers are able to recall that they saw *this car* or *this beach* and not *that one* (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle et al., 2010a, 2010b). Similarly, they are able to recall the state of the object: They know that they saw *this car with the door open* and not *this car with the door closed* (Brady et al., 2008)

What is in the representation of pictures that allows for this remarkable ability? It does not seem to be raw perceptual features, since meaningless textures are remembered far more poorly than scenes (Wolfe, Horowitz, & Michod, 2007). Nor can the remembered representation be a merely categorical coding. If observers were just encod-

J. M. Wolfe (✉) · Y. I. Kuzmova
Harvard Medical School and Visual Attention Lab,
Brigham & Women's Hospital,
64 Sidney St. Suite. 170,
Cambridge, MA 02139-4170, USA
e-mail: wolfe@search.bwh.harvard.edu

ing "beach," they would not be able to discriminate between *this* beach and *that* beach on later test. We might gain some insight into the nature of the memory representation if we combined the picture memory task with the work of Torralba (2009). In this article, we will use the term *image size* to refer to the size on the screen and *image resolution* to refer to the grain of the underlying information. If we used this terminology to describe the Torralba experiment, we would say that he showed observers stimuli of fixed size (256×256) and varied resolution. In the experiments reported here, we looked at the effects of both variables. A number of other studies have looked at memory as a function of size (see Table 1 in Uttl, Graf, & Siegenthaler, 2007). All of them have used 2- or 3-fold variations in size; we extend the range of variation to 32-fold. Additionally, whereas previous studies have employed objects or faces as stimuli, here we used scenes, meaning that the objects present may have been represented by a very few pixels.

If we hold everything else constant, how might picture memory change as we reduce the resolution and/or size of the image? Three hypotheses are possible:

1. Memory performance, like categorization performance in Torralba (2009), could be a smooth function of image resolution. In that case, memory performance with tiny images should qualitatively resemble Torralba's report of categorization performance with the same images. Experiment 1 supports this hypothesis.
2. Memory performance could be dependent on the number of pixels presented (i.e., the physical size presented on the screen). We know that this cannot be true to the exclusion of Hypothesis 1, because abstract textures and real scenes with the same number of pixels produce very different memory performance (Wolfe et al., 2007). Nevertheless, it could be that a low-resolution image is more memorable if it is up-sampled to a larger size than if it is presented at a smaller size. However, Experiment 1 does not support this hypothesis; the large and small sizes produced similar performance if resolution was fixed.
3. When the images are directly viewed, it is possible, up to some level of degradation, to recognize the lower-resolution version of an image and a full-resolution version of the same image as related to each other. Seeing that low resolution is a version of full resolution is the same as seeing that full is a version of low. Experiment 2 shows that this is *not* the case in memory. If observers see a full-resolution image in the memory phase, they perform quite well if tested with a low-resolution image. However, if observers see a low-resolution image first and a full-resolution image second, they perform poorly on the memory task.

Experiment 1

Participants

Seventy-three observers were tested in the 11 conditions of Experiment 1. Observers ranged in age from 18 to 53 (average = 28). All had at least 20/25 visual acuity (with correction as needed) and could pass the Ishihara color test. All gave informed consent and were paid $10/h for their time. There were 10 observers per condition; thus, some observers were tested in multiple conditions (49 observers completed one condition, 15 completed two, 7 three, 1 four, and 1 five). There was no evidence that performance varied systematically as a function of number of testing sessions.

