How Low Can You Go? Detecting Style in Extremely Low Resolution Images

Rachel A. Searston
The University of Adelaide

Matthew B. Thompson
Murdoch University

John R. Vokey
University of Lethbridge

Luke A. French and Jason M. Tangen
The University of Queensland

Humans can see through the complexity of scenes, faces, and objects by quickly extracting their redundant low-spatial and low-dimensional global properties, or their *style*. It remains unclear, however, whether semantic coding is necessary, or whether visual stylistic information is sufficient, for people to recognize and discriminate complex images and categories. In two experiments, we systematically reduce the resolution of hundreds of unique paintings, birds, and faces, and test people's ability to discriminate and recognize them. We show that the stylistic information retained at extremely low image resolutions is sufficient for visual recognition of images and visual discrimination of categories. Averaging over the 3 domains, people were able to reliably recognize images reduced down to a single pixel, with large differences from chance discriminability across 8 different image resolutions. People were also able to discriminate categories substantially above chance with an image resolution as low as 2×2 pixels. We situate our findings in the context of contemporary computational accounts of visual recognition and contend that explicit encoding of the local features in the image, or knowledge of the semantic category, is not necessary for recognizing and distinguishing complex visual stimuli.

Public Significance Statement
Accurate recognition and discrimination of complex visual stimuli is critical to human decision making in medicine, forensic science, aviation, security, and defense. This study highlights the sufficiency of redundant low-spatial and low-dimensional information for visual recognition and visual discrimination of 3 large-scale natural image sets.

Keywords: visual recognition, visual discrimination, ensemble, gist, perceptual expertise

> For me, music and life are all about style.
>
> —Miles Davis

Humans have a remarkable ability to pick up on redundant visual *stylistic* information in impoverished images and under noisy conditions. Radiologists, for example, can detect abnormal mammograms when given a half-second glimpse (Evans, Georgian-Smith, Tambouret, Birdwell, & Wolfe, 2013; Evans, Haygood, Cooper, Culpan, & Wolfe, 2016). Fingerprint experts can identify prints that belong to the same finger at a glance and filtered with artificial noise (Thompson & Tangen, 2014). They can even detect pairs of prints left by *different* fingers of the same person (Searston & Tangen, 2017c) and can search an array of fingerprints for a target more rapidly and accurately than a novice (Searston & Tangen, 2017a, 2017b). Looking to more conventional perceptual domains, earlier demonstrations in face recognition revealed that "coarse quantized" or reduced resolution versions of a familiar face image (e.g., the *Mona Lisa*) remained recognizable at just 16 × 16 pixels (Harmon & Julesz, 1973; Morrone, Burr, & Ross, 1983). Torralba (2009) revived this resolution reduction approach and showed that people can identify the semantic category in a variety of indoor (e.g., kitchen, office) and outdoor (e.g., forest, beach) scenes with 81.8% accuracy when each of the images were 32 × 32 color pixels. This same resolution seems to be sufficient for a similar level of image recognition performance (Wolfe & Kuzmova, 2011). We extend on this body of empirical work here to explore *how* we accomplish these perceptual feats with so little information at hand.

Prior work that has pushed the limits of human visual recognition with natural images seems to suggest that global properties are sufficient for classification and recognition. Indeed, contemporary computational accounts of visual recognition consistently emphasize the importance of the ensemble, the gist, and the spatial envelope for interpreting the contents of natural images, referencing "spatially distributive," "non-spatially localized," or "low frequency" global properties (Brady, Shafer-Skelton, & Alvarez, 2017; Greene & Oliva, 2009; Oliva & Torralba, 2001). A rapid and automatic process of extracting coarse-grained global information (e.g., color blobs) is thought to constrain subsequent selective attention and analysis of finer-grained local image properties (e.g., the subtle dimpling that bookends Mona Lisa's smile; Oliva & Torralba, 2006; Schyns & Oliva, 1994). In scene recognition terms, Torralba (2009) describes *gist* as a summary of the semantic content of the scene (i.e., its category label), its spatial layout, and a few objects that compose the scene. In face recognition terms, holistic or *configural* information refers to the spatial relationships between face features (Richler, Mack, Gauthier, & Palmeri, 2009). Although these descriptions are homing in on the same concept, we think there is inconsistency in the literature about whether references to such global properties are attempting to describe information that is distributed *within* a particular image, or information that is distributed *across* images, or a mixture of both. These two different kinds of global information, and their relative necessity for visual recognition, appear to be conflated or unspecified in current theorizing.

We use visual *style* here to describe the residual redundant information distributed within and across extremely low resolution image sets. A complete exposition of the concept is beyond the reach of this article; however, we conceive of visual stylistic information as representing the earliest principle components of a given pixel space. Visual style embodies the covariant structure, and not simply the average, of such global properties as low spatial frequency information, configural information, and low-level dimensions like hue, luminance, orientation, and texture. Style can be distinguished from other terms used to describe global information because it does not embody any semantic properties (e.g., category labels), or local features (e.g., high-level spatial information), and can include both spatial (e.g., configuration) and dimensional (e.g., color) information.

The resolution reduction method first applied by Harmon and Julesz (1973) and Morrone et al. (1983) provides a simple but powerful means of testing the sufficiency of global information for performing a range of perceptual tasks. Reducing the resolution of an image or downsampling effectively eliminates the noisiest, least frequent pixels. To the extent that one image shares a pixel-space with other images in a set, reducing their image resolution also has the effect of reducing high-dimensional noise across the entire set—leaving behind the most redundant, low-dimensional category information. The remaining redundant information might be described as a distributed configuration of global features, like hue and luminance, that covary across images. Indeed, Morrone et al. point out that high spatial frequency information is minimal compared with low spatial frequency information in extremely low "blocked" resolution images (like those produced using a simple nearest neighbor resampling procedure). By reducing local image features so drastically, and retaining the visual structure across the image set, we can examine the sufficiency of this information for a range of fundamental perceptual tasks: visual object recognition, visual categorization, and, as we focus on here, visual discrimination.

Torralba (2009) presented 28 participants with 240 images of indoor and outdoor scenes that were all the same physical size (i.e., all upsampled to 256 × 256 pixels) but varied in the number of pixels, ranging from 4 × 4 to 128 × 128. They used a 12-alternative forced choice task in which participants were asked to classify each image as a member of one of 12 semantic categories (six indoor and six outdoor scenes). Chance level in their experiment was 8.3% and, as a colored image lower bound, 4 × 4 pixels was enough to produce a correct categorization rate of 18.4%. Wolfe and Kuzmova (2011) built on this finding by using a two-alternative forced choice image recognition memory task (with an average lag of 32 trials) instead of the categorization task. When accounting for the different chance levels, they found a similar pattern of performance as Torralba across image resolutions ranging from 8 × 8 pixels up to 256 × 256 (but did not go any lower than 8 × 8 pixels). The fact that people can categorize complex scenes with a resolution as low as 4 × 4 pixels, and recognize them with 8 × 8 pixels, suggests that they are remarkably sensitive to global features within and distributed across images. Less clear is the basis for this sensitivity: Is semantic coding of the image category (the conceptual back end) necessary for above-chance sensitivity at extremely low resolutions? Is the global information retained in the low-resolution images (the perceptual front-end) sufficient in the absence of such semantic coding of the category or any local image features?

We address these questions by probing people's sensitivity to global image (Experiment 1) and category (Experiment 2) information using a resolution reduction methodology similar to prior

scene recognition work. However, we add some additional image resolutions and use an adapted experimental design to test just how low people can go without any cues to the semantic category or content of an image. First, in Experiment 1, we extend the work of Wolfe and Kuzmova (2011). Here, we test people's image recognition memory performance using eight image resolution conditions, reducing the images down from 128 × 128 to a single pixel. We also use three large sets of images from a variety of different domains (i.e., paintings, birds, and faces) to test the general lower bounds of human visual recognition with extreme image reduction. Wolfe and Kuzmova suggest that correctly encoding the semantic category—the conceptual back end—aids picture memory, resulting in better performance with higher resolutions; that is "knowing what you are seeing helps you remember it." We test how well people can recognize natural images without necessarily knowing what they are seeing by reducing the image resolution down to a single pixel and counterbalancing the categories across old and new images. With this design, even if people can recognize an "owl," a "Cubist" painting, or a "female" face from a single colored pixel, or configurations of four, eight, and even 16 pixels, the category label would not be diagnostic of whether they had seen that particular patch of pixels before or not, and thus cannot be relied on for recognition.

