When a never-seen but less-occluded image is better recognized: Evidence from same-different matching experiments and a model

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

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

Empirical studies of human object recognition have found that a query image different from that previously seen almost invariably gives rise to inferior recognition performance than a query image identical to that previously seen. In the present study of same-different matching, we demonstrate that a query image of a face or a Chinese character that was less occluded than that previously seen can yield more accurate positive identification than a query image identical to that previously seen. However, when occlusion of the second image was further reduced, or when the faces were inverted, this effect disappeared. These findings indicate that the representation of a partially occluded object is effectively less occluded, and that the ability of the visual system to overcome occlusion is limited and dependent on its familiarity with the perceived object. A model with limited capacity to overcome occlusion was proposed to qualitatively account for the results.

Introduction

In a visual memory test, recognition accuracy is typically higher if the test image is identical to the study image previously seen than if different from it. In object recognition, for instance, a viewpoint change from a study to a test image invariably yields a decrement in recognition performance, either large (Tarr, Williams, Hayward, & Gauthier, 1998) or small (Biederman, 2000) in magnitude. In memory research in general, maximum recognition performance is typical when a test item is identical to that previously presented (Hirshman & Bjork, 1988).

An intriguing question, then, is what happens when a study image is more impoverished than a test image, for example, when a test image is less occluded. Will the highest recognition accuracy still be obtained when the study and test are exactly identical? Or is it possible that a less occluded test image can give rise to higher accuracy? If the latter, will recognition accuracy keep improving monotonically with less and less occlusion?

These questions will be addressed in the present study. We hypothesized that the frequency of correctly responding "this object has been seen," corresponding to the hit rate in signal detection theory (Green & Swets, 1974), would depend on the similarity of the test image to the *representation* of the study image. For example, if a less occluded test image gives rise to a higher hit rate than does a test image identical to the study image, we would infer that the internal representation of the more occluded study image is effectively less occluded, and hence matches better to the less occluded test image. We now review the relevant literature in order to motivate the present study.

Amodal completion

Perceptually inferring what is behind occlusion is called amodal completion (Kanizsa, 1979). The representation of a partially occluded shape, presumably the result of amodal completion, was investigated by Sekuler and Palmer (1992) in a priming study. A partially occluded shape, such as a $^3/_4$ pie chart (or a pacman) that is perceived as a corner occluding a disk, was first presented. Two shapes were then presented, and observers were asked whether they were the same or different from each other. Discrimination speed was found to be approximately equal when the two shapes were both pacmen and when they were both disks. The authors suggested that the representation of a pacman was its perceived counterpart, the disk, which facilitated subsequent comparison between two disks. However, given that the test sequence was repeated, there was a confound: the pacman may have been simply associated in memory with the disks that followed, independent of any perceptual completion. Note

also that the discrimination speed of the two disks was equal to, but not greater than, that of the two pacmen. In the present study, each trial was shown only once, thereby avoiding the confound. More accurate, rather than equally fast, recognition was sought in the present study when the study and test items were different, relative to when they were the same.

Nakayama, Shimojo, and Silverman (1989) studied the role of amodal completion in face recognition by either allowing or disabling amodal completion (see also He & Nakayama, 1992; Kellman & Shipley, 1991). These researchers stereoscopically switched the relative depth between a face image and a curtain blind. Face recognition deteriorated when a fragmented face was perceived in front of a wall, as compared to when an amodally completed face was perceived behind curtain blind. In the present study, because of the presumed strong perceptual completion of faces, faces were also chosen as the initial class of experimental objects.

The boundary extension effect discovered by Intraub and Richardson (1989) can also be considered as amodal completion of a photo's boundary. What is unique about this effect is that the photo of a natural scene needs to be a close-up view. For instance, after seeing a close-up photo of a plate of spaghetti, an observer's drawing from memory recall expanded the boundary of the original photo, as if the mind's eye view zoomed out. In terms of signal detection theory (Green & Swets, 1974), however, bias and sensitivity were not teased apart. Accordingly, it remains unresolved whether this effect may be entirely due to observers' preference for a farther view when a photo appears too close up. In the present study, the issue of sensitivity versus bias would be specifically addressed.

An observer's ability to amodally complete was quantified by Kersten (1987) using natural images, including faces. A grayscale image was partially occluded by randomly positioned large pixels that were easily distinguishable from the natural image behind. Observers repeatedly estimated the grayscale value of an occluded pixel with feedback, until they were correct. The number of guesses was then used to quantify an observer's ability to predict an occluded pixel value from the rest of the image. It was found that, when occlusion was sparse (1%), the value of an occluded pixel was well predicted by its nearest neighbors. The prediction was robust regardless of whether the mean or median was computed. The model in the present paper was inspired by the Kersten study, in that the value of an occluded pixel was also estimated by the model using the nearest unoccluded pixels.

The overall aim of the current study, that is, determining whether perceptual abstraction via amodal completion may better characterize the internal representation of a seen image than the image itself, can be traced back to the classic study of Posner and Keele (1970). There, participants were trained to categorize random-dot configurations, each of which was created by randomly perturbing the dot positions from a predetermined configuration,

termed a prototype. During training, the prototypes were never shown. Participants were tested immediately after training and one week afterwards. Classification errors for the trained exemplars were found to increase in one week's time (.14 to .39), whereas those for the prototypes changed little (.35 to .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, also seemed represented, and in a more stable manner. However, as indicated by the numerical error rates, the never-seen prototypes were never categorized more accurately than the trained exemplars. In the present study, we sought to find conditions under which unseen stimuli may be recognized more accurately than the studied stimuli.

In a companion paper to the present study, Lu and Liu (2008) designed old-new rating experiments to study the effect of occlusion on recognition memory. In contrast to the same-different matching task in the present study, a standard technique in memory research was used to study longer-term internal representations. Participants first rated the attractiveness of a face or natural scene, and then rated how likely a scene had been seen. In comparison to the present study's red pixel occluders, red rectangles were used. Lu and Liu (2008) demonstrated that an "old" face or scene whose image had not been seen but less occluded was more accurately recognized as "old" than a more occluded image, identical to that previously seen.