Method

In a condition, observers saw a sequence of 200 scenes. Each scene, with the obvious exception of the first, could be either *new* or *old*—a repeat of a previously presented

Table 1 Stimuli for experiment 1

Condition	Stimulus size (deg)	Stimulus size (pixels)	Resolution
8 × 8 small	.25 × .25	8	8
16 × 16 small1	.50 ×.50	16	16
16 × 16 small2	.50 ×.50	16	16
32 × 32 small1	1 × 1	32	32
32 × 32 small2	1 × 1	32	32
64 × 64 small1	2 × 2	64	64
64 × 64 small2	2 × 2	64	64
16 × 16 big	8 × 8	256	16
32 × 32 big	8 × 8	256	32
64 × 64 big	8 × 8	256	64
Original	8 × 8	256	256

"Small1" and "small2" refer to two ways of reducing a 256 × 256 original to a smaller onscreen size. For the present purposes, the two versions serve as an internal replication.

scene. No scene was presented more than twice in the sequence of 200 trials. A scene could be repeated 2, 4, 8, 16, 32, 64, or 128 trials after its first appearance (*lag*). The probabilistic algorithm that generated these sequences produced different sequences for each observer. On average, 58% of the trials were new scenes. Of the 42% that were old scenes, 6%–7% of trials were presented at each of the lags 2, 4, 8, 16, 32, and 64, with 4% of trials having a lag of 128. The average lag was 32 trials. On each trial, an image was presented for 2 s, and observers gave a "new"/"old" keypress response with no time pressure. Feedback was provided after each trial.

Scenes were taken from the stimuli developed for Torralba (2009) and represented a diverse collection of indoor and outdoor scenes. Two distinct sets of scenes were employed, each of which contained the same perceptually and semantically distinct categories of indoor and outdoor scenes (e.g., the locker room in Set 1 was matched by a different locker room in Set 2). As noted, some participants took part in more than one condition. When that happened, the two conditions were separated by at least 2 weeks and employed different sets of scenes.

The 11 conditions differed in the size and resolution of their images. Within each condition, all of the stimuli were of the same size and resolution. As noted, *size* refers to the onscreen size, and *resolution* refers to the dimensions of the $N \times N \times 3$ matrix that defined the image in the computer, as is shown in Table 1. Stimulus size (in degrees) gives the visual angle at a standard viewing distance of approximately 57 cm. Stimulus examples can be seen in Fig. 1. Further details about the stimuli can be found in Torralba (2009).

The stimuli were centrally presented square photographs on a black background. The experiment was run on 17-in. CRT monitors with their screen resolution set to 1,280 × 960 pixels and a refresh rate of 85 Hz.

Results

Average d' values are shown in Fig. 1 as a function of resolution. One can readily see that performance declines as resolution declines, and it does not seem to make much difference if the resolution-16 images are big or small. All of the conditions with resolution of less than 256 produce performance that is lower than that produced by the full-sized image. Except for the size 64 (small1) condition, all ts(9) > 3.9, all ps < .005, uncorrected for multiple comparisons. The size 64 (small1) condition yields $t(9) = 2.54$, $p = .03$. This would not survive correction for multiple comparisons, so it is possible to argue that the representation was not significantly worse for the 64-pixel blurred image than for the full-sized image.

Although performance declines as resolution declines, it remains above chance for all conditions [all ts(9) > 7.0, all ps