In Experiment 2, we extend the work of Torralba (2009). Identifying the semantic category of an image below 4 × 4 pixels poses quite a challenge. It may be possible, however, to detect the remaining global properties distributed *across* images at extremely low resolutions using a two-alternative category discrimination task—in which participants do not necessarily need to know the semantic categories in order to distinguish their perceptual features. Our guiding hypothesis across the two experiments is that global image features, distributed within and across images, are sufficient for above-chance detection of natural images and categories, even when the semantic content is no longer diagnostic or interpretable. Our preregistered predictions for both experiments are available on the Open Science Framework, along with our experiment code, deidentified data, and analysis script (https://osf .io/thk2a).

Experiment 1: Recognizing Natural Images at Extremely Low Resolutions

In Experiment 1, we test whether people can recognize natural images based on their most redundant visual features by gradually reducing their resolution down to 1 pixel. As noted by Wolfe and Kuzmova (2011), reducing the resolution of images appears to limit people's capacity to rely on conceptual understanding to aid image recognition. But even simple biological systems whose cognitive architecture is not likely to support the construction and use of abstract concepts—ants, bees, wasps, pigeons, and fish— can learn to recognize previously traveled routes (Wehner, Boyer, Loertscher, Sommer, & Menzi, 2006), specific artistic styles (Watanabe, 2001; Watanabe, Sakamoto, & Wakita, 1995; Wu, Moreno, Tangen, & Reinhard, 2013), photographs that contain people and photographs that do not (Aust & Huber, 2001; Herrnstein & Loveland, 1964), faces of conspecifics (Sheehan & Tibbetts, 2011; Siebeck, Parker, Sprenger, Mäthger, & Wallis, 2010), abnormal mammograms (Levenson, Krupinski, Navarro, & Wasserman, 2015), and four-letter words (Scarf et al., 2016; Scarf,

Corballis, Güntürkün, & Colombo, 2017; see Vokey, Jamieson, Tangen, Searston, & Allen, 2018, for a familiarity-based account of these findings). As such, it seems plausible that people, too, may be able to recognize images above chance under extremely low resolution conditions without knowledge of the original image category or its content.

In this experiment, participants viewed 24 blocks (8 image resolutions × 3 domains) of 24 trials. Within each block of trials, the first 12 consisted of a sequence of 12 images presented on a computer screen one after the other, and participants were asked to remember them. The second 12 consisted of another 12 images displayed one after the other, and participants were asked to identify whether each image was presented in the original sequence of 12 or not. We examined participants' average discriminability (*A*), response bias (*b*), and rate correct (RCS), collapsing across domains (i.e., paintings, birds, faces). Using a metric of discriminability, in which .5 is chance and 1 is perfect performance, we based our preregistered predictions on the results obtained by Wolfe and Kuzmova (2011), whose task and stimuli are most similar to this experiment. We predicted that participants' discriminability of old and new items would come close to .51 for the 1 pixel images, to .56 for the 2 × 2 pixel images, and .60 for the 4 × 4 pixel images, with a further discrimination improvement as the number of pixels in the images increased.

Method

We compiled three large sets of images for the two experiments, which we refer to as *domains*. Each domain consisted of two subcategories: birds (hawks and owls), faces (males and females), and paintings (cubist and impressionist). Figure 1 provides a sample of images from each domain and subcategory. To keep the size of the images constant while varying image resolution, all of
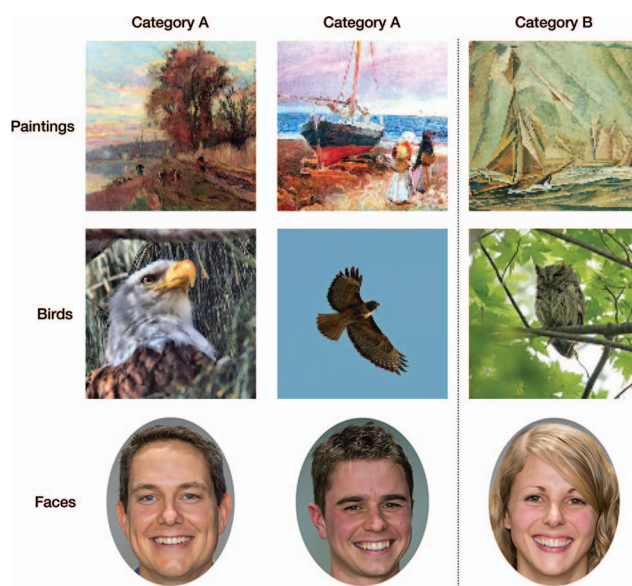

Figure 1. The images to the left of the dotted line are sample "same" category image pairs from the three domains at 128 × 128 pixels. The images to the right depict the corresponding "different" category samples. See the online article for the color version of this figure.

the images were first downsampled to eight different resolutions (128 × 128, 64 × 64, 32 × 32, 16 × 16, 8 × 8, 4 × 4, 2 × 2, and 1 × 1) and then upsampled to 256 × 256. Figures 2 and 3 provide an example image from each domain and image resolution.

Paintings. The paintings were from a larger art collection created for a previous project. The full collection contains 5,184 paintings, made up of 18 different paintings by 72 different artists in each of four different artistic styles (cubism, impressionism, realism, and renaissance; 288 artists in total). We only used the 1,296 cubist and 1,296 impressionist paintings here to equate the number of subcategories within the three image sets (2,592 painting in total). The paintings were collected manually from a variety of websites. We cross referenced the paintings to avoid duplicates and only downloaded images that were larger than 500 × 500 pixels. All of the paintings in the art collection were cropped to the center of the shortest dimension using a 1:1 (square) aspect ratio and resized (using nearest neighbor scaling) to 256 × 256 pixels. Any signatures on the paintings were also removed using the "Content Aware" fill tool in Photoshop.

Birds. The bird images were a subset of 1,502 images extracted from the Cornell Lab of Ornithology's NABirds V1 collection (downloaded from http://dl.allaboutbirds.org/nabirds). Seven hundred fifty-one images, or half of the subset, were natural photographs of birds from the Accipitridae family (e.g., hawks, eagles, kites, harriers, Old World vultures), and the other half were natural photographs of birds from the Strigidae family (owls). The subset contained 16 different species, with eight species in each family. The species were randomly selected to match the number of images between the two families, so both sets of eight species contained 120, 103, 103, 99, 73, 72, 95, and 86 images per species, respectively. We resized each of the images to 256 × 256 pixels using a nearest-neighbor scaling algorithm and removed any species labels using the "Content Aware" fill tool in Photoshop.