Overview of the current study

We parametrically varied the proportion of an image's occlusion, using pixel occlusion to take advantage of local redundancy in natural images. Rather than using large gray occluding pixels as in the study of Kersten (1987), we used small red pixels occluding a grayscale image, and each image could be occluded as much as 60%. Hence, although the unit of occlusion was a small pixel, occluding pixels were often contiguous. Our hypothesis that faces strongly complete perceptually is based on evidence of the human ability to reliably perceive faces in impoverished images (Moore & Cavanagh, 1998). To preview the present findings, in a same-different sequential matching task, we found that a never seen but 50% occluded face image gave rise to more accurate recognition than the image of the same face that was 60% occluded and identical to that previously seen. Interestingly, further removal of occlusion from 50% only worsened the performance. The effect disappeared when all faces were turned upside-down. A similar effect was found using Chinese characters as perceived by participants who could not read Chinese. Finally, we proposed a model with limited capability to recover from occlusion that can qualitatively explain all patterns of empirical findings.

Experiments

Experiment 1 (pilot): 25 occlusion combinations

Stimuli

Seventy-five grayscale face images were used, which were from the FERET database of the National Institute of Standard and Technology (Phillips, Moon, Rizvi, & Rauss, 2000; Phillips, Wechsler, Huang, & Rauss, 1998). Each face was from a different individual. No spectacles, makeup, or facial hair was present in any of the images. Only the inner portion of a face was visible through an oval aperture (195 × 145 pixels, 5.3° × 3.9° in visual angle), such that no ears or hair was visible. In order to reduce the possibility that participants rely on local features for recognition, each image was, before occlusion was applied, low-pass filtered with a two-dimensional Gaussian kernel (standard deviation = 5 pixels, window size = 10 × 10 pixels). Each face was centered within a dark background (.29 cd/m^2) of 6.4° × 9.1° in visual angle. The average luminance of the face region was 17.1 cd/m^2.

Occlusion was created by randomly replacing pixels in a low-pass filtered image with red pixels (1.6 × 1.6 min of arc apiece). The red pixels had the luminance of 21.2 cd/m^2 and CIE color chromaticity of x = .684, y = .316. There were five occlusion levels in terms of occluded image area: 20%, 30%, 40%, 50%, and 60%. An important constraint on the occlusion was that, for each face, additional occlusion was generated by adding red occluding pixels, while keeping the existing ones intact. Accordingly, the red pixels in a more occluded image were always a superset of those in a less occluded image of the same face (Figure 1). This is termed the hierarchical design of occluding pixels in this paper.

Procedure

The experiment used a same-different sequential matching task. In each trial, the image sequence was as follows (Figure 2): the first face image (1000 ms), mask (200 ms), fixation (50 ms), the second face image (500 ms), mask (200 ms), and fixation. To further reduce the possibility
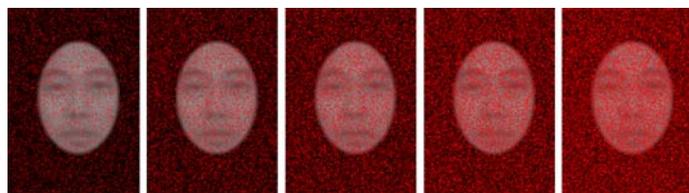

Figure 1. The same face being occluded by increasing number of red pixels that cover 20%, 30%, 40%, 50%, and 60% of an image, respectively. Occlusion was increased by adding red pixels while keeping the existing ones intact.
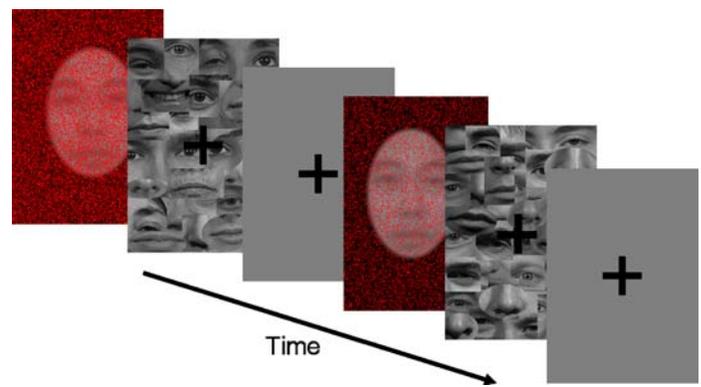

Figure 2. Image sequence in one trial: the first face image, mask, fixation, the second face image, mask, and fixation. The participant was asked: "Do the two faces show the same person?"

that participants rely on local features, the second face was the mirror reflection of the first in a "same" trial; the corresponding red pixels were mirror reflected also. For example, when both the occlusion and face were the same, the first and second images were identical except for a mirror reflection. The display location of the second face image was also randomly shifted within a horizontal and vertical range of 5 to 10 pixels. Participants decided whether the faces shown in the two images were the same or different. No feedback was provided.

Occlusion of each of the two images in a trial was at one of the five levels, generating a total of 25 occlusion combinations. In either a "same" or a "different" trial, the hierarchical design of red occluding pixels was retained for the first and second images. For example, if the first image was 60% occluded and the second image 50% occluded, the distribution of the red pixels in the second image was a subset of that in the first image. Distributions of red pixels of different faces were uncorrelated otherwise. Among the 75 faces in total, 25 were randomly chosen to be in the 25 "same" trials, one for each occlusion combination. By randomly choosing 25 faces from the remaining 50, the second set of 25 "same" trials was similarly created. Finally, 25 "different" trials were created by using 25 first images randomly selected from the 50 "same" trials, and using the remaining 25 faces as the second images.

An important control provided by this design can be illustrated as follows. Let us label a trial as "60–50 same" if the first image was 60% occluded, the second 50% occluded, and the same face was shown. If a "60–60 same" trial is outperformed by a "60–50 same" trial, this result cannot be due to any prior presentation of the "50" image.

Participants were informed that a "same" trial was twice as likely as a "different" trial. The selection, pairing, and occlusion of faces were randomized across participants. These 75 trials consisted of a single block and were presented in a random sequence. The same block was repeated five more times, each with a different random

sequence. In this paper, however, only data from the first block will be presented. Each participant took approximately 40 min to complete the experiment.

