When a never-seen but less-occluded image is better recognized: Evidence from old–new memory experiments

Hongjing Lu

Department of Psychology, University of California, Los Angeles, CA, USA

Zili Liu

Department of Psychology, University of California, Los Angeles, CA, USA

In studies of visual memory, an image identical to that previously seen is invariably recognized more accurately than any image that is different. This identity superiority provides the empirical foundation for the image-based theory of object and scene recognition. Here we present evidence to the contrary. In an "old–new" recognition task, a face less-occluded than its "old" counterpart was recognized more accurately than the "old" face itself. The same effect was also found with natural scenes. The superiority of a less-occluded image weakened when occlusion was further removed, indicating the visual system's ability to recover from occlusion is limited. When the images were inverted, the effect disappeared. Our findings support the classic notion that the visual system performs active abstraction and organization on sensory information in order to encode it into a memory representation.

Introduction

A fundamental problem in visual cognition is the nature of internal memory representations of objects and scenes (for a review, see Palmeri and Gauthier, 2004). The representation can be partially revealed by the characteristics of behavioral performance in an object or scene recognition task. Behavioral performance in turn helps revise or refute a theory when reliable counterexamples are found. This is the approach taken by the current study.

In object recognition, the image-based theory predicts that the further a viewpoint is rotated away from the viewpoint from which the object was originally seen, the less accurate recognition performance will be (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Tarr & Bülthoff, 1998; Tarr, Williams, Hayward, & Gauthier, 1998). More generally, according to Poggio and Edelman (1990), the theory predicts that the less similar an image is to that previously seen, the worse performance will be. Moreover, the competing theory proposed by Biederman and colleagues also predicts that an image different from that previously seen (due to object rotation, translation, reflection, or size change) gives rise to at best equal, but never better, performance than an image identical to that previously seen (Biederman, 2000; Biederman & Cooper, 1991, 1992).

Here we present evidence contrary to both of these leading theories of object recognition. Similarity is a thorny concept (Goldstone & Son, 2005), but for present purposes we simply made the minimal assumption that two identical images are more similar to each other than are two different images. We sought to identify situations in which two distinct images give rise to better recognition than two identical images. In order to achieve that, we took advantage of the active nature of the visual system in perceptual organization and abstraction, which may be particularly important when input information is impoverished, e.g., with partial occlusion (He & Nakayama, 1992; Kanizsa, 1979; Kellman & Shipley, 1991; Nakayama, Shimojo, & Silverman, 1989). The visual system's ability to recover from occlusion is called amodal completion.

Amodal completion was investigated by Sekuler and Palmer (1992), who found in a same–different matching task that a pie chart inducer primed a disk as well as the disk itself (but not better). These investigators argued that the inducer was perceptually completed into a disk, which in turn primed the subsequent disk stimulus. However, the trials in the experiment were repeated, so an observer had seen multiple trials with the inducer presented as the prime and followed by the disks. As a result, an association may have been formed between the inducer and the disks, confounding the role of perceptual completion.

Perceptual abstraction was also investigated in the classic study of Posner and Keele (1970). Participants were trained to classify random-dot patterns into three

categories, each of which was created by randomly perturbing the dot positions from a predetermined pattern, termed a prototype. During training, the prototypes were never shown. Participants were tested either immediately or one week later. It was found that proportion of classification errors for the trained exemplars increased in one week's time (0.14 to 0.39), whereas error rates for the prototypes changed little (0.35 to 0.38). Posner and Keele suggested that the representations of the categories were not simply the trained exemplars; rather, the average of the trained exemplars, or the prototypes, seemed to also be represented, and in a more stable manner. It should be noted that in this study, the never-seen prototypes were not better categorized than the trained exemplars.