Faces. The faces are a collection of 1,772 faces images (886 female and 886 male) extracted from the 10k US Adult Faces Database (Bainbridge, Isola, & Oliva, 2013). The entire database contains 10,168 natural face photographs, and we obtained a

Figure 2. Panels A, B, and C depict participants' mean discriminability (*A*), response bias (*b*), and rate correct scores (in seconds) recognition memory task as a function of image resolution (*x*-axes), along with their polynomial trend over pixels at the top of the three panels. All plots represent the 50 participants' responses, collapsing over the three domains: paintings, birds, and faces. Panel D shows the receiver operating characteristic curves for the eight image resolutions, overlaid with the "best-fitting" curve assuming binormal distributions (the dotted line indicates chance performance). Finally, the raincloud plots in Panel E depict a half violin plot of participants' mean proportion correct scores across the eight image resolutions overlaid with jittered data points from each individual participant, the mean proportion correct per resolution (the black dot), and standard error of the mean per resolution. See the online article for the color version of this figure.
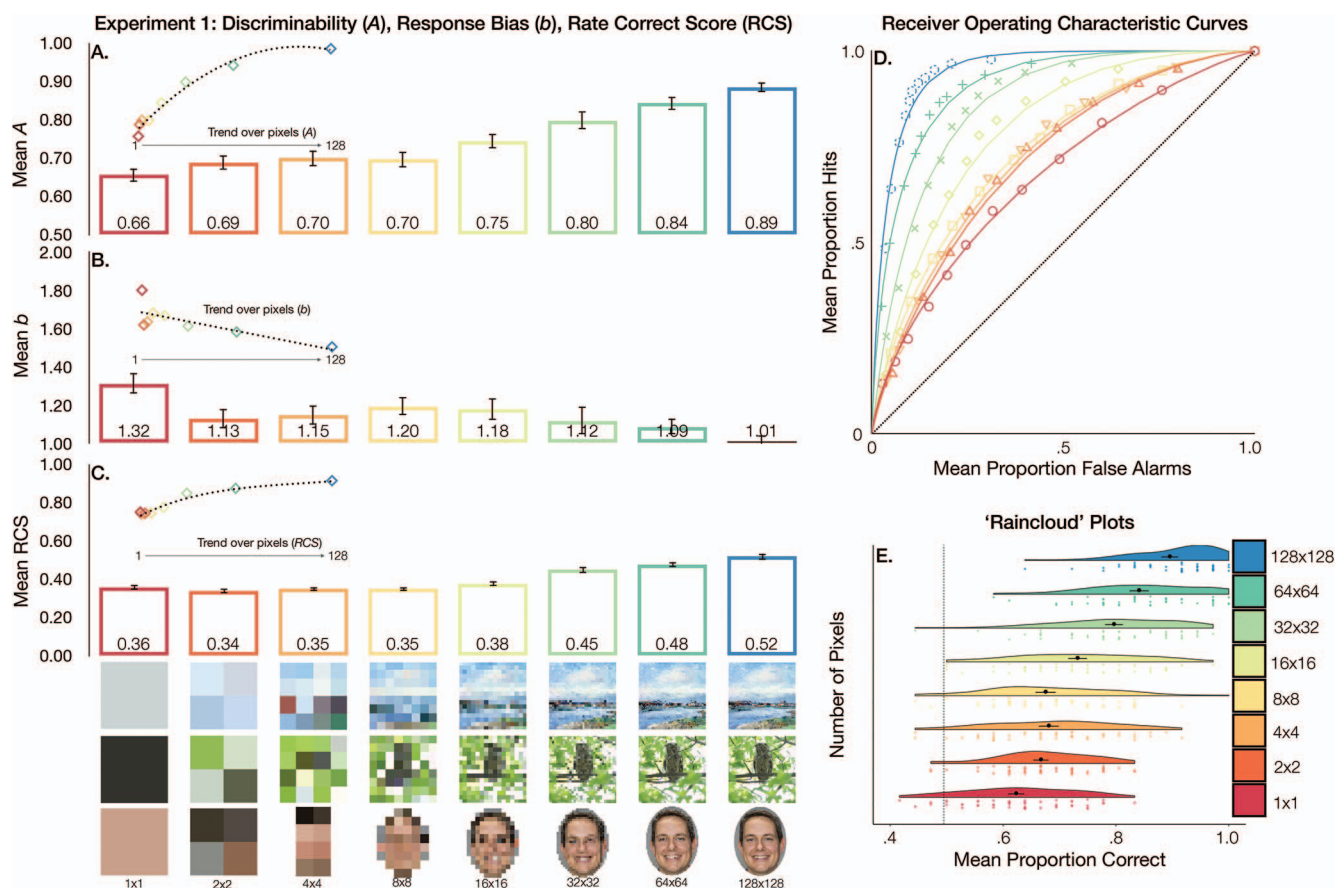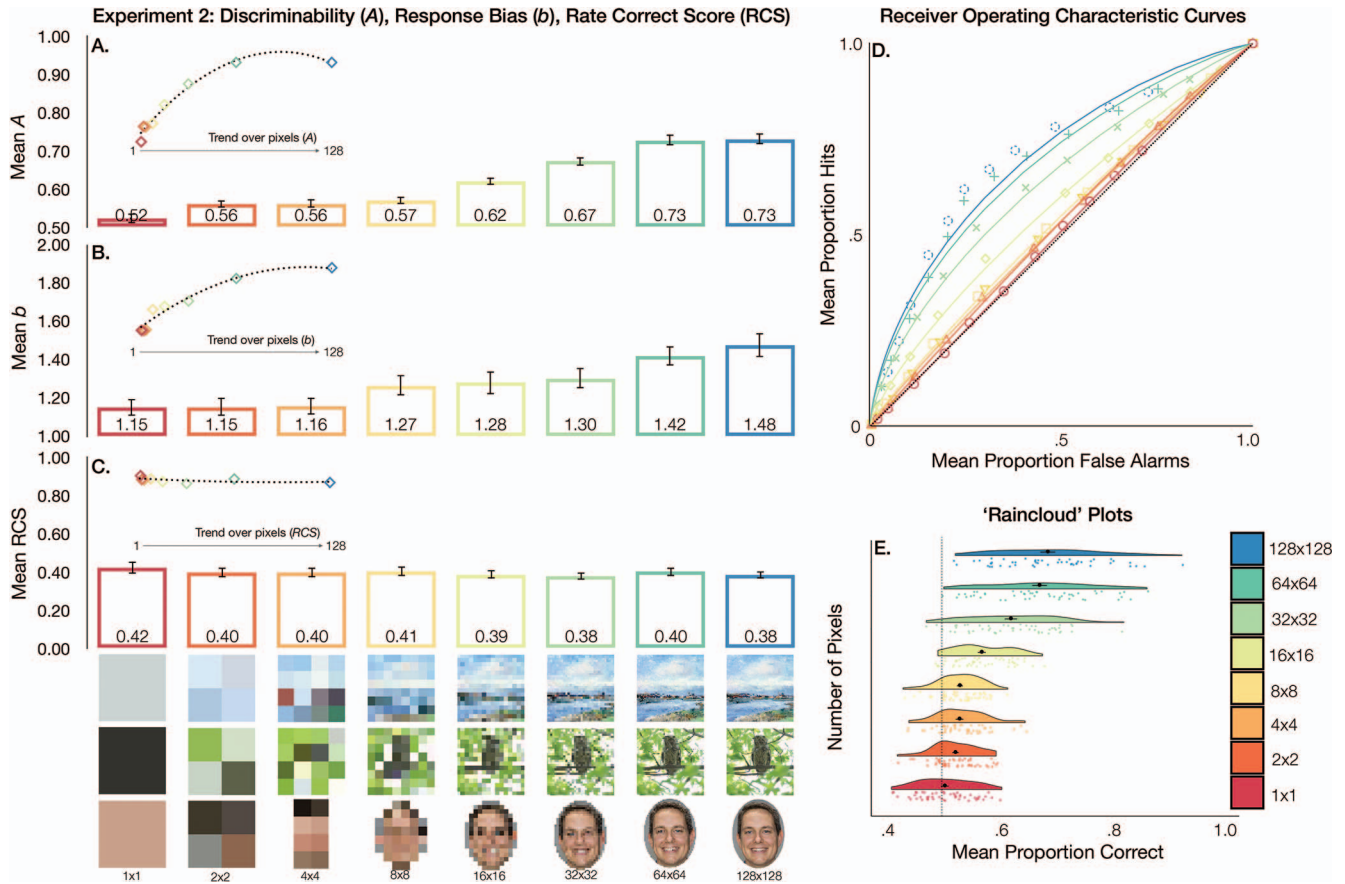
Figure 3. Panels A, B, and C depict participants' mean discriminability (*A*), response bias (*b*), and rate correct scores (in seconds) discrimination task as a function of image resolution (*x*-axes), along with their polynomial trend over pixels overhead. All plots represent the 50 participants' responses, collapsing over the three domains: paintings, birds, and faces. Panel D shows the receiver operating characteristic curves for the eight image resolutions, overlaid with the "best-fitting" curve assuming binormal distributions (the dotted line indicates chance performance). As in Experiment 1, the raincloud plots in Panel E depict a half violin plot of participants' mean proportion correct scores across the eight image resolutions, overlaid with jittered individual data points, the mean proportion correct per resolution (the black dot), and standard error of the mean per resolution. See the online article for the color version of this figure.

reduced set of 2,222 images from the database creators along with attribute data (e.g., memorability, attractiveness, happiness, friendliness). We then removed the famous faces from this subset, further reducing it to 2,063 faces (886 female, 1,177 male), and extracted the height and width dimensions of the remaining images. All of the images were 256 pixels tall, except one, which we scaled down to match the height of the others. The width of the images ranged from 153 to 157 pixels, and we resized them to 256×256 pixels. We deleted one image in the male domain that did not match these dimensions. Finally, we randomly sampled 886 male faces from the remaining 1,176, creating an equal number of male and female images.