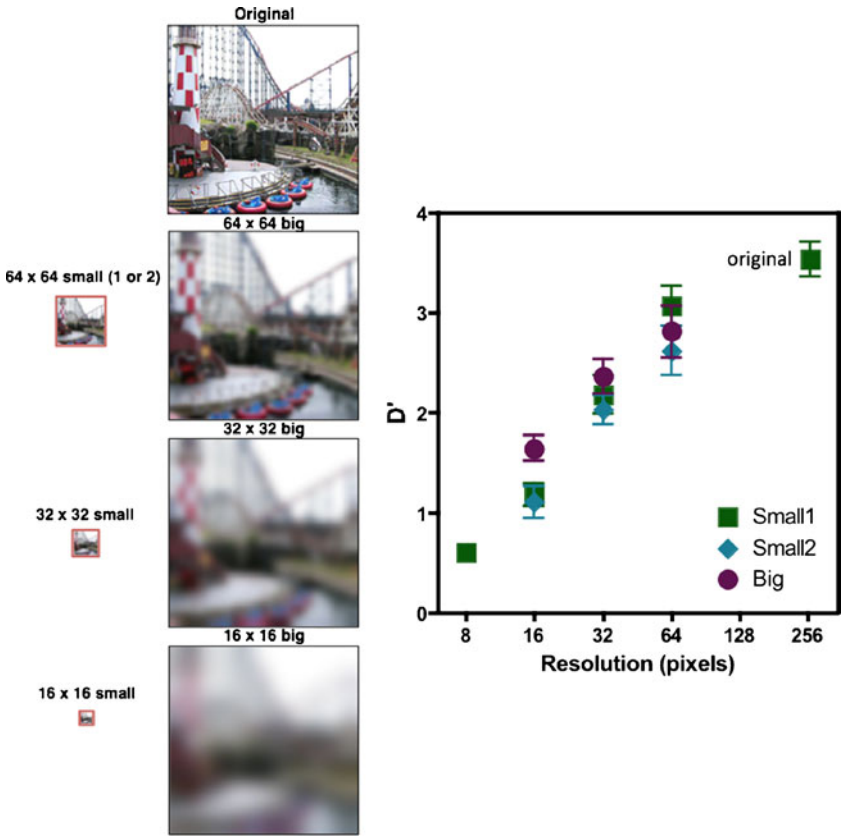

Fig. 1 Stimuli for the experiments are shown on the left. An original 256 × 256 image can be shrunk to produce the "small" versions shown on the far left. The shrunken representation can then be up-sampled to produce the "big" blurred images in the second column. Data on the right show a smooth decline in memory performance as the resolution changes

< .0001], even the extremely small 8×8 images (d' = 0.6, *SD* = 0.2). The mode of presentation does not seem to make much difference. A between-subjects ANOVA for the three resolutions (16, 32, 64) represented in each display mode (small1, small2, and big) revealed a main effect of resolution [$F(2, 81) = 50.2$, $p < .0001$] but no effect of display mode [$F(2, 81) = 2.7$, $p = .07$], and the interaction was not significant [$F(4, 81) = 0.96$, $p = .43$]. At 16 pixels, performance appears to be somewhat better for the large, up-sampled image than for either of the small, 16 × 16 images [both ts(18) > 2.6, ps < .02], although the level of significance is not convincing, given issues of multiple comparison. On the other hand, using a between-subjects design loses some power, so, if it were sufficiently interesting, this difference could be pursued with a stronger design. The similar decline of resolution in all conditions is the main effect, however.

Figure 2 shows that performance declines with lag. (Note that Fig. 2's *x*-axis is on a $\log_2$ scale.) The lower-resolution stimuli may seem to fall off more rapidly, but this is hard to assess, because the high-resolution stimuli are at near-ceiling performance for the shorter lags.

Discussion

The results show, somewhat unsurprisingly, that memory performance declines with declines in the quality of the images to be remembered. The lower-resolution images are remembered less effectively overall (Fig. 1) and are forgotten more quickly (Fig. 2). The nature of that decline is of more interest, because it could provide some insight into what is remembered when observers encode hundreds of images in a picture memory experiment. The function relating memory performance to image quality can be compared to the Torralba (2009) function relating categorization performance to image quality. This comparison is shown in Fig. 3.

Note that the memory and categorization tasks differ in their chance performance levels. The memory task is a 2-alternative, old versus new, forced choice. Torralba's (2009)

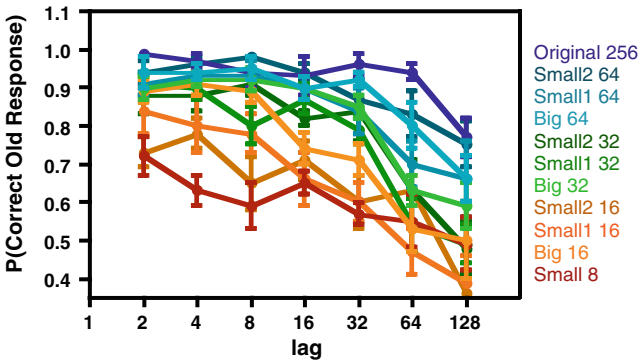

Fig. 2 Percent correct "old" responses as a function of lag between the first and second appearances of an image