In the current study, we were seeking to identify conditions in which a never-seen image gives rise to better categorization than an image that has been seen. A result close to what we were seeking is the boundary extension effect in scene perception (Intraub & Richardson, 1989; see also Balas & Sinha, 2007). To illustrate with an example, after seeing a close-up photo of a bowl of spaghetti, a participant reproduced from memory the scene by drawing in a way as if the mind's eye had "zoomed out," such that additional space and content beyond the original boundary were drawn. More recently, it was also found that after a close-up view of a scene was shown, a zoomed-out view was rated as more similar to the original view than was the view identical to the original, which was rated as being too close (Park, Intraub, Yi, Widders, & Chun, 2007). In terms of signal detection theory (Green & Swets, 1974), however, whether this effect was due to sensitivity or bias remains unsolved.

In the current study, we specifically investigated whether the putative effect we were seeking was due to discrimination sensitivity or response bias. To anticipate, we found that for faces and natural scenes, discrimination sensitivity was greater when a test image was less occluded than when it was identical to the image previously seen. However, further reduction of occlusion worsened recognition performance, implying limited capability of the visual system to amodally complete patterns. No face or scene was retested, thereby ensuring that the results could not be due to prior exposure of the less-occluded images.

In a companion study (Lu & Liu, 2008), two partially occluded faces were sequentially presented in a same–different matching task. The grayscale face images were occluded by randomly positioned red pixels. It was found that when occlusion was reduced from 60% in the first image to 50% in the second image, both the hit rate and sensitivity d' were greater than when the two images were identical and 60% occluded. Further reduction of occlusion worsened the performance. The superiority of less-occluded images also disappeared when the faces were inverted. These results are consistent with those found in the current study.

Experiments

Experiment 1: Faces

Stimuli

Grayscale face images (257×257 pixels apiece) were obtained from the Max Planck Institute for Biological Cybernetics, Tübingen. No spectacles, hair, or ears were visible in any image. The inner portion of a face was visible through an oval aperture of 160×130 pixels, subtending $3.5° \times 4.3°$ in visual angle from a viewing distance of 57 cm.

Occlusion was created by 328 randomly positioned and non-overlapping red rectangles that covered 35%, 30%, or 25% of the image area. All rectangles were approximately equal in area. Their shapes were randomized by varying the width from 5 to 15 pixels. Due to pixel quantization, additional fine adjustment was made to precisely fix the total occluded area. If occlusion was 35%, an "old" image remained exactly the same from the first (study) to second (test) phase. If occlusion was 30% or 25%, the size of each rectangle was reduced without changing its center position. Occlusion pattern from one image to the next, and from one participant to the next was randomized.

Procedure

Recognition memory was tested using a six-scale old–new rating paradigm. Participants first saw partially occluded faces and then rated how likely each face had been seen before. The experiment consisted of three blocks. In each block, there were two phases. In the first (study) phase of 52 trials, participants rated the attractiveness of a face on a six-point scale (Figure 1A). In each trial, a face image was shown for 6 sec. It was then replaced by a rating scale for which the participant responded using the mouse. One second after the response, the next trial started automatically. There were 13 faces in the study phase, each presented four times. Nine of the 13 faces were occluded by 35% in area and were tested in the second phase. Two were occluded by 30%, and the remaining two by 25%. These four faces served as fillers in order to show participants the range of occlusion, but were not tested in the second phase. The sequence of the trials was randomized.