Apparatus

Participants binocularly viewed from a chin-rest the stimuli 57 cm away, through a dark tube that abutted the computer display. The experiment was conducted in dark rooms using calibrated computer monitors with a refresh rate at 75 Hz. MatLab software (Math Works, Inc.) and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) were used.

Participants

Fifty-seven psychology undergraduate students at the University of California, Los Angeles (UCLA) participated for course credit. Thirty-two of them were tested with upright faces, and 25 with inverted faces.

Results and discussion

Figure 3 shows the hit rates across the 25 occlusion combinations. Using hit rates was necessary because there were only two "same" trials per occlusion combination per participant, so discrimination sensitivity d' was impossible to calculate. Using hit rates was also appropriate because this measure directly tested the main hypothesis. If d' was used instead, one would still have to ascertain whether a higher d' was due to more hits, or more correct rejections. It is apparent that the data in the first four panels of Figure 3 did not support the main hypothesis. This is because, in each panel, the hit rate when the two images were identical (20–20, 30–30, 40–40, and 50–50) was no worse than otherwise. However, the data in Panel 5 did provide supporting evidence. After confirming the effect in Panel 5 in subsequent experiments, we will return to explain the results in all panels. For now, we focus on Panel 5.

In Panel 5, when the first image was 60% occluded, the hit rate was higher at 60–50 when the second image was 50% occluded than when it was identical to the first image at 60–60 (.96 vs. .82). Levene's test revealed that the data violated the assumption of equal variance ($F(4, 155) = 18.10, p < .001$). Accordingly, we performed the Dunnett's T3 multiple comparison test without assuming equal variance and found that this difference had a p value of .15 ($t(31) = 2.31$).

In order to test whether this hit rate difference was specific to upright faces, we repeated this experiment with inverted faces while keeping everything else unchanged. As shown in dashed lines in Figure 3, the original difference found with upright faces (solid line) in Panel 5 disappeared. This result indicates that the difference depended on upright faces, an object class that is familiar

and presumably has strongly constrained internal representations. It further indicates that the hit rate difference was not due to cues that are invariant of image inversion, such as spatial frequency change from 60% to 50% occlusion.

Another noteworthy pattern of the results was that when the occlusion of the first and second images was different, the participants had a bias to respond "different." This was particularly the case in the inverted 50–20 "same" trials (Panel 4), where the hit rate was below chance.

The results of Experiment 1 thus revealed a trend that the best recognition performance was not always achieved when identical images were presented (with a mirror reflection). Instead, less occluded images could yield better recognition. We believe that the lack of statistical significance with multiple comparisons was due to the small number of trials in each occlusion combination. We therefore performed Experiment 2 in an effort to confirm the reliability of the trend observed in Experiment 1.

Experiment 2: The first image with 60% occluded

This experiment was identical to Experiment 1, except that the occlusion of a first image was always 60% so that there were more trials in each of five occlusion combinations. Thirty-nine fresh UCLA students were tested with upright faces, and 22 with inverted faces.

As shown in Figure 4, with upright faces, the hit rate at 60–50 was .12 higher than at 60–60 when the second image was identical to the first image (with a mirror reflection), confirming the trend obtained in Experiment 1, where the difference was .14. Levene's test again revealed a trend toward unequal variance in the data ($F(4, 190) = 2.08, p = .09$). Consequently, the Dunnett's T3 test was used for multiple comparisons without assuming equal variance. The hit rate difference between 60–50 and 60–60 was statistically significant: $t(38) = 3.69, p = .02$. In Experiment 2, sensitivity d' could also be calculated, since there were sufficient trials. Dunnett T test was used after confirming the equal variance assumption by Levene's test ($F(4, 190) < 1$). Discrimination sensitivity d' confirmed the hit results. Specifically, the 60–50 condition yielded higher d' than 60–60: $t(38) = 3.15, p = .02$. There was no statistically significant difference in bias between 60–50 and 60–60, $t(38) < 1$.

The dashed lines in Figure 4 show the results from the control condition when all faces were inverted. It is apparent that the effect with upright faces was now reversed—the hit rate at 60–50 was lower than at 60–60: $t(21) = 2.10, p = .04$ (Dunnett's T test). The bias was only marginally different between 60–60 and 60–50: $t(21) = 1.76, p = .09$. It is worth noting that at 60–60, the hit rate was comparable between the upright and inverted faces (.79 and .81).
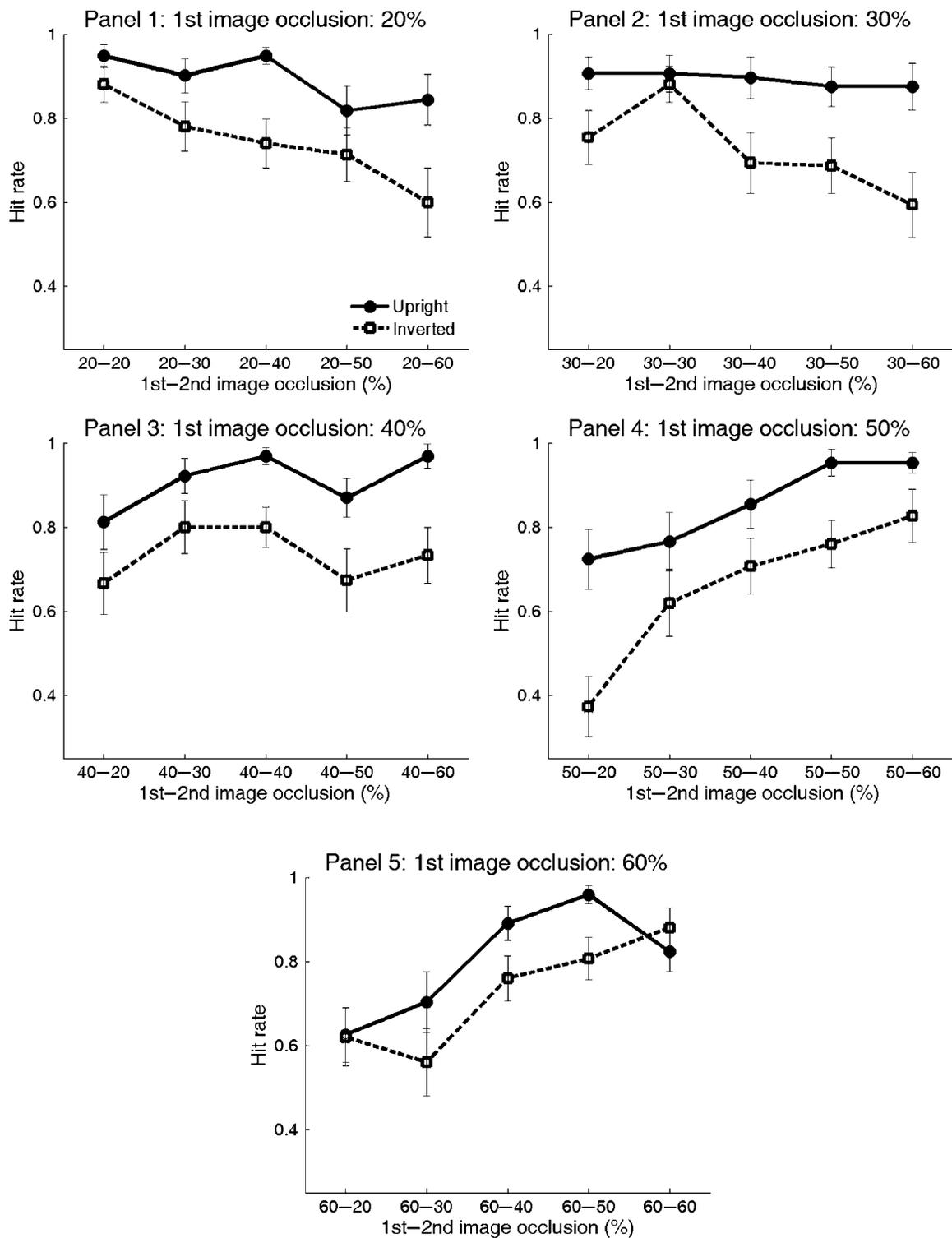
Figure 3. Hit rates with upright (solid lines) and inverted faces (dashed lines). Data from the 25 occlusion combinations are organized into five panels. In each panel, the first image's occlusion area is constant while the second image's occlusion varies. Error bars represent standard errors of the mean here and in subsequent plots.