Image resolution. We generated eight different versions of every image in our three sets, at 1 pixel, 2×2, 4×4, 8×8, 16×16, 32×32, 64×64, and 128×128 pixels, using a nearest-neighbor resampling procedure in MATLAB (see Figures 2 and 3 for an example image from each domain at each of the eight image resolutions). All images were then upsampled to 256×256 pixels.

Although upsampling a set of images using nearest-neighbor interpolation produces the blocky-looking images illustrated in Figures 1 and 2, the number of pixels in each condition is clear. One intuition about this technique is that the pixel edges or "blocked" appearance of the images introduced at the lower resolutions may serve as high spatial frequency noise masking the lower spatial dimensions. However, Morrone et al. (1983) showed that high spatial frequency information is minimal at extremely low resolutions, and that adding high spatial frequency noise can even help unmask low-dimensional information—similar to squinting, blurring, shrinking, or looking from a distance.

Participants. A sample of 50 undergraduate psychology students (33 female, 17 male) from The University of Queensland participated for course credit. Participants' mean age was 20.46 years (*SD* = 4.73). Ethics approval was received from The University of Queensland Behavioral and Social Sciences Ethical Review Committee (2014001677). Effect sizes were not reported in previous resolution reduction studies, making it

difficult to precisely predict magnitude of effects in our experiment ahead of time. Torralba (2009) tested 28 participants who were distributed equally across two groups (grayscale and color images). Wolfe and Kuzmova (2011) tested 73 subjects who were distributed across 11 groups (24 participants were tested in multiple conditions to make 10 participants per group). Considering our use of a within-subjects design and multiple image sets, we anticipated that our relatively large prespecified sample of 50 subjects would provide sufficient sensitivity to detect any meaningful differences. Indeed, a sensitivity analysis based on this sample suggested that the smallest effect size we can detect, with 80% power for our within-subjects analyses of variance (ANOVA), accounting for the 576 trials, was $f = .04$ ($\eta^2 = 0.002$ or a "small" effect). For the one-sample analyses comparing with chance (i.e., $A = .5$), the smallest effect size we can detect with 80% power was estimated to be $d = .40$ or a "small" to "medium" effect.

Procedure. Participants completed the experiment (programmed in LiveCode 8.0.0) on a MacBook Air laptop, with a screen resolution of 1440×900 pixels at 72 dpi. Participants first read an information sheet about the experiment, and then watched an instructional video on the nature of the materials, the image recognition task, and the response scale. Each participant then viewed a series of images, presented one after the other on a solid gray background (R: 209, G: 209, B: 209) to provide contrast and ensure that each image would not blend into the background (i.e., none of the 5,678 images across the three matched that particular shade of gray at 1 pixel). Participants completed 576 trials in total, broken into 3 (image set: birds, paintings, and faces) $\times$ 8 (image resolution: 1, 2×2, 4×4, 8×8, 16×16, 32×32, 64×64, and 128×128 pixels) blocks of 24. Each 24-trial block was a "memory" block divided into a learning phase and a test phase. In each learning phase, participants viewed 12 images appearing on screen for 4 s at a time, separated by a 500-ms interval. In each test phase, participants also viewed 12 images with an old–new response scale appearing on screen. Six of the test images were *new* and had not been shown previously in the experiment, whereas the other six were *old* and had been displayed at some point during the previous learning phase. On a given test trial, participants saw a single image on the screen and rated the extent to which they thought they had seen the image before or not. The image remained on screen until a response was made, but a "speed-up" prompt was delivered if the response time was slower than 4 s. Their responses were recorded by pressing one of 12 buttons on a forced-choice scale, with "old" responses ranging from 1 (*sure old*) to 6 (*unsure old*) and "new" responses ranging from 7 (*unsure new*) to 12 (*sure new*). The images were sampled equally from the two subcategories in each image set and presented in a different random order within the learning and test phases of the experiment. Participants were not provided with the category labels for any of the image sets, either in the instructional video or during the course of the experiment.

Results

Of primary interest was the effect of resolution on participants' image recognition memory performance when averaging over the three image sets or "domains" (birds, faces, and paintings). We

therefore report analyses of the effect of image resolution on participants' aggregated discriminability (A; see Zhang & Mueller, 2005, on the correct computation of A instead of A′), the associated nonparametric measure of response bias (b), and rate correct (RCS; see Vandierendonck, 2018, for an overview of RCS and other speed-accuracy metrics). Our R code and output for all plots, and primary and subsidiary analyses performed for both experiments, have been compiled into a single R Markdown HTML document, stored publicly on the Open Science Framework (https://osf.io/thk2a). Separate exploratory analyses of participants' performance with paintings, birds, and faces, in each of the experiments, have also been included in the R Markdown document as supplemental to the main preregistered analyses reported on here.

Discriminability. In our preregistered analysis plan, we specified that we would compute nonparametric discriminability (A′) because it is intuitive to interpret (e.g., 1 is perfect performance and .5 is chance) and does not assume normally distributed data. We use a corrected formula for A′ here that was introduced by Zhang and Mueller (2005) and is denoted as A. The new formula can be correctly interpreted as the average of the maximum area and minimum area under the proper receiver operating characteristic curve constrained by the hits and false alarms. We first converted participants' confidence ratings into hits and false alarms, and then computed their average discriminability (A) collapsing across domain, for all eight image resolutions. We also computed participants' empirical area under the curve (AUC; see Figure 2) to explore whether the pattern of results would change when we accounted for their full range of confidence ratings. Because these additional exploratory analyses yielded the same pattern of results (see the supplemental R Markdown document available on the Open Science Framework for a breakdown of the results and plots of the AUC data across image resolutions: https://osf.io/thk2a), we only report our planned analyses of discriminability using A here.

As shown in Figure 2A and Figure 2D, participants' discriminability of old and new images tended to increase with increasing image resolution: 1 pixel ($A = .66$, $SD = .12$), 2×2 ($A = .69$, $SD = .13$), 4×4 ($A = .70$, $SD = .15$), 8×8 ($A = .70$, $SD = .14$), 16×16 ($A = .75$, $SD = .14$), 32×32 ($A = .80$, $SD = .16$), 64×64 ($A = .84$, $SD = .11$), and 128×128 pixels ($A = .89$, $SD = .09$). As planned, a one-way within-subjects ANOVA using participants' A scores revealed a significant main effect of image resolution, $F(7, 343) = 32.19$, $p < .001$, $\eta^2G = .26$ (for a breakdown on using generalized eta-squared as the preferred effect size in an ANOVA context, see Bakeman, 2005). Given that the data distributions for some image resolutions appeared to be skewed, we also conducted a permutation test, permuting participants' A scores 1,000 times, to determine whether a more precise sampling distribution would produce the same result, and it did ($p < .001$). Polynomial contrasts further revealed a significant linear ($p < .001$) and quadratic ($p < .001$) trend to participants' discriminability (A) over image resolutions. This trend was observed even when we accounted for the quadratic increases in the number of pixels across conditions: discriminability (A) increased as the pixel levels went up as integer powers of two (i.e., 2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, and 2^7), with larger increases observed at the higher resolutions (marked by the slight concavity seen in Figure 2A; see the supplemental R Markdown document available on the Open

Science Framework for a full description of all seven orthogonal contrasts: https://osf.io/thk2a).