In the test phase of 18 trials, a partially occluded face was shown together with a six-point scale below it. The participant rated how likely the face (but not necessarily the image) had been seen in the study phase (Figure 1B). The nine "old" faces plus nine "new" faces were presented in a random order without repetition. Three of the nine "old" faces were occluded by 35% (identical to those in the study phase), three by 30%, and three by 25% (Figure 1C). The nine "new" faces were similarly occluded in image area, but the distribution of red rectangles was random. The 22 faces in one block (13 + 9) were from 22 people, so that no face was retested. Each
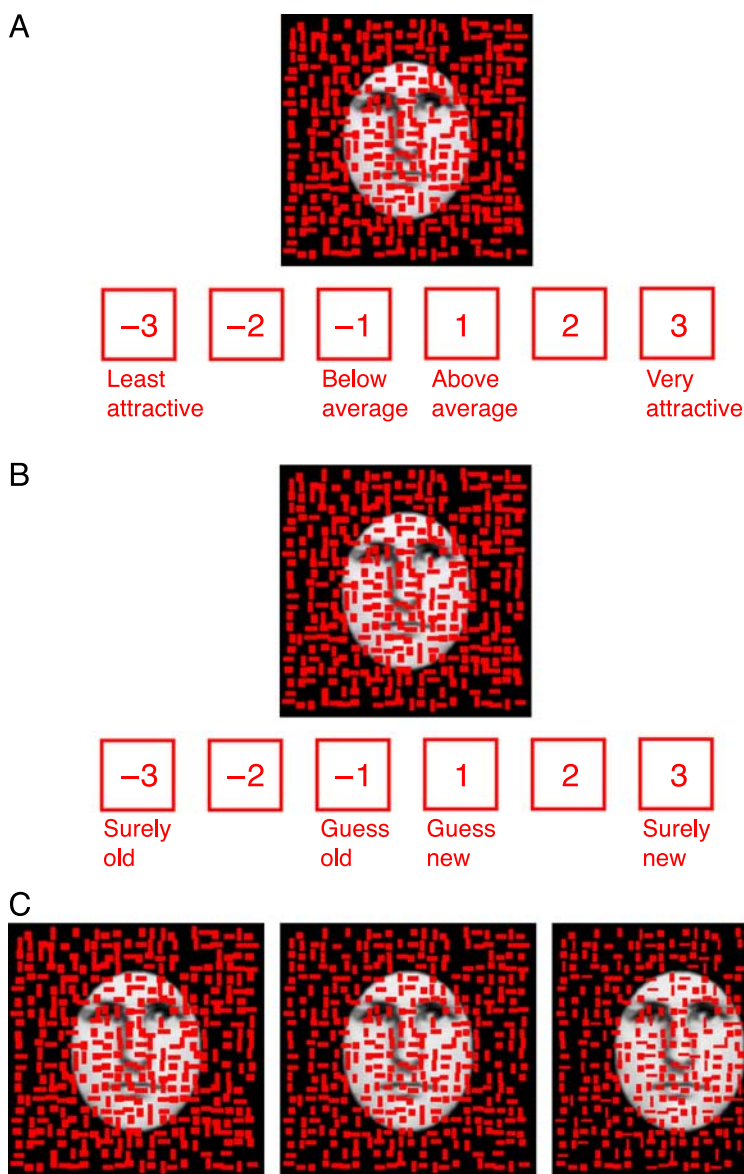
Figure 1. Example face stimuli in Experiment 1. (A) An example stimulus used for attractiveness rating. (B) An example stimulus used for old–new rating. (C) An example of the face in A and B being occluded by, from left to right, 35%, 30%, or 25%. Reduction of occlusion was achieved by shrinking each red rectangle while keeping its center location unchanged. In panels A and B, the entire background was also black.

block used a separate set of 22 faces, yielding a total of 66 faces in the experiment. The selection of the 22 faces in each block and the old–new assignment of faces were randomized across participants. For half of the participants, the direction of the old–new scale from −3 to +3 indicated increasing certainty of "new"; for the remaining participants, the opposite was true. It took about 40 min for a participant to complete the experiment. The experiment was conducted in dark rooms.

Participants

One hundred and one University of California Los Angeles (UCLA) undergraduate students participated for course credit. Another 37 students participated in the control experiment when all faces were inverted. They were instructed that their old–new judgment should be solely based on the faces, but not necessarily on the images; and that the odds of an "old" face were 50–50.

Results

Our main hypothesis concerned whether a less occluded "old" face could be better recognized than the same face image identical to the previously seen. Accordingly, the most direct test of this hypothesis is to compare the hit rates at the original occlusion and each of the two less occluded levels. In order to avoid the confound of any bias

difference, we computed bias-free hit rates by choosing on each receiver operating characteristic (ROC) in Gaussian space (which was a straight line with unitary slope) the point where Z(hit rate) $= -Z$(false-alarm rate). (We also verified that the slope of the ROC did not violate the unitary assumption, $p = 0.08$.) It turned out that the bias-free hit rate at 30% occlusion was significantly higher than at 35% (72.5% vs. 68.5%, $t(100) = 3.14$, $p = 0.002$). The hit rate at 25% occlusion (70.0%), however, was comparable but not significantly higher than at 35% ($t(100) = 1.22$, $p = 0.23$), as shown in Figure 2, left. This implies that the representation of a 35% occluded face effectively reduced the occlusion to approximately 30%, but little further.

