Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 45 (2005) 3262–3276
Vision Research
www.elsevier.com/locate/visres

The recognition of partially visible natural objects in the presence and absence of their occluders

Jeffrey S. Johnson [a,*], Bruno A. Olshausen [a,b]

[a] *Center for Neuroscience, University of California Davis, Davis, CA, USA*
[b] *Redwood Neuroscience Institute, Menlo Park, CA, USA*

Received 21 December 2004; received in revised form 13 May 2005

Abstract

The visual system is adept at compensating for the missing information in scenes that results from occlusion, but how this is done is not fully understood. In particular, the role of the occluding object in visual processing and its effect on the subsequent recognition of the occluded object is unclear. We report three human behavioral experiments suggesting that the recognition of partially visible objects is facilitated when the missing object information is replaced by an occluder rather than simply removed. Furthermore, we provide EEG evidence suggesting that the processes responsible for facilitated recognition occur relatively early in the visual stream.

Keywords: Object recognition; Occlusion; Amodal completion; Deletion; Natural images; ERP

1. Introduction

The world that we live in is a cluttered one. In contrast to the controlled realm of the laboratory, it is only the rare object which is seen in isolation in our daily lives. Many objects are partially occluded by other intervening objects. Despite this, we do not have the impression as we view the world that it is filled with object fragments—the objects that we see appear to be complete ones. Although our impressions of object wholeness could conceivably arise at a purely conceptual level, there are strong ecological reasons to think that our visual system should understand the natural rules of occlusion and have developed some mechanisms at the perceptual level to account for occlusion and other forms of missing information in a scene in the early or intermediate stages of visual processing (Nakayama, He, & Shimojo, 1995).

There are two main types of completion effects that compensate for missing or ambiguous information in the retinal image: *modal* and *amodal* completion (Michotte, Thinès, & Crabbé, 1964/1991). Modal completion is a process that results in effects such as illusory contours (Kanizsa, 1979) and neon color spreading (van Tuijl, 1975). Modal completion is perceptually salient despite having no physical counterpart in the retinal image. Neural correlates of modal completion have been demonstrated in V2 (von der Heydt, Peterhans, & Baumgartner, 1984) and recently, in V1 (Lee, 2003). Amodal completion is the term used to describe the continuation of object contours and surfaces behind occluders, a process which does not manifest a perceptual counterpart. Because occlusion events are common and the illusory conjunction of unrelated contours is rare, amodal completion is more applicable to natural images than modal completion. There are suggestions that the two types of completion are mediated by the same mechanisms (Kellman, Yin, & Shipley, 1998, but see Singh, 2004), but the neural correlates of amodal completion that have been seen are weak in comparison

---

[*] Corresponding author. Tel.: +1 530 757 8789.
E-mail address: jsjohnson@ucdavis-alumni.com (J.S. Johnson).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.06.007

to those arising from modal completion (Lee, 2003; Sugita, 1999), and the fact that amodal effects do not result in visible contours suggests that they may be post-perceptual.

Despite the ubiquity of occlusion in the world, most models of visual object recognition are not specifically equipped to account for amodal completion effects, instead focusing their efforts on the goodness-of-match of a feedforward analysis of the image with an object model (Fukushima, 1980; Mel, 1997; Riesenhuber & Poggio, 1999; Ullman & Bart, 2004; Ullman, Vidal-Naquet, & Sali, 2002; VanRullen & Thorpe, 2002). When object fragments are missing, these models do not perform completion, relying instead on matching only the present fragments to the object model. Such an approach is relatively easy to implement, but does not take into account depth-based image cues which distinguish between objects which are partially visible due to occlusion and objects which are partially visible because some of the object is missing. Similarly, a model that performs completion indiscriminately (e.g., Kellman, Guttman, & Wickens, 2001) cannot distinguish between these two cases without resorting to higher-level information to prune inappropriate completions after they have been made (Kellman, 2003). Models such as these predict that recognition (or at least early visual processing) of partially visible objects will not be affected by the presence or absence of an occluder.

On the other hand, a model such as Biederman's Recognition-By-Components (Biederman, 1987; Hummel & Biederman, 1992) makes an explicit attempt to determine which edges in the scene should be bound together using local contour junction rules. Going a step further, some models (Fukushima, 2005; Lee & Mumford, 2003; Nakayama et al., 1995) begin by establishing, with the help of feedback, a global surface-based representation of the scene at low levels of the system. Such models employ inferred depth relations from the earliest representations of the scene and explicitly predict that the visual system will treat image fragments as a single object under occluded conditions (when amodal completion should occur), but separately when global image structure suggests completion is not appropriate. These models suggest that the presence or absence of a depth-appropriate occluder plays a crucial role in determining whether completion occurs, thereby having an effect on the recognition of a partially visible object.