We conducted a further finer-grained analysis comparing participants' mean discriminability with each image resolution to chance performance ($A = .5$) and applied corrections to all p values to control the family-wise error rate (Holm, 1979). This analysis revealed that participants performed significantly greater than chance for all image resolution conditions, including the 1 pixel images, $t(49) = 9.36$, $p < .001$, $d = 1.32$. We report d as defined by Cohen's (1988) original formula (i.e., the difference between our sample mean and chance performance, divided by our sample standard deviation as a substitute for σ) because it deliberately ignores information about the study design and thus allows other researchers to meaningfully compare our effects with different studies. Effect sizes remained large across the image resolution conditions: 2×2 pixel images, $t(49) = 10.32$, $p < .001$, $d = 1.46$; 4×4 pixel images, $t(49) = 9.76$, $p < .001$, $d = 1.38$; 8×8 pixel images, $t(49) = 9.70$, $p < .001$, $d = 1.37$; 16×16 pixel images, $t(49) = 12.58$, $p < .001$, $d = 1.78$; 32×32 pixel images, $t(49) = 13.10$, $p < .001$, $d = 1.85$; 64×64 pixel images, $t(49) = 21.60$, $p < .001$, $d = 3.05$; and 128×128 pixel images, $t(49) = 29.76$, $p < .001$, $d = 4.21$. Equivalent nonparametric one-sample tests (e.g., Wilcoxon signed-ranks test) produced the same significant results ($p < .001$, applying the Holm correction) across image resolutions (see the supplemental R Markdown file for data analytic code and output: https://osf.io/thk2a).[1]

Response bias. We computed participants' average response bias (b) for each image resolution condition from their hits and false alarms, once again averaging over domains (see Figure 2B). A b score of 1 indicates no response bias, with scores greater than 1 indicating a tendency to say "old" more, and scores less than 1 indicating a tendency to say "new" more. Participants in our experiment tended to say "old" more than "new" ($b = 1.15$, $SD = .37$). This response bias was strongest for the 1 pixel images (see Figure 2B): 1 pixel ($b = 1.32$, $SD = .39$), 2×2 ($b = 1.13$, $SD = .35$), 4×4 ($b = 1.15$, $SD = .36$), 8×8 ($b = 1.20$, $SD = .35$), 16×16 ($b = 1.18$, $SD = .39$), 32×32 ($b = 1.12$, $SD = .53$), 64×64 ($b = 1.09$, $SD = .28$), and 128×128 pixels ($b = 1.01$, $SD = .22$). A one-way within-subjects ANOVA using participants' b scores revealed a significant main effect of image resolution on response bias, $F(7, 343) = 3.18$ $p = .003$, $\eta^2 G = .05$. Polynomial contrasts further revealed a significant linear ($p < .001$) trend to participants' response bias (b) scores over image resolutions. Response bias (b) decreased as the pixel levels went up (see Figure 2B).

We conducted further analyses comparing participants' mean b scores with each image resolution to that expected with no bias ($b = 1.00$). This analysis revealed that participants' response bias toward saying "old" was significantly greater than no bias for all image resolution conditions (correction applied to all p values). The largest effect was observed with the 1 pixel images, $t(49) = 5.72$, $p < .001$, $d = .81$. Smaller effect sizes were observed for the other seven image resolutions: 2×2 pixel images, $t(49) = 2.62$ $p = .012$, $d = .37$; 4×4 pixel images, $t(49) = 2.96$, $p = .005$, $d = .42$; 8×8 pixel images, $t(49) = 4.00$, $p < .001$, $d = .57$; 16×16 pixel images, $t(49) = 21.25$, $p < .001$, $d = .46$; and 64×64 pixel images, $t(49) = 2.25$, $p = .029$, $d = .32$. No significant response bias was observed for the 32×32 pixel images, $t(49) = 1.63$, $p = .110$, or 128×128 pixel images, $t(49) = .43$, $p = .66$.

Rate correct. Finally, we conducted exploratory analyses of participants' rate correct scores (RCSs; see Figure 2C). Unlike mean response times, RCSs provide an indication of participants' speed *relative* to their accuracy. Compared with other speed-accuracy measures, RCS is also easy to interpret and can be described here as the number of correct responses per second. Participants' rate correct remained stable over the lower resolutions, but increased from 32×32 and up (see Figure 2C): 1 pixel (RCS = .36, $SD = .09$), 2×2 (RCS = .34, $SD = .07$), 4×4 (RCS = .35, $SD = .07$), 8×8 (RCS = .35, $SD = .09$), 16×16 (RCS = .38, $SD = .13$), 32×32 (RCS = .45, $SD = .13$), 64×64 (RCS = .48, $SD = .09$), and 128×128 pixels (RCS = .52, $SD = .11$). A one-way within-subjects ANOVA based on participants' mean RCSs also revealed a significant main effect of image resolution, $F(7, 343) = 47.03$, $p < .001$, $\eta^2 G = .34$. As with discriminability, polynomial contrasts of the image resolution conditions showed a significant linear ($p < .001$), quadratic ($p < .001$), and quartic ($p = .011$) trend over image resolutions. Rate correct (RCS) increased as the pixel levels went up, and this linear trend was marked by four bends or two step changes at the higher resolutions (see Figure 2C).

Discussion

We found that people could reliably recognize images above chance (with large effect sizes) at every image resolution, including 1 pixel. As predicted, people's discriminability of old and new images tended to increase as image resolution increased. On average, participants were also more likely to say that they had seen an image before ("old") than not ("new"), and their rate correct (i.e., correct responses per second) also increased with higher resolutions.

Looking back in memory, the lower resolution images may appear "older" as a whole because there are fewer distinguishing features between the old and the new images in those conditions. But how are participants remembering 1×1, 2×2, 4×4 pixel images reliably? The old and new images were randomly and equally sampled from both categories in all three domains for each memory block. That is, we controlled for the categorical structure of the image sets by counterbalancing the categories across the old and the new images (e.g., in a given painting's memory block, half of the old paintings were randomly sampled from old cubist artists and half were randomly sampled from old impressionist artists, and likewise for the new paintings). Participants may have gleaned the semantic category on any particular image in the experiment, and they may have even noticed that the images tended to be of particular kinds over the course of the experiment. But it was impossible for participants to rely on what they knew about the categorical structure of the image *sets* to recognize the *images*, because the categories did not covary systematically with "old" and "new" images. In other words, as a participant in this experiment, knowing that half of the images in a given memory block contained "owls" and half contained "hawks" was not a diagnostic

---

[1] We report the results of parametric analyses throughout the main text for ease of comparison across studies; however, because of the observed skewness in our data (see the density plots in the supplemental R Markdown file: https://osf.io/thk2a), we repeated all follow-up one-sample t tests in both experiments with a Wilcoxon's signed-rank test, consistent with our reporting of nonparametric signal detection statistics, and use permutation tests as subsidiary analyses to the ANOVAs.

cue as to whether they had seen that image before—because these categories were uncorrelated with the categories "old" and "new."

What information contained *within* the particular images, then, are participants relying on to recognize them? At the higher resolutions, knowing that an image contains a snowy owl in flight, with flecks of black plumage, on a backdrop of clear blue sky, may help to recognize it as "old" or "new" by virtue of having multiple possible commonalities with the image represented in memory. At the lower resolutions (1 pixel, 2 × 2 pixels, etc.), however, it seems unlikely that participants *know* there is a feathered, carnivorous bird in the image, and it is therefore unlikely that they are relying on such semantically meaningful content to recognize the image. We propose, instead, that participants are relying on the global configural and low-dimensional properties distributed across the image at extremely low resolutions. That is, participants are probably recognizing the "blocky" patterns present at extremely low resolutions without knowing the semantic category and without recognizing any meaningful local content in the images. If this explanation holds, the low-resolution results provide evidence that we do not *need* to know what we are seeing in order to remember it (Wolfe & Kuzmova, 2011). The fact that people can remember images that have been downsampled to a single pixel, even 2 × 2 or 4 × 4 pixels, reliably above chance, suggests that global features distributed across the image can be sufficient for recognition.