Because the assignment of "old" and "new" were completely randomized across participants, a higher hit rate is expected to lead to higher discrimination sensitivity, d'. To directly verify this prediction, d' was calculated for each occlusion level per participant. The d' value at occlusion 30% was indeed significantly higher than at occlusion 35% (1.25 vs. 1.03, $t(100) = 2.82$, $p = 0.006$). The sensitivity at occlusion 25% (1.12) was slightly greater than, but not significantly different from, that at 35% ($t(100) = 1.20$, $p = 0.23$). These results were further confirmed using a different sensitivity measure, A_z, which is the area under the ROC in standard space. Specifically, occlusion at 30% gave rise to higher A_z than at 35% ($t(100) = 2.21$, $p = 0.03$); occlusion at 25% was marginally higher than at 35% ($t(100) = 1.74$, $p = 0.08$).

Taken together, these results show that a never-seen but less-occluded face image was better recognized than the image identical to that previously seen. This result was not critically dependent on the way sensitivity was measured, and also generally held under Bonferroni multiple comparison correction. However, when occlusion was reduced further, recognition ceased to improve, indicating limited capability of the visual system's amodal completion.

In comparison, when all faces were inverted, reduced occlusion ceased to make a difference (Figure 2, right). The hit rates for the three occlusion levels were 0.60, 0.59, and 0.59, respectively. No significant difference could be found in any pairwise comparison. The lack of difference for inverted faces suggests that this superior recognition of
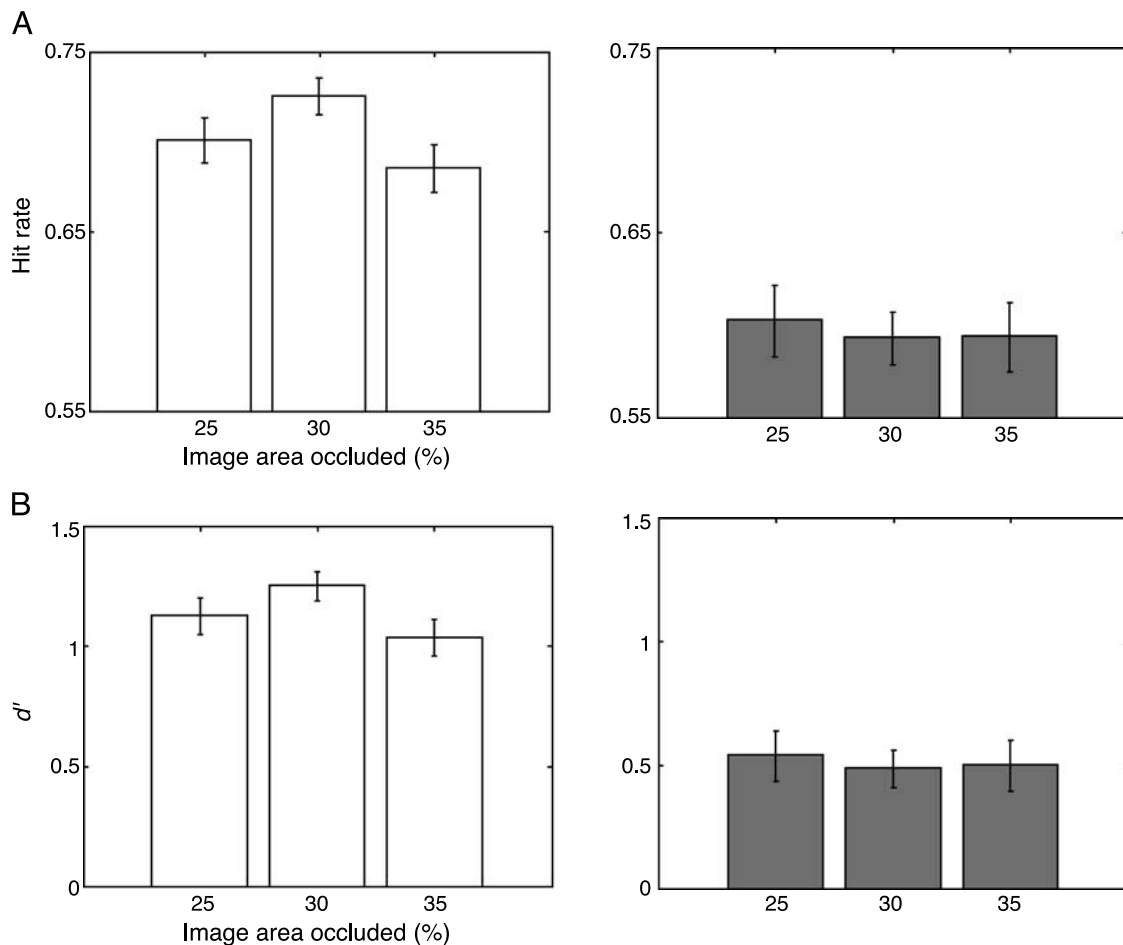