Taken together, Experiment 2 revealed that the major effect—a never seen, but less occluded image gave rise to a higher hit rate and greater discrimination sensitivity than the image identical to that previously seen—was dependent on upright faces, probably due to the high familiarity of faces. This experiment further demonstrated that this effect was spatially limited, in that it was based on "removal" of only 1/6 of the occlusion (from 60% down to 50%).
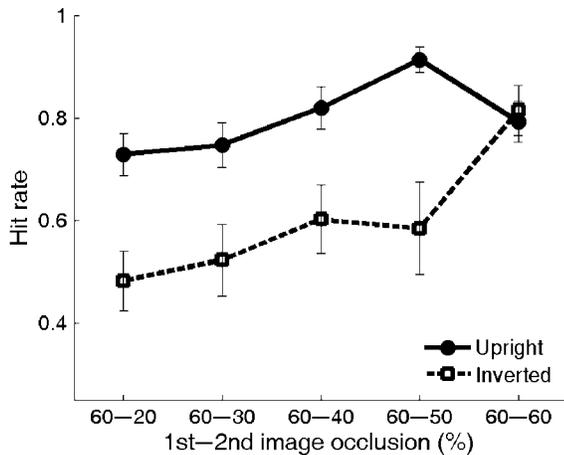
Figure 4. Hit rates with upright and inverted faces in Experiment 2, when the first image was 60% occluded.

Experiment 3: Recognition without mirror reflection

In Experiments 1 and 2, when the two faces in a trial showed the same person, they were mirror images of each other. In Experiment 3, we tested whether the effect found so far could still exist without mirror reflection. This experiment was otherwise identical to Experiment 2. Twenty-five fresh UCLA students were tested with upright faces.

As shown in Figure 5, the effect in Experiments 1 and 2 was replicated. Namely, the hit rate at 60–50 was statistically higher than at 60–60, $t(24) = 2.22$, $p = .04$ (since we specifically planned to compare 60–50 versus 60–60, no correction for multiple comparisons was made). The bias difference between 60–50 and 60–60 was not statistically significant ($t(24) = 1.65$, $p = .11$).

Experiment 4: Recognition of Chinese characters

Experiment 4 was designed to assess whether the key effect found thus far was specific to upright faces. Since the effect was not found with inverted faces, perhaps high stimulus familiarity was required. On the other hand, the effect may be obtainable from stimuli with Gestalt cues such as good continuation, which may not be prominent for complex face images. To test this possibility, Experiment 4 employed simple Chinese characters by testing participants who could not read Chinese.

Two hundred simple Chinese characters were used, each of which had less than five strokes. Figure 6 shows an example character at five occlusion levels. The experimental design was identical to Experiment 1 with 25 occlusion combinations, except the display duration of the second image was 250 ms (rather than 500 ms). No mirror reflection or inverted characters were used.

UCLA undergraduate students were similarly recruited as in the previous experiments, except that the word "face" in the recruitment description was replaced by either "Chinese character" or "shape." Reading ability in Chinese was not mentioned in the recruitment advertisement. Before the experiment started, participants were asked whether they could read any Chinese, and were labeled either as readers or non-readers. Among the 149 participants tested, 29 were readers, 120 were not.

Figure 7 shows the hit rates for the 120 non-readers. The key effect again occurred as in the upright face experiments. When the first image was 60% occluded, the hit rate at 60–40 was higher than at 60–60 ($t(119) = 2.68$, $p = .047$, Dunnett's T test for multiple comparisons). This finding indicates that the key effect obtained in the earlier experiments is not limited to face recognition only.

The pattern of performance of the 29 readers apparently depended on their fluency. For instance, both authors are native Chinese and their performance was perfect across all levels of occlusion. Because the readers' fluency was not objectively measured and the number of readers was relatively small, the study of perceptual abstraction as a function of expertise is beyond the scope of the present study and will not be reported in this paper.