Does the visual system take these depth relations into account when performing amodal completion? One method to determine the effects, if any, that the presence of an occluder has on the recognition of partially visible objects is to construct two sets of images, one with occluded objects and another containing the same object fragments with the occluder removed and the previously occluded regions open to the background. This latter form of image, which by virtue of its depth relations is amodal-inappropriate, we will call "deleted". An early demonstration by Bregman (1981) suggested that a set of outline letterforms which is partially obscured by an occluder in two dimensions is subjectively easier to perceive than the same letter fragments with the occluder removed. Psychophysical studies on similar stimuli have suggested the opposite (Brown & Koch, 2000; Brown & Koch, 1993), showing that subjects are, in general, slower to identify occluded letter fragments than deleted ones. Another study (Gerbino & Salmaso, 1987) using a matching task with outline shape stimuli determined that subjects were faster and more accurate in matching an intact template shape to an occluded version of the shape than to a deleted one. These conflicting studies do not leave a clear picture of the relative difficulty of recognition of occluded and deleted line objects and invite questions about to what extent the visual system entertains depth relations when performing amodal completion.

Another open question is at what stage of the visual pathway the completion of occluded objects is performed. The idea that it may be accomplished quite early is supported by psychophysical studies in visual search of occluded objects (Rensink & Enns, 1998). Neurophysiological evidence of amodal contour responses about 100 ms after presentation of occluded images in macaque V1 cells (Lee, 2003) also supports an early view. However, because occlusion is a function of the relative depth of the objects in the scene, the ability to distinguish an occluded object from a deleted one would appear to rely on an initial determination of a scene's depth relations. Peterson and Gibson (1994) have found behavioral evidence that the determination of depth relations can be contingent upon object contour cues, suggesting that depth relations may be assigned relatively late in visual processing. Furthermore, human event-related potential (ERP) studies on contour closure have suggested that the processes involved in forming a unified percept of a deleted line object are measured no earlier than 230 ms after presentation on electrodes over occipital cortex (Doniger et al., 2000).

The vast majority of the above evidence regarding the efficacy and timecourse of the visual processing of partially visible objects has come from studies of simple shapes and line objects. A notable exception is Nakayama, Shimojo, and Silverman (1989), which shows that photographic face fragments interrupted by bars are easier to recognize when the bars are stereoscopically occluding the face than when the same bars are presented behind the face in a deleted fashion. Such an effect could well rely on stereoscopic depth cues and prove non-replicable in 2D images, even if the relative depth of the objects can be inferred. The studies described here are dedicated to investigating the role of occluders in the recognition of 2D partially visible natural objects. Is there any advantage at all in the recognition of occluded natural objects relative to their deleted counterparts? If so, how long must the visual system have access to the images to allow this difference to emerge? How early are differences evident in the EEG? The three experiments reported here show evidence that considerable recognition advantages exist for 2D stimuli containing partial objects which are appropriate for amodal completion, and these advantages can be seen with as little as 40 ms of masked image presentation. We also show that ERP differences between occluded and deleted stimuli occur as early as 130 ms after presentation, and suggest they may be correlated with amodal completion processes.

2. Methods

A total of 64 volunteers, 45 females and 19 males (aged 18–33 years, average 20.9) participated in the three experiments reported in this study. All participants had normal or corrected to normal vision. Volunteers were given either monetary compensation or university course credit for their participation. All participants gave informed consent, and the UC Davis Human Subjects IRB approved all studies.

2.1. Stimuli

The image stimuli used in all experiments were composed of a cutout object, a solid-colored background, and eight or nine solid-colored ovals (Fig. 1). There were three types of images. "Intact" images (used only in Experiment 2) consisted of a complete cutout object placed in front of ovals. "Occluded" images consisted of a cutout object placed behind ovals that occluded a specified percentage of the object's image pixels, such that the object was partially visible. "Deleted" images consisted of a similar partially visible object placed in front of ovals. One occluded and one deleted version of each source object was created for each experiment; in these the visible portion of the object was identical. The occluded and deleted (and