Figure 2. Results of Experiment 1 with upright (left column) and inverted faces (right column). *Top row*: bias-free hit rates. *Bottom row*: discrimination sensitivity d'. Error bars represent *SEM*.

less-occluded face images may depend on the high familiarity of upright faces such that when faces are inverted, recovery of occluded regions becomes much weakened.

Experiment 2: Natural scenes

This experiment was similar to Experiment 1 except that instead of faces, each of the three blocks used images from one of three categories (categorized by one author): buildings, animals or objects, and other natural scenes (from the Corel Stock Photo Library) (Figure 3, top). Each block consisted of 34 images (384 × 256 pixels apiece, subtending 10.4° × 7° in visual angle). In the study phase, occlusion of 30 out of the 34 images was 50%. There were 488 red occluding rectangles per image. This number was chosen so that the average area of a rectangle at 50% occlusion was 50/35 of that of a rectangle at 35% occlusion in the face experiment. There were also two images that were 40% occluded and two 30% occluded. These were fillers that were not tested in the test phase. Each image was presented only once in the study phase. In the test phase, the occlusion was at 50% (images identical to those in the study phase), 40%, or 30% (Figure 3, bottom). The images were paired according to similarity, such that one was assigned as "old" and the other "new." Figure 4 shows two example pairs. Due to the authors' error, the assignment of "old" and "new" images was randomized once for each pair, and then was consistent across participants. The assignment of pairs to occlusion level was randomized for each participant.