A model

Model assumptions and results

We now provide a computational model, inspired by Kersten (1987), to account for empirical findings with 25 occlusion combinations, and with upright and inverted faces. We first list the model assumptions:

1. The model's ability to recover from occlusion was limited, depending on the familiarity and complexity
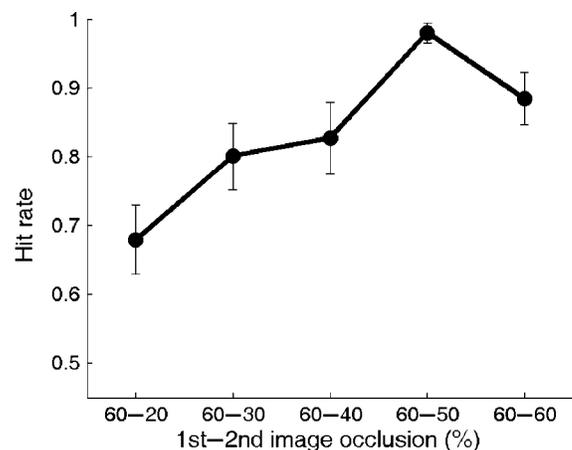

Figure 5. Hit rates with upright faces in Experiment 3, when the first image was 60% occluded and no mirror reflection was used.
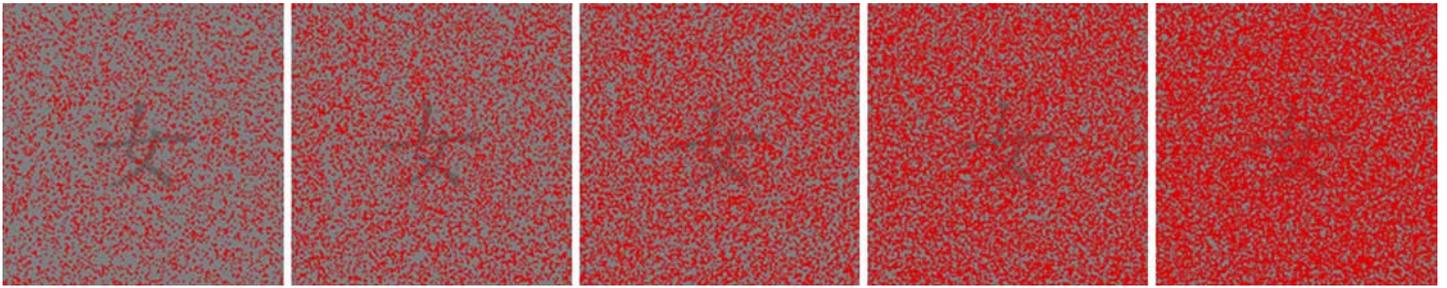
Figure 6. An example Chinese character at five occlusion levels.

of the objects. Specifically, the number of recoverable pixels that were occluded was set as a free parameter. By exploring the parameter space of the model, we found that if 45% to 55% of an image's total pixels were recoverable for upright faces, and 0% to 1% for inverted faces, all results in the face experiments could be accounted for qualitatively.

2. When an occluded pixel was recoverable, the recovered value was the average of its four nearest and unoccluded neighbors. In comparison, when an occluded pixel was unrecoverable, its value was uncertain, modeled as being randomly sampled from the uniform distribution between 0 and 255. Whether a pixel was recoverable or not was determined randomly in each image.

3. When the first and second images were compared, the mirror reflection and the relative positional shift were assumed known.

4. After the values of all occluded pixels had been assigned, the difference between the first and second images was computed. As a result, a histogram of the resulting pixel value differences was obtained. The histograms obtained from the "same" trials were called "same" histograms. Otherwise they were called "different" histograms. Both "same" and "different" histograms were typically symmetric and centered at zero. A "different" histogram was more spread out than its "same" counterpart that had many more zero values.

5. Recognition performance was characterized by how the average "same" and "different" histograms could be best discriminated. Recall that in signal detection theory, discrimination between signal and noise is characterized by the ideal observer's sensitivity. Analogously, the area under the Receiver Operating Characteristic (ROC) was computed to characterize discrimination sensitivity between the average "same" and average "different" histograms, for each occlusion combination.

It should be noted that the qualitative pattern of recognition performance was not critically dependent on this measure of the ROC area. The qualitative results were unchanged when either the analogous "hit rate" (when the

likelihood ratio criterion was set at unity to maximize accuracy), or the Chi-square distance, or the Kullback-Leibler distance (Kullback & Leibler, 1951) between the average "same" and average "different" histograms was used. The two average histograms were numerically computed from the 75 face images used in the human experiments.

Figure 8 depicts the model results for all conditions, when the completion capacity was set at 50% of the total pixels for upright faces, and 0% for inverted faces. In Appendix A, a simplified example is provided to illustrate why the dashed curve (for inverted faces) in Panel 5 is monotonically increasing. Once this example is understood, the rest of the results can be similarly understood.

Finally, it should be noted that the model results matched human performance qualitatively, but not quantitatively. In particular, whereas human performance at 60–60 was comparable for both upright and inverted faces, the model performance differed substantially (95% vs. 78%). It should also be noted that stimulus presentation time was assumed irrelevant to the model, which is over simplified because a 60–20 "same" trial, for example, was not as easy as a 20–60 "same" trial for humans.

The same model similarly accounted for the findings from Experiment 4 with Chinese characters, though the parameter values of the completion capacity were different from those in the face experiments. This variation in parameter values is reasonable because the ability to perceptually complete likely depends on stimulus complexity and familiarity.