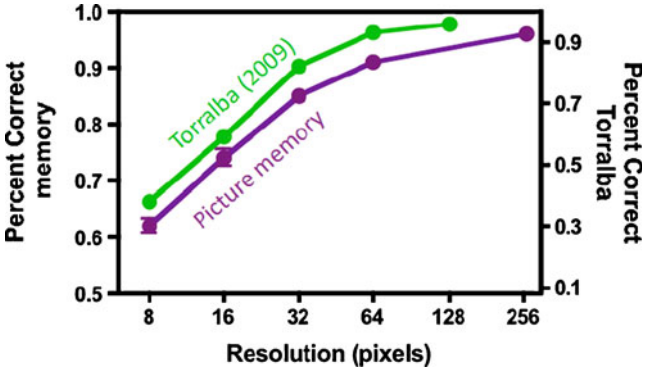

Fig. 3 Memory performance as a function of resolution, compared to data replotted from Fig. 2 of Torralba (2009)

categorization task, however, was a 12-alternative forced choice. To compare the functions, each has been plotted on an axis from chance to perfect performance. Chance is 1/12 for the Torralba experiment and .5 for the present study. Plotted in this way, we see that the functions are very similar in shape, with what appears to be a modest advantage for the categorization task. This seems plausible and consistent with the hypothesis that the representation that is remembered is similar to the representation that is used for categorization. As with the advantage for scenes over meaningless textures, knowing what you are seeing helps you remember it. The data suggest that if you encode the category correctly, you will remember the image.

It is of potentially greater interest that the small and large versions of the same-resolution images produce essentially the same results. The results did not need to take this form. Making an image smaller is like moving it farther away. In terms of spatial frequency, this shifts the Fourier spectrum to higher frequencies. Holding the size the same and decreasing the resolution, on the other hand, is like blurring the image. This manipulation selectively attenuates the higher spatial frequencies. Even though the changes in the Fourier domain are very different, it has been shown for other stimuli (Hayes, Morrone, & Burr, 1986; Loftus and Harley, 2005) that these manipulations produce similar results on recognition tasks. This has led to the idea that, within the limits imposed by the contrast sensitivity function, "cycles per object" rather than "cycles per degree" drive behavior (Hayes, Morrone, & Burr, 1986; Loftus and Harley, 2005). In the present experiments, "cycles per scene" might be the correct metric; we note with Torralba (2009) that individual objects may well have dropped below the recognition limit if they had been presented in isolation.

Experiment 2

In Experiment 1, observers saw items of the same size and resolution on both first and second exposures. What would

performance look like if observers saw a degraded stimulus first and a full-resolution stimulus second, or vice versa? In studies with single objects (Uttl et al., 2007) and faces (Kolers, Duchnicky, & Sundstroem, 1985), performance was worse when small stimuli were presented first and large stimuli second than when that order was reversed. Here, we look at that effect at a large size differential (8:1) and dissociate size and resolution.

Method

In Experiment 2, two stimulus types were intermixed in a block of 200 trials. In Experiment 2a, the stimuli were either small 32 × 32 images or full-sized, full-resolution 256 × 256 images. If an observer saw a small 32 × 32 version of a scene on that scene's first appearance, the observer saw the 256 × 256 version on the scene's second appearance, and vice versa. As in the first experiment, not all scenes were repeated. Observers were fully informed about the design. Half of the stimuli were at the small size and half at the large, presented in a pseudorandom order that did not allow observers to guess "old" versus "new" on the basis of size, although observers might draw some conclusions based on memory for the size of the first appearance. That is, if you are looking at a small image and you think you saw it as a small image before, you might be inclined to say that the current image was "new," since there were no small–small pairs. This design allowed us to determine whether a scene encoded as a small stimulus can be recovered when tested with a large stimulus, and vice versa. Since there are two conditions in each 200-trial block, this design reduced the number of stimuli per condition to 100 trials. To compensate, the number of observers was increased to 15. Other aspects of the method were the same as in Experiment 1.