Experiment 2: Discriminating Natural Images at Extremely Low Resolutions

In Experiment 2, we move from an image recognition task to a discrimination task to probe the sufficiency of global features distributed across images for discriminating *categories* rather than *images*. Here, we showed people two images on screen, side-by-side. Their task was to decide whether the two images were from the *same* category or two *different* categories. We tested how well 50 undergraduate psychology students could discriminate categories across the same three domains (paintings, birds, and faces) and eight image resolutions (128 × 128, 64 × 64, 32 × 32, 16 × 16, 8 × 8, 4 × 4, 2 × 2, and 1 pixel) introduced in Experiment 1. Again, we examined participants' discriminability (*A*), response bias, and rate correct, collapsing over domain. Using a metric of discriminability, in which .5 is chance and 1 is perfect performance, we based our preregistered predictions on the results obtained by Torralba (2009), whose task and stimuli were most similar to this experiment. We predicted a discriminability of .54 for the 1 pixel images, .60 for the 2 × 2 pixel images, and .65 for the 4 × 4 pixel images, with a further discrimination improvement as the number of pixels in the images increased.

Method

Participants. A separate, prespecified sample of 50 undergraduate psychology students (34 female, 16 male) from The University of Queensland participated for course credit. Participants' mean age was 19.86 years (*SD* = 3.51). A sensitivity analysis based on this sample suggested that the smallest effect size we can detect with 80% power, accounting for the 768 trials and within-subjects design, was *f* = .037 (η^2 = 0.001 or a "small" effect). Sensitivity for the one-sample analyses was identical to Experiment 1.

Procedure. The stimuli and procedure were identical to Experiment 1, with the exception of the task. Participants were asked to rate the extent to which they thought the two images on each trial were from the same category or from two different categories. They indicated their response on a 12-point, forced-choice rating scale, with buttons ranging from 1 (*sure different*) to 6 (*unsure different*) for the "different" responses and buttons ranging from 7 (*unsure same*) to 12 (*sure same*) for the "same" responses. After they pressed one of these 12 buttons, the two images disappeared and were replaced 500 ms later with two new images to be rated. The two images remained on screen until a response was made, but a "speed-up" prompt was delivered if they took longer than 4 s to respond.

Each participant continued this process for 768 trials, consisting of 256 pairs of bird images, 256 pairs of paintings, and 256 face pairs, presented as three consecutive blocks in a different random order for each participant. Each 256-trial domain block included eight smaller blocks of 32 image pairs at each of the eight image resolutions. Each 32-trial resolution block was also presented in a different random order within each domain and to each participant. The 32 trial blocks were further subdivided into half matching and half mismatching image pairs. For example, the correct response on a trial displaying two different impressionist paintings would be "same." But the correct response on a trial displaying an impressionist and cubist painting would be "different." Matching and mismatching trials are presented in a different random order within each block of 32, and the images were sampled equally from both subcategories in each domain across all conditions (e.g., cubist and impressionist, hawk and owl, male and female). As in Experiment 1, participants were not provided with the category labels for any of the image sets at any point in order to minimize reliance on higher level semantic knowledge. We opted for the label-free visual discrimination task, as we were interested in capturing participants' perceptual sensitivity, in its purest form, to the similarities and differences in the image sets at reduced resolutions. For clarity of instruction, participants were asked to judge whether the two images were from the same category or two different categories, but they could theoretically perform the task without knowing the names of the categories they were discriminating.

Results

As in Experiment 1, we report analyses of the effect of image resolution on participants' aggregated nonparametric discriminability (*A*), the associated nonparametric measure of response bias (*b*), and RCSs (number of correct response per second). Again, exploratory analyses of participants' AUC data are not reported, as they revealed the same pattern of results as our primary planned analyses of nonparametric discriminability (*A*).

Discriminability. As illustrated in Figure 3A (see also Figure 3D), participants' discriminability tended to increase with increasing image resolution in a similar fashion to the predicted pattern: 1 pixel (*A* = .53, *SD* = .09), 2 × 2 (*A* = .56, *SD* = .07), 4 × 4 (*A* = .56, *SD* = .08), 8 × 8 (*A* = .57, *SD* = .07), 16 × 16 (*A* = .62, *SD* = .07), 32 × 32 (*A* = .67, *SD* = .08), 64 × 64 (*A* = .73, *SD* = .10), and 128 × 128 pixels (*A* = .73, *SD* = .10). A one-way within-subjects ANOVA using participants' *A* scores also revealed a significant main effect of image resolution, $F(7, 343) = 49.84$, $p < .001$, $\eta^2_G = .44$. Polynomial contrasts of the image resolution

conditions revealed a significant linear ($p < .001$) and quadratic ($p < .001$) increase in discriminability (A) with increases in the number of pixels. Discriminability (A) increased as the pixel levels went up, and this trend is further characterized by larger increases at higher resolutions (see Figure 3A).

Further one-sample analyses (applying the Holm correction to all p values) revealed that participants' discriminability was reliably greater than chance for all image resolution conditions, except the 1 pixel images, $t(49) = 1.85$, $p = .070$, $d = .26$. Effect sizes were large and tended to increase across the remaining seven image resolutions: 2×2 pixel images, $t(49) = 5.98$, $p < .001$, $d = .85$; 4×4 pixel images, $t(49) = 5.60$, $p < .001$, $d = .79$; 8×8 pixel images, $t(49) = 6.85$, $p < .001$, $d = .97$; 16×16 pixel images, $t(49) = 12.49$, $p < .001$, $d = 1.77$; 32×32 pixel images, $t(49) = 14.95$, $p < .001$, $d = 2.12$; 64×64 pixel images, $t(49) = 15.83$, $p < .001$, $d = 2.24$; and 128×128 pixel images, $t(49) = 15.60$, $p < .001$, $d = 2.21$. Discriminability (A) was significantly greater than chance ($p < .001$) for all image resolutions, including the 1 pixel images ($p = .033$) using the Wilcoxon signed-ranks test and Holm correction (see the supplemental R Markdown file for all data analysis code and output: https://osf.io/thk2a).

Response bias. In Experiment 2, participants tended to say "different" more than "same" ($M = 1.28$, $SD = .38$). This response bias also increased as image resolution *increased* (opposite from what we predicted): 1 pixel ($M_b = 1.15$, $SD = .29$), 2×2 ($M_b = 1.15$, $SD = .32$), 4×4 ($M_b = 1.16$, $SD = .31$), 8×8 ($M_b = 1.27$, $SD = .36$), 16×16 ($M_b = 1.28$, $SD = .42$), 32×32 ($M_b = 1.30$, $SD = .37$), 64×64 ($M_b = 1.42$, $SD = .36$), and 128×128 pixels ($M_b = 1.48$, $SD = .44$). A one-way within-subjects ANOVA using participants' aggregated b scores revealed a significant main effect of image resolution on their response bias, $F(7, 343) = 6.66$, $p < .001$, $\eta^2 G = .09$. Similar to participants' discriminability, polynomial contrasts showed a significant linear ($p < .001$) and quadratic ($p = .040$) increase in participants' response bias (b) with increasing pixel levels.

Further analyses revealed that participants' response bias toward saying "different" was significantly greater than no bias for all image resolution conditions, including the 1 pixel images, $t(49) = 3.65$, $p < .001$, $d = .52$. Effect sizes tended to increase across the eight image resolutions: 2×2 pixel images, $t(49) = 3.39$, $p = .001$, $d = .48$; 4×4 pixel images, $t(49) = 3.67$, $p < .001$, $d = .52$; 8×8 pixel images, $t(49) = 5.19$, $p < .001$, $d = .73$; 16×16 pixel images, $t(49) = 4.72$, $p < .001$, $d = .67$; 32×32 pixel images, $t(49) = 5.80$, $p < .001$, $d = .82$; 64×64 pixel images, $t(49) = 8.19$, $p < .001$, $d = 1.16$; and 128×128 pixel images, $t(49) = 7.64$, $p < .001$, $d = 1.08$.