Twenty-seven UCLA undergraduate students participated for course credit. Fifty students at the Chinese University of Science and Technology, Hefei, China, also participated. A control experiment was also run with all images shown upside-down while everything else remained constant, with 138 UCLA undergraduate students participating for course credits. In all conditions, the sequence of the three blocks was randomized across participants.
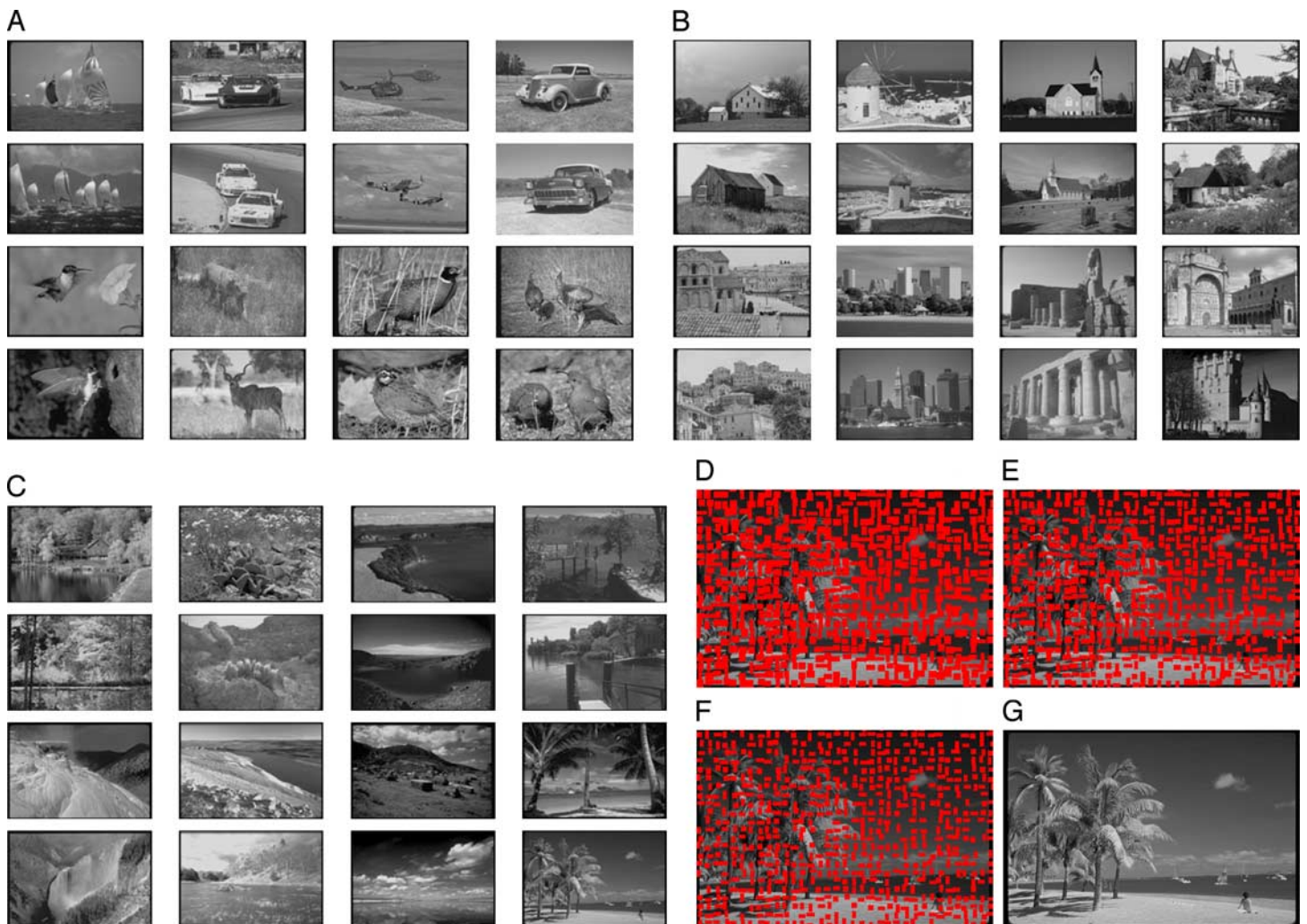

Figure 3. Example stimuli in the scene study (Experiment 2). (A) Objects (top two rows) and animals (bottom two rows). (B) Buildings. (C) Other natural scenes. (D–G) An example scene from category C at occlusion level 50%, 40%, 30%, and 0%, respectively.

Figure 4. Two pairs of example images. One image from a pair was assigned as "old" and the other as "new."

Results

Upright images

Because the assignment of "old" and "new" images was by mistake randomized only once and then held constant across participants, a sensitivity analysis might be confounded by any differences between the particular images assigned as "old" versus "new." Accordingly, we analyzed the unbiased hit rates to test our hypothesis directly. In all analyses reported below, the three image categories gave rise to no significant main effect or interactions. Consequently, data were collapsed across these three categories. The lack of differences among categories implies that the subjective classification of the images into categories did not influence any of the results concerning occlusion levels.