Additional model predictions for new experimental designs

Recall that in the experimental design thus far, the distributions of the red occluding pixels were hierarchical. That is, additional occlusion was created by adding red occluding pixels without relocating existing ones. Alternatively, whenever occlusion changes, the position of a red pixel could be randomly re-positioned. Only when the two images share the same occlusion and face, would the images be identical with exactly the same distributions of
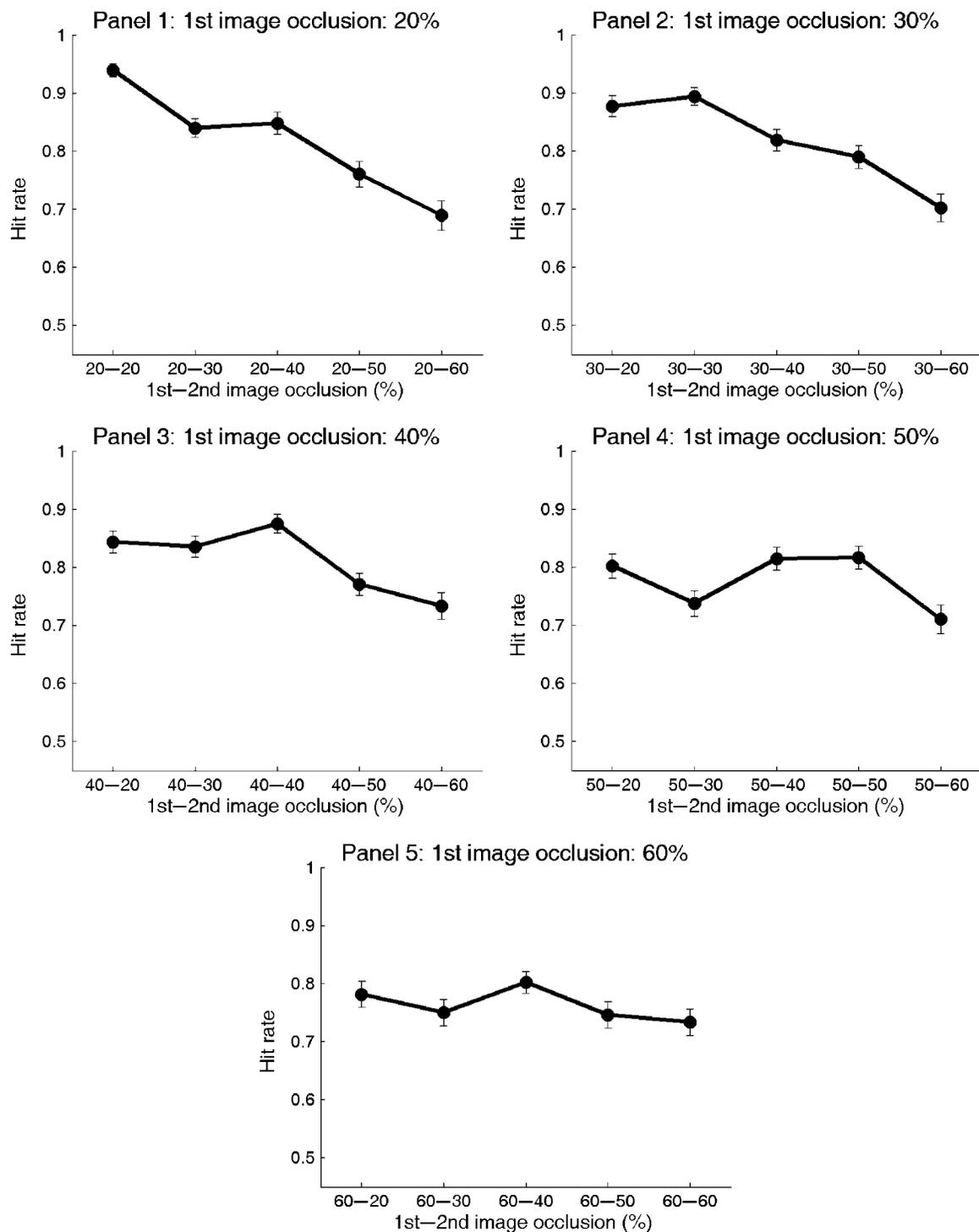
Figure 7. Hit rates with Chinese character in Experiment 4. Data from the 25 occlusion combinations are organized into five panels.

the red pixels. With this design, the model's overall performance deteriorated, while its qualitative pattern of results remained unchanged. Figure 9 shows the simulation results and the corresponding new results from human participants. The corresponding human experiment with the new occlusion design will be described in Experiment 5.

Experiment 5: Random distributions of the red occluding pixels

Occlusion was created by completely and randomly redistributing red pixels (when both the faces and occlusion were the same, the two images remained identical). The design and procedure were otherwise
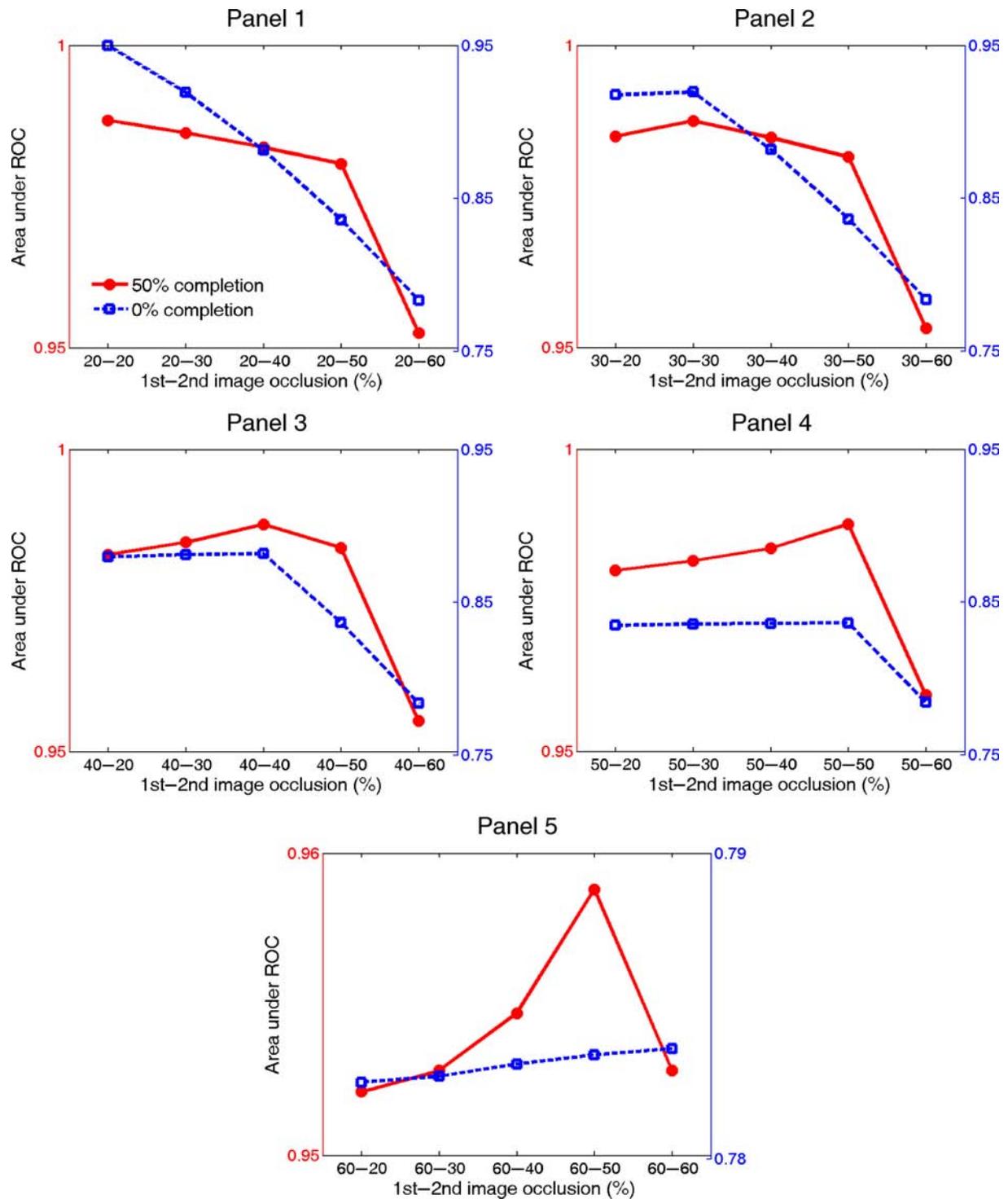
Figure 8. Simulation results for the 25 occlusion combinations. The results plotted qualitatively match human performance. The numerical values of completion capacity were: 50% for upright faces (red curves, y-scales on the left), and 0% for inverted faces (blue curves, y-scales on the right). Note that the y-scales in Panel 5 are different from the rest of the panels.