Rate correct. Participants' mean rate correct remained fairly stable across image resolutions, unlike Experiment 1 (see Figure 3C): 1 pixel (RCS = .42, $SD = .24$), 2×2 (RCS = .40, $SD = .18$), 4×4 (RCS = .40, $SD = .18$), 8×8 (RCS = .40, $SD = .18$), 16×16 (RCS = 39, $SD = .16$), 32×32 (RCS = 38, $SD = .13$), 64×64 (RCS = .40, $SD = .16$), and 128×128 pixels (RCS = .39, $SD = .12$). A one-way within-subjects ANOVA using participants' mean RCSs showed no significant main effect of image resolution, $F(7, 343) = .99$, $p = .442$, $\eta^2 G = .01$, and no significant linear ($p = .473$) or quadratic ($p = .712$) trend in rate correct with quadratic increases in pixel level.

Discussion

We found that people could reliably discriminate image categories above chance at image resolutions down to 2×2 pixels and less reliably with 1 pixel. As predicted, people's discriminability tended to increase as image resolution increased. Participants' response bias also tended to increase as image resolution increased, and they tended to say "different" more often than "same" at all resolutions. There was no evidence for a difference in rate correct across image resolution conditions.

The benefit of randomizing the resolution blocks was in negating the possibility that any observed increases in discriminability with increasing image resolution was a result of participants learning from category exemplars presented earlier in the experiment. Had we increased the resolution blocks incrementally over the experiment, for instance, the observed increased performance would be perfectly confounded with the number of prior category exemplars participants had seen in the experiment. The drawback of randomizing the resolution blocks is that some participants viewed some higher resolution blocks before the lower resolution blocks, purely based on a random generation process. And participants' exposure to high-resolution category exemplars early in the experiment may have inflated their discriminability of the lower resolutions.

If learning in the higher resolution blocks was leaking over into the lower resolution blocks, we would expect to see correlated discriminability across the different blocks. To test this possibility, we correlated participants' discriminability (A) scores in each resolution block and with other resolution block. Participants' discriminability was highly correlated among the three highest resolution blocks: 32×32 and 64×64 ($r = .638$, $p < .001$), 32×32 and 128×128 ($r = .515$, $p < .001$), and 64×64 and 128×128 ($r = .588$, $p < .001$). We also found a moderate correlation between the 2×2 pixel block and the 16×16 block ($r = .364$, $p = .009$). But discriminability in the 1×1 pixel block was not correlated at all with discriminability in any other block (relating to the 2×2 pixel block and up: $r = -.092$, $r = -.084$, $r = .051$, $r = .013$, $r = -.023$, $r = -.060$, and $r = -.020$, respectively). Discriminability in the 2×2 pixel block was also not correlated with any other block, except 16×16 pixels (correlations with the 4×4 pixel block and up: $r = .094$, $r = .096$, $r = .364$, $r = -.049$, $r = -.042$, and $r = .143$, respectively). Nor was discriminability in the 4×4 pixel block correlated with any other (correlations with the 8×8 pixel block and up: $r = .159$, $r = .131$, $r = .051$, $r = -.010$, and $r = .115$, respectively). And discriminability in the 8×8 pixel block (correlations with the 16×16 pixel block and up: $r = .113$, $r = .087$, $r = .126$, and $r = .252$, respectively) was not correlated with the 16×16 pixel block (correlations with the 32×32 pixel block and up: $r = -.120$, $r = .018$, $r = .126$, and $r = .081$, respectively) or any of the higher resolutions.

These correlational data show that discriminability at the lower resolutions is not predicted by discriminability at the other resolutions. Apart from the three highest resolutions, discriminability in each block is independent of the other blocks. That is, any familiarity gained with the higher resolution exemplars does not predict performance at the lower resolutions. Prior knowledge of the category exemplars is, therefore, not an adequate explanation for why participants can discriminate categories at extremely low image resolutions. Participants' RCSs also did not vary signifi-

cantly across image resolutions in the discrimination task, indicating that the effect of image resolution on discriminability cannot be accounted for by participants emphasizing accuracy over speed to a greater extent given more information. One plausible explanation for the lack of an effect on rate correct is that the increasing discriminability with increasing image resolution was offset by increasing complexity in the images—they were slower to respond in the higher image resolution conditions even though they were more accurate.

With image resolutions as low as 2 × 2 or 4 × 4 pixels, the once-nameable local features within the images were no longer present and thus cannot be used as the basis for discrimination. All that remained was a few patches of color that represent the most redundant information within and across the images in each of the categories. The hawks and owls, for example, appeared as different shades of green and brown, and the male and female faces as shades of beige. Even as the image resolution increased, the local features—a fine-brushed sailboat, a leafy branch, a flourishing pair of eyebrows—remained obscure.

These findings suggest that people are sensitive to visual information that is distributed across images as well as within images even when local information is sparse. Although semantic coding at extremely low image resolutions seems unlikely, we speculate that the increase in discriminability at 16 × 16 pixels, as indicated by the significant quadratic trend, may mark a point at which participants start to take advantage of semantic coding in their decision process. Similarly, the change in participants' tendency to say "different" more with the higher image resolutions might be explained by an increase in the identification of distinguishing objects and features that are local to the images. These post hoc observations could be tested in future experiments by gauging the point at which participants tend to recognize category labels and objects in an image with gradually increasing image resolution. An interesting prediction based on this interpretation of our results is that semantic coding may influence the decision process at lower and lower image resolutions with the accumulation of experience in a domain as participants develop visual expertise with natural categories (e.g., faces, birds, medical images, fingerprints; Mackenzie, Dodd, Tomarken, & Gauthier, 2018).

General Discussion

In two experiments, we illustrated that the global *stylistic* information retained at extremely low image resolutions was sufficient for recognizing images and discriminating categories with above-chance accuracy across three domains—paintings, birds, and faces. We averaged participants' performance over the three large-scale natural image sets and used a random image sampling procedure for each participant and block of trials. We therefore expect our findings to generalize well to other natural image sets that have different categorical structures. We distinguish between the style of an image (e.g., a single Monet painting) and the style of a category (e.g., a set of impressionist paintings) here because their relative importance appears to depend on the visual task at hand.

Experiment 1, for instance, points to the importance of stylistic information distributed across an image for visual recognition. People were able to reliably recognize natural images reduced down to 4 × 4 pixels, 2 × 2 pixels, and even a single pixel without any meaningful local objects in the image to rely on. The old and

new images were randomly sampled from each of the broader categories in the set, such that their global structure could not be used to aid image recognition. At extremely low resolutions, the images also contained little to no conceptually meaningful local features reminiscent of the original, full-sized images. With a single colored pixel, 2 × 2 pixels, or 4 × 4 pixels, it seems unlikely that participants are encoding any semantic properties of the original image in order to correctly recognize it. Thus, although an increased ability to perceive meaningful objects and categories in images may help people to remember them (Wolfe & Kuzmova, 2011), our findings suggest that one need not know to remember.

Experiment 2, on the other hand, points to the importance of stylistic information distributed across images for discriminating between different visual categories. Participants were able to reliably discriminate images with extremely low resolutions, such as discriminating between a painting by Monet and a painting by Picasso when given only four pixels to go on. Again, at these extremely low resolutions, there were no original local features in the image for people to rely on in order to make a discrimination. As with Herrnstein and Loveland's (1964) pigeons, people appear to show sensitivity to the covariant structure that is distributed across members of a category in the absence of knowing about their semantic content.