Each participant's unbiased hit rate was computed in the same manner as in Experiment 1. (We also verified that the slope of the ROC did not violate the unitary assumption, $p = 0.06$). As shown in Figure 5, the hit rate at 40% occlusion was reliably greater than at 50% occlusion (69.2% vs. 67.8%, $t(76) = 2.17$, $p = 0.03$); however, the hit rate at 30% occlusion (68.7%) was not statistically different from, though numerically greater

than, that at 50% occlusion, $t(76) = 1.40$, $p = 0.17$. Sensitivity analyses could be impacted by differences between "old" and "new" items; nonetheless, the pattern was in fact very similar to that found for hit rates. The d' values at 30%, 40%, and 50% occlusion were 1.22, 1.32, and 1.22, respectively. The d' difference between occlusion levels 40% vs. 50% approached significance, $p = 0.09$, and the A_z difference was reliable, $p = 0.03$; the difference between occlusion levels 30% vs. 50% was not reliable by either sensitivity measure.

Inverted images

In order to test whether the effect of occlusion level was specific to upright orientation of images (as we found for faces in Experiment 1), the data from the 138 UCLA control participants were similarly analyzed. The only difference between the control and experimental conditions was image inversion. When all images were inverted, the effect of occlusion level was eliminated. Hit rates for the three occlusion levels were 66.7%, 66.8%, and 67.8%, respectively. There were no statistically significant differences between hit rates at occlusion
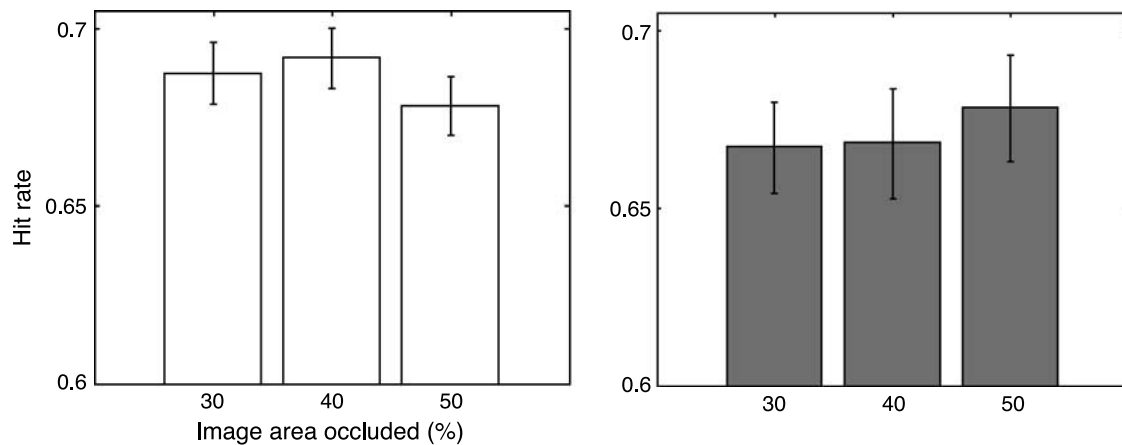
Figure 5. Hit rates for upright (left) and inverted (right) scenes as a function of total image area occluded.

30% versus 50%, or at 40% versus 50% ($t(137) < 1$). Indeed, the hit rate at 50% occlusion was numerically the highest across the three occlusion levels (Figure 5). The d' values for inverted scenes were 0.95, 0.93, and 0.89 for occlusion levels 30%, 40% and 50%, respectively, significantly lower than those obtained using upright scenes ($F(1, 213) = 25.34$, $p < 0.001$).

Discussion

We found evidence that a never-seen but less-occluded face can give rise to more accurate recognition than a more-occluded image of the same face—the image that had actually been seen. Similar results were found with natural scenes. As far as we know, this is the first demonstration of this type of effect that has been shown to be based on changing discrimination sensitivity and not on changing decision criterion or bias.

It might be argued that the observed effect was not surprising, as less occlusion would be expected to lead to better recognition regardless of the occlusion level of the studied image. However, any general advantage of lower occlusion would imply that the less occlusion the better. Instead, we consistently obtained most accurate recognition performance when the occlusion level was at an intermediate level. Moreover, the occlusion effects were found only for upright and not for inverted images. Thus, our findings appear to support an interpretation involving a perceptual process that has some limited capacity to "clean up" an occluded view of a recognizable image, yielding a less-occluded internal representation of it.