identical to that of Experiment 1. Twenty-four fresh UCLA students participated.

As shown in Figure 9B (solid curve), the qualitative pattern of the original effect in Experiment 1 was replicated. The hit rate at 60–50 was higher than at 60–60 ($t(23) = 2.39$, $p = .03$ without correction for multiple comparisons), consistent with the model prediction. Furthermore, a clear trend was found for a main effect that this new design reduced the overall recognition performance as compared to the original hierarchical
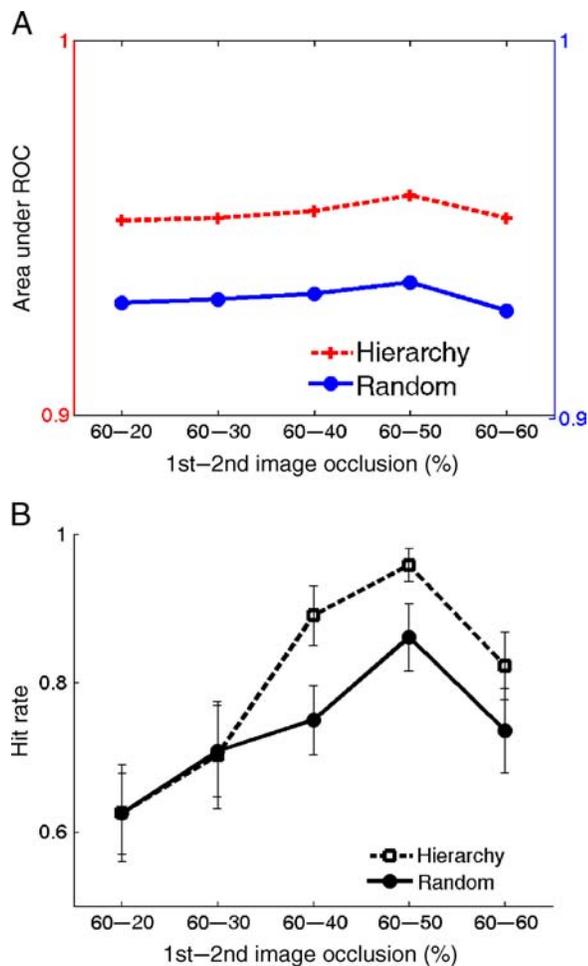
Figure 9. (A) Comparison between simulation results in two experimental designs. The designs were hierarchical (red dashed line) in Experiment 1, in which the red pixels in a less occluded image were a subset of those in a more occluded image in a "same" trial; and random (blue solid line) in Experiment 5, in which the red pixels were randomly and independently distributed (the two images remained identical in a 20–20, or 30–30, 40–40, 50–50, or 60–60 "same" trial). The simulation results are plotted as the area under ROC as a function of the second image occlusion level with the assumption that up to 50% of the total number of pixels were recoverable. (B) Human hit rates in the hierarchical design (dashed line) in Experiment 1 and in the random design (solid line) in Experiment 5, when the first image was 60% occluded.

design ($t(54) = 1.75$, $p = .045$, one-tailed), consistent with the simulation results of the model in Figure 9A.

Discussion

In the literature on object recognition in particular and on memory research in general, an image identical to that previously seen has been invariably found to give rise to better recognition than an image different from that previously seen. The present study, together with its companion study (Lu & Liu, 2008) demonstrate for the first time, to our knowledge, that a different but less-occluded image could yield more accurate recognition than a more occluded-image that is identical to that previously seen.

One might argue that this result is not surprising because less occlusion means "more stimulus information," and hence better recognition performance. However, whether additional information in a second image is useful or not depends on how the first image is processed. At one extreme, if the first image has been stored as a template, then any additional information from the second image should not improve recognition. A reasonable scenario for template memorization is when occlusion is severe such that the occluded object is unrecognizable. At the other extreme, if the visual system were infinitely capable of removing any occlusion, then the less occluded the second image is, the better recognition will be. What we found for upright faces was inconsistent with either of these possibilities. The finding was, however, consistent with the traditional notion that the visual system organizes stimulus information into a more coherent and abstract representation than a template, with limited capability. For example, in size perception, visual recognition is determined by perceived, rather than retinal, size (Bennett, 2007; Milliken & Jolicoeur, 1992).