Our notion of style in the image recognition memory task refers to the distributed visual information *within* the image. No single feature or visual cue (e.g., hue, luminance, orientation) embodies the image's style, but when you look across several features or cues in the image, its style begins to take shape. Extending this description to the discrimination task in Experiment 2, the style of the categories being distinguished cannot be defined by any particular feature or visual cue, or even a set of features *within* a particular image. Instead, performance on the task depends on a series of features and visual cues that covary *across* images. An image or category's style is fuzzy, ill-defined, and not necessarily tied to any semantic description of its features. Indeed, the style of an image or visual category is analogous to Wittgenstein's (1953/2001) description of family resemblance when describing the covariation of features among the varied instances of "games"—board games and ball games have some commonalities, but also many differences, and it is only when you look across several instances that a resemblance emerges. How people perform on similar tasks, with the within-image and across-image global properties of the stimuli pitted against one another, could be tested in future studies and might help elicit a model of how different kinds of global properties are used in visual cognition.

Previous work has shown that categorization and recognition of scenes decreases as image resolution decreases, remaining above chance at 8 × 8 pixels (Torralba, 2009; Wolfe & Kuzmova, 2011). These findings are interpreted to suggest that the representation of a scene used for recognition memory may be similar to the representation used for categorization, and that images are better remembered when they are meaningful (Wolfe, Horowitz, & Michod, 2007). These findings, and ours, are consistent with computational vision accounts that propose a coarse-to-fine process by which ensemble or gist-like "blobs" of information are coded first and then used to guide subsequent analyses and recognition of finer details (Oliva & Torralba, 2006; Schyns & Oliva, 1994; Torralba & Oliva, 2003). Unlike gist, however, which is thought to embody a scene's semantic description, spatial layout,

and a few select objects that compose the scene (Torralba, 2009), our data provide evidence that explicit encoding of the semantic category or any local features in an image is not necessary for recognizing natural images or distinguishing among natural categories. Visual recognition and discrimination under extremely low resolutions appear to be more about style.

References

Aust, U., & Huber, L. (2001). The role of item- and category-specific information in the discrimination of people versus non-people images by pigeons. *Animal Learning & Behavior, 29,* 107–119. http://dx.doi.org/10.3758/BF03192820

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142,* 1323–1334. http://dx.doi.org/10.1037/a0033872

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37,* 379–384. http://dx.doi.org/10.3758/BF03192707

Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 1160–1176. http://dx.doi.org/10.1037/xhp0000399

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Routledge.

Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L., & Wolfe, J. M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin & Review, 20,* 1170–1175. http://dx.doi.org/10.3758/s13423-013-0459-3

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 10292–10297. http://dx.doi.org/10.1073/pnas.1606187113

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20,* 464–72. http://dx.doi.org/10.1111/j.1467-9280.2009.02316.x

Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science, 180,* 1194–1197. http://dx.doi.org/10.1126/science.180.4091.1194

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science, 146,* 549–551. http://dx.doi.org/10.1126/science.146.3643.549

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6,* 65–70.

Levenson, R. M., Krupinski, E. A., Navarro, V. M., & Wasserman, E. A. (2015). Pigeons (*Columba livia*) as trainable observers of pathology and radiology breast cancer images. *PLoS ONE, 10,* e0141357. http://dx.doi.org/10.1371/journal.pone.0141357

Mackenzie, A. S., Dodd, M. D., Tomarken, A. J., Gauthier, I. (2018). How faces (and cars) may become special. *Vision Research.* Advance online publication. http://dx.doi.org/10.1016/j.visres.2017.12.007

Morrone, M. C., Burr, D. C., & Ross, J. (1983). Added noise restores recognizability of coarse quantized images. *Nature, 305,* 226–228. http://dx.doi.org/10.1038/305226a0

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175. http://dx.doi.org/10.1023/A:1011139631724

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36. http://dx.doi.org/10.1016/S0079-6123(06)55002-2

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research, 49,* 2856–2861. http://dx.doi.org/10.1016/j.visres.2009.08.025

Scarf, D., Boy, K., Uber Reinert, A., Devine, J., Güntürkün, O., & Colombo, M. (2016). Orthographic processing in pigeons (*Columba livia*). *Proceedings of the National Academy of Sciences of the United States of America, 113,* 11272–11276. http://dx.doi.org/10.1073/pnas.1607870113

Scarf, D., Corballis, M. C., Güntürkün, O., & Colombo, M. (2017). Do "literate" pigeons (*Columba livia*) show mirror-word generalization? *Animal Cognition, 20,* 999–1002. http://dx.doi.org/10.1007/s10071-017-1116-4

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial scale-dependent scene recognition. *Psychological Science, 5,* 195–200. http://dx.doi.org/10.1111/j.1467-9280.1994.tb00500.x

Searston, R. A., & Tangen, J. M. (2017a). The emergence of perceptual expertise with fingerprints over time. *Journal of Applied Research in Memory & Cognition, 6,* 442–451. http://dx.doi.org/10.1016/j.jarmac.2017.08.006

Searston, R. A., & Tangen, J. M. (2017b). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLoS ONE, 12,* e0178403. http://dx.doi.org/10.1371/journal.pone.0178403

Searston, R. A., & Tangen, J. M. (2017c). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin & Review, 24,* 1324–1329. http://dx.doi.org/10.3758/s13423-016-1211-6

Sheehan, M. J., & Tibbetts, E. A. (2011). Specialized face learning is associated with individual recognition in paper wasps. *Science, 334,* 1272–1275. http://dx.doi.org/10.1126/science.1211334

Siebeck, U. E., Parker, A. N., Sprenger, D., Mäthger, L. M., & Wallis, G. (2010). A species of reef fish that uses ultraviolet patterns for covert face recognition. *Current Biology, 20,* 407–410. http://dx.doi.org/10.1016/j.cub.2009.12.047

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: Experts can do a lot with a little. *PLoS ONE, 9,* e114759. http://dx.doi.org/10.1371/journal.pone.0114759

Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience, 26,* 123–131. http://dx.doi.org/10.1017/S0952523808080930

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14,* 391–412. http://dx.doi.org/10.1088/0954-898X_14_3_302

Vandierendonck, A. (2018). Further tests of the utility of integrated speed-accuracy measures in task switching. *Journal of Cognition, 1,* 8. http://dx.doi.org/10.5334/joc.6

Vokey, J. R., Jamieson, R. K., Tangen, J. M., Searston, R. A., & Allen, S. W. (2018). A visual familiarity account of evidence for orthographic processing in pigeons (*Columbia livia*): A reply to Scarf, Corballis, Güntürkün, and Colombo (2017). *Animal Cognition, 21,* 425–431. http://dx.doi.org/10.1007/s10071-018-1166-2

Watanabe, S. (2001). Van Gogh, Chagall and pigeons: Picture discrimination in pigeons and humans. *Animal Cognition, 4,* 147–151. http://dx.doi.org/10.1007/s100710100112

Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior, 63,* 165–174. http://dx.doi.org/10.1901/jeab.1995.63-165

Wehner, R., Boyer, M., Loertscher, F., Sommer, S., & Menzi, U. (2006). Ant navigation: One-way routes rather than maps. *Current Biology, 16,* 75–79. http://dx.doi.org/10.1016/j.cub.2005.11.035

Wittgenstein, L. (2006). *Philosophical investigations.* Oxford, UK: Blackwell. (Original work published 1953)

Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research, 47,* 955–964. http://dx.doi.org/10.1016/j.visres.2006.11.025

SEARSTON, THOMPSON, VOKEY, FRENCH, AND TANGEN

Wolfe, J. M., & Kuzmova, Y. I. (2011). How many pixels make a memory? Picture memory for small pictures. *Psychonomic Bulletin & Review, 18,* 469–475. http://dx.doi.org/10.3758/s13423-011-0075-z

Wu, W., Moreno, A. M., Tangen, J. M., & Reinhard, J. (2013). Honeybees can discriminate between Monet and Picasso paintings. *Journal of Comparative Physiology A, Neuroethology, Sensory, Neural, and Behavioral Physiology, 199,* 45–55. http://dx.doi.org/10.1007/s00359-012-0767-5

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70,* 203–212. http://dx.doi.org/10.1007/s11336-003-1119-8

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.

- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.

- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.

- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/journals/resources/review-manuscript-ce-video.aspx.