Experiment 2b was identical to Experiment 2a, except that the stimuli were big and small 32 × 32 resolution images. If an observer saw a big 32 × 32 version of a scene on its first appearance, the second appearance was a small 32 × 32 image, and vice versa. Thus, in Experiment 2b, the big and small stimuli contained essentially the same amount of information at resolution 32. Twelve observers were tested in Experiment 2b. In all other respects, Experiment 2 followed the procedures of Experiment 1.

Results

Figure 4 reveals a striking asymmetry between the two conditions of Experiment 2a. Specifically, memory performance was terrible when the initially encoded image was small (32 × 32 pixels). Performance in this condition did not differ from chance [$t(14) = 1.012$, $p = .33$]. When the initial encoding was of the full-resolution, 256 × 256 image, performance on a 32 × 32 test stimulus was comparable to performance at a resolution of 32 pixels in Experiment 1 (cf. Fig. 1).

This impression is borne out statistically. On first appearance, the 256 × 256 stimuli were more likely to be labeled as "new" than were the 32 × 32 stimuli [89% vs. 70%; $t(9) = 5.75$, $p = .0003$]. On second appearance, the 256 × 256 stimuli were much more likely to be incorrectly labeled as "new" (because the first appearance of that scene was the degraded 32 × 32 version) [64% vs. 27%; $t(9) = 7.76$, $p < .0001$]. Of course, the d' difference, shown in Fig. 4, was highly statistically reliable [$t(9) = 10.5$, $p < .0001$].

These results can be compared to those from the Experiment 1 condition in which 32 × 32 images were used on both first and second appearances. Interestingly, performance on the 32→256 condition of Experiment 2a was significantly worse than performance in the 32→32 condition

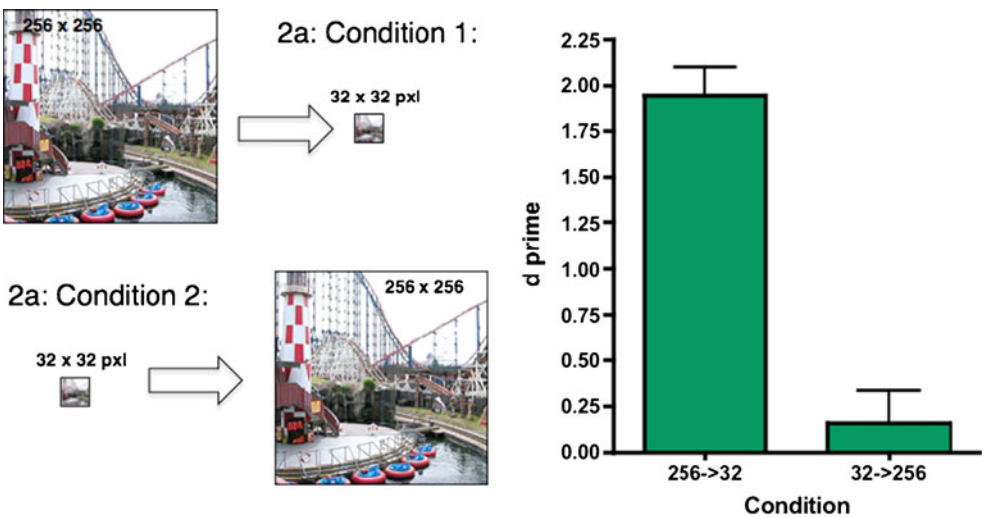

Fig. 4 Results of Experiment 2a. When observers initially encode a full-resolution image, memory is quite good when probed with a small, 32 × 32 image. However, when observers initially see a 32 × 32 image, they are at chance when later asked about a full-resolution image