Theoretically, any additional information revealed by reduced occlusion may or may not be useful for recognition, depending on how the visual system is presumed to process an occluded image. At one extreme,

if the visual system makes no predictions regarding what is behind occlusion (imagine that a face is replaced by white noise), then the additional information revealed is useless. In that case, identical images would be expected to always give rise to most accurate recognition. However, our results indicate that the visual system treats occluded regions as not completely uncertain, but with certain expectations. These visual expectations appear to be fairly crude and relatively more detailed only near occlusion boundaries. This local nature of occlusion recovery is consistent with previous findings in the literature on perceptual completion (Kanizsa, 1979). Because we anticipated the limited ability of the visual system to recover from occlusion, we chose to use small rectangles as occluders. (Imagine the difficulty of recovery when only the top half of a face or scene is available.) The visual system's expectations also apparently depend on the general familiarity of faces and scenes, as demonstrated by the elimination of the occlusion effect with image inversion. This pattern of results implies that the effect we have found was not completely due to local, lower-level cues, which are commonly assumed to be responsible for amodal completion.

Our results support the traditional notion that perception is an active process that involves organizing and abstracting from sensory information (Gibson, 1991; Gregory, 1970; Helmholtz, 1866/1924). Indeed, given that our experimental paradigm is based on standard techniques used in memory research, it is worth noting that our findings are akin to the "generation" phenomenon in memory research. For example, the word "banana" can be better recognized in a test when "b_n_n_" had been presented in previous word-completion task than when "banana" itself had been presented as an intact word (Hirshman & Bjork, 1988; Jacoby, 1978).

Our criticism of image-based theories of object recognition is that they place insufficient emphasis on the structure and organization of object representations, instead emphasizing the appearance or "snapshot" aspects

of an image or view. Our results do not contradict prior findings concerning object recognition since most studies have investigated the issue of viewpoint invariance in object recognition rather than in overcoming partial occlusion. However, image-based theories of object recognition have been proposed not only to account for viewpoint variation, but also for general variations in image projection from a three-dimensional object. For this reason, we believe that our results have important implications for theories of object recognition.

Acknowledgments

Commercial relationships: none.
Corresponding author: Zili Liu.
Email: zili@psych.ucla.edu.
Address: UCLA Department of Psychology, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095, USA.

References

Balas, B., & Sinha, P. (2007). "Filling-in" color in natural scenes. *Visual Cognition, 15,* 765–768.

Biederman, I. (2000). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision, 13,* 241–253. [PubMed]

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20,* 585–593. [PubMed]

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 121–133.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89,* 60–64. [PubMed] [Article]

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32,* 2385–2400. [PubMed]

Gibson, E. J. (1991). *An odyssey in learning and perception.* Cambridge, MA: MIT Press.

Goldstone, R. L., & Son, J. Y. (2005). Similarity. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 13–36). Cambridge, UK: Cambridge University Press.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Huntington, NY: Robert E. Krieger Publishing Company.

Gregory, R. L. (1970). *The intelligent eye.* London: Weidenfeld and Nicolson.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359,* 231–233. [PubMed]

Helmholtz, H. v. (1866/1924). *A treaties on physiological optics* (vol. 1). New York: Optical Society of America.

Hirshman, E. L., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 484–494.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 179–187. [PubMed]

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667.

Kanizsa, G. (1979). *Organization in vision.* New York: Praeger Publishers.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23,* 141–221. [PubMed]

Lu, H., & Liu, Z (2008). *When a never-seen but less-occluded image is better recognized: Evidence from same–different matching experiments and a model.* Manuscript submitted for publication.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18,* 55–68. [PubMed]

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews, Neuroscience, 5,* 291–303. [PubMed]

Park, S., Intraub, H., Yi, D. J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron, 54,* 335–342. [PubMed] [Article]

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263–266. [PubMed]

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Sekuler, A. B., & Palmer, S. E. (1992). Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General, 121,* 95–111.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67,* 1–20. [PubMed]

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1,* 275–277. [PubMed]