One might also argue that the pixel occlusion used in the current study was atypical, in the sense that the "recovery from the red noise" may have more to do with transparency (because of effective spatial summation) than with amodal completion. Since larger pixels were used as occlusion without controversy (Kersten, 1987), the criticism must be based on the size of an occluding pixel. However, given that pixel size is a continuum, it appears unlikely that one mechanism (transparency) is at work below a certain size threshold and another distinctly different mechanism (amodal completion) is at work above the threshold. By definition, amodal completion is to infer what is behind the visible. So in this sense, referring to the red pixels in the current study as occluders is technically correct. It remains an open question whether transparency perception and amodal completion are separate mechanisms when occluders are small or whether the distinction is meaningful, given our limited understanding of either process. Furthermore, spatial summation as a possible mechanism of transparency perception is unimportant in the present study because it is independent of image inversion. We used the term "occlusion" in the present paper because the study was the first in a series of experiments that manipulated occlusion size. When red rectangles were used as occluders in an old-new memory task with faces and natural scenes (Lu & Liu, 2008), results similar to those of the present study were obtained.

Appendix A

Qualitative explanation of the model

The purpose of this appendix is to present a simplified example to qualitatively approximate the model simulation. We will use the comparison between 60–50 and 60–60 conditions with 0% completion capacity. Once this example is understood, the other predictions of the model results can be better understood as well. Note that some assumptions will be introduced in the appendix simply for the ease of analytic derivations.

A pixel value in the difference image can be categorized into one of the following three groups: 1) when the two corresponding pixels in the 1st and 2nd images are both unoccluded, the pixel value difference is assumed to follow a uniform distribution within $[-r_1, r_1]$; 2) when one corresponding pixel is unoccluded and the other occluded but unrecovered (replaced with a random sample from 0 to 255), the difference is assumed to follow a uniform distribution within $[-r_2, r_2]$; 3) when both corresponding pixels are occluded and unrecovered, the difference is assumed to follow a uniform distribution within $[-r_3, r_3]$ (simplified from convolution between two uniform distributions). Given the nature of these three groups, we assume that $r_1 < r_2 < r_3$, which is verified by simulations.

Let x_i denote the proportion of Group i pixels ($i = 1, 2, 3$). Then in the 60–50 histogram, $x_1 = 0.4$, $x_2 = 0.1$, and $x_3 = 0.5$. In the 60–60 histogram, $x_1 = 0.4$, $x_2 = 0$, and $x_3 = 0.6$. Note that $x_1 + x_2 + x_3 = 1$.

Note also that the only difference between the "same" $s(j)$ and "different" $d(j)$ histograms (j indicates integer pixel value) is that the Group 1 pixels are zero in the "same" histogram, and are uniformly distributed within $[-r_1, r_1]$ in the "different" histogram. We can now write the probability distribution $d(j)$, while noting that $s(j)$ can be similarly expressed when letting $r_1 = \frac{1}{2}$:

$$d(j) = \begin{cases} 0, & \text{when } r_3 \prec |j|; \\ \dfrac{x_3}{2r_3}, & \text{when } r_2 \prec |j| \le r_3; \\ \dfrac{x_3}{2r_3} + \dfrac{x_2}{2r_2}, & \text{when } r_1 \prec |j| \le r_2; \\ \dfrac{x_3}{2r_3} + \dfrac{x_2}{2r_2} + \dfrac{x_1}{2r_1}, & \text{when } |j| \le r_1. \end{cases} \qquad \text{(A1)}$$

Therefore, as a measure of distribution difference, the *Kullback-Leibler* distance, *KL(d, s)*, is effectively determined by the difference of probability distributions for the Group 1 pixels between the "same" and the "different" conditions.

$$
\begin{aligned}
KL(d,s) &= \sum_j d(j)\log\frac{d(j)}{s(j)} \\
&= \frac{1}{2}\left(\frac{x_1}{r_1} + \frac{x_2}{r_2} + \frac{x_3}{r_3}\right)\log\left(\frac{\dfrac{x_1}{r_1} + \dfrac{x_2}{r_2} + \dfrac{x_3}{r_3}}{2x_1 + \dfrac{x_2}{r_2} + \dfrac{x_3}{r_3}}\right) \\
&\quad + \frac{2r_1 - 1}{2}\left(\frac{x_1}{r_1} + \frac{x_2}{r_2} + \frac{x_3}{r_3}\right)\log\left(\frac{\dfrac{x_1}{r_1}}{\dfrac{x_2}{r_2} + \dfrac{x_3}{r_3}} + 1\right).
\end{aligned}
$$
$$\text{(A2)}$$

It is then straightforward to use Taylor expansion to show that this distance monotonically decreases as a function of x_2 (while x_1 is kept constant), so long as $x_1 + x_2 \le 1$ and $r_2 < r_3$. Recall in the 60–60 condition, $x_1 = 0.4$ and $x_2 = 0$; in the 60–50 condition, $x_1 = 0.4$ and $x_2 = 0.1$. Hence, the KL distance is greater in the 60–60 condition than in the 60–50, indicating that the 60–60 condition outperforms 60–50 when the completion capacity is 0%. This result from the analytic derivations is consistent with the simulation result.

Acknowledgments

Editor's note

This paper was submitted and reviewed as part of the special issue on Perceptual organization and neural computation http://www.journalofvision.org/8/7/i/.

References

Bennett, D. J. (2007). Does an estimate of environmental size precede size scaling on a form-comparison task? *Perception, 36,* 375–390. [PubMed]

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

Brainard, D. J. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

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

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

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]

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]

Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A, Optics and Image Science, 4,* 2395–2400. [PubMed]

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22,* 79–86.

Lu, H., & Liu, Z. (2008). When a never-seen but less-occluded image is better recognized: Evidence from old-new memory experiments. *Journal of Vision, 8*(7):31, 1–9, http://journalofvision.org/8/7/31, doi:10.1167/8.7.31. [PubMed] [Article]

Milliken, B., & Jolicoeur, P. (1992). Size effects in visual recognition memory are determined by perceived size. *Memory & Cognition, 20,* 83–95. [PubMed]

Moore, C., & Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Cognition, 67,* 45–71. [PubMed]

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]

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face recognition algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,* 1090–1104.

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. (1998). The FERET database and evaluation procedure for face recognition algorithms. *Journal of Image and Vision Computing, 16,* 295–306.

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