of Experiment 1 [$t(23) = 8.07$, $p < .0001$]. Performance on the 256→32 condition did not differ from the 32→32 condition [$t(23) = 0.37$, $p = .72$]. Thus, 32→256 performance was low not because observers could not encode a 32 × 32 image, but because what was encoded could not be found and matched to a subsequent 256 × 256 image.

Is this asymmetry a matter of size or resolution? Perhaps observers simply cannot match a large image to a previously encoded small image. Experiment 2b falsified this hypothesis. In this experiment, where both the large and small stimuli were based on resolution-32 images, there was no asymmetry. Performance was very similar in the two conditions. As shown in Fig. 5, it did not seem to matter here whether the large stimulus was seen first or second.

The difference in *d'*s between conditions was not significant [$t(11) = 0.83$, $p = .42$]. Performance in Experiment 2b was somewhat worse than performance in the 32 × 32 condition of Experiment 1 [d' in Experiment 1 = 2.03; difference with 32→256, $t(20) = 2.69$, $p = .014$; difference with 256→32, $t(20) = 2.28$, $p = .034$]. Of most interest, performance in the 32→256 condition of Experiment 2b was far better than performance on the 32→256 condition of Experiment 2a [$t(25) = 5.8$, $p < .0001$]. Recall that the difference is that the 256×256 stimulus was degraded in Experiment 2b but not in 2a, so here we have the curious finding that degrading the test stimulus improved memory performance even though (or, more likely, because) the second stimulus in Experiment 2a was of higher quality. The 256→32 conditions did not differ reliably between Experiments 2a and 2b [$t(20) = 1.81$, $p = .08$].

Discussion

Experiment 2 points to an interesting distinction between the perceptual and mnemonic representations of these scenes. Looking at the larger scenes in Figs. 4 and 5, we would have no difficulty in agreeing that the images in Fig. 4 are sharper, more detailed renditions of those in Fig. 5. If asked to pick which sharp image was the original of some blurred or small version, we would have no difficulty making that choice. However, if we need to rely on our memory for the degraded image, we are then unable to match the degraded version with a sharp image at the time of test. The 32→256 condition of Experiment 2a produced chance performance, which was not due to an inability to encode pictures at a resolution of 32 pixels; Performance with such stimuli was well above chance in Experiment 1 and in the 32→256 condition of Experiment 2b. Rather, it appears that there is something in the representation of the original 256 × 256 image that, in the memory of the 32-pixel images, is not merely degraded, but absent. As a result, the 256-pixel image looks "new" even when it is old. It is interesting that the results are as asymmetric as they are, since the representation of the 256-pixel image seems to contain some version of the representation of the 32-pixel image. In the 256→32 condition of Experiment 2a, performance was quite good, which it would not have been if the internal representations of the low- and high-resolution images were fundamentally different.

How could this come to pass? Uttl et al. (2007) argued that the related asymmetry in their data was consistent with the Loftus and Harley (2005) "distance-as-filtering" hypothesis. However, that hypothesis does not seem to predict that performance would drop to chance in the 32→256 condition of Experiment 2a. One might imagine that the filtered representation of the small item could be found, albeit imperfectly, in the larger test stimulus. A different possibility is that the 32 × 32 image is misclassified on its first appearance, since at this size and resolution the image lends itself to various speculative semantic interpretations (Bruner and Potter, 1964). When the readily interpreted 256 × 256

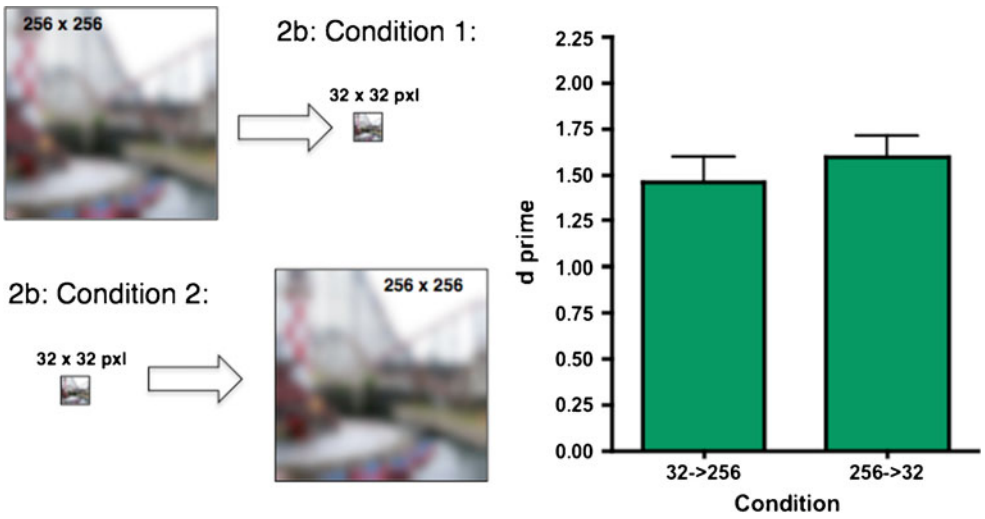

Fig. 5 Results of Experiment 2b. Although they differ in size, the large and small stimuli of this experiment both have a resolution of 32. In this case, it does not matter whether the large or the small stimulus is presented first. Both pairings produce reasonable memory

stimulus appears later, the observer is left with the impression that this is the first time he or she has seen that particular scene. Thus, recall fails in this 32→256 condition. In the 256→32 case, however, the observer, primed by the 256-pixel image, is able to see the degraded 32-resolution image as the same scene. In the 32→256 case of Experiment 2b, one may imagine that the misclassification of the large degraded stimulus might be more likely to match the misclassification of the small degraded image, and performance rises above chance.

Another way to phrase this is that the representation of the low-resolution image is not just a subset of the representation of the high-resolution image. If it were, one might expect the observer to be able to recognize the presence of the subset in the superset in the 32→256 condition of Experiment 2a. However, the stored representation of the 32-resolution image is not recalled when the 256-resolution image is presented.

To summarize, the degradation of memory for scenes as resolution is reduced closely parallels the degradation of forced choice scene categorization, suggesting a similarity between the visual information that supports categorization and the information that is encoded into memory. The asymmetry in the results of Experiment 2a suggests that the representation of a low-resolution scene is not a simple subset of the representation of the higher-resolution original.

Author Note This research was supported by Grants NIH-EY017001 and ONR-MURI N000141010278. We thank Antonio Torralba for the use of his images. These images, at multiple resolutions, can be downloaded from http://people.csail.mit.edu/torralba/tmp/tinyMemory/. We thank Talia Konkle for insightful comments on an earlier version of this article.

References

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

Bruner, J., & Potter, M. (1964). Interference in visual recognition. *Science, 144*, 424–425.

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception, 15*(5), 595–602.

Kolers, P. A., Duchnicky, R. L., & Sundstroem, G. (1985). Size in the visual processing of faces and words. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 726–751.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General, 139* (3), 558–578. doi:10.1037/a0019165

Konkle, T., Brady, T., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*, 1551–1556.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review, 12*(1), 43–65.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of conistency with schema expectations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 587–595.

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

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73–74.

Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience, 26*, 123–131.

Uttl, B., Graf, P., & Siegenthaler, A. L. (2007). Influence of object size on baseline identification, priming, and explicit memory. *Scandinavian Journal of Psychology, 48*, 281–288.

Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research, 47*, 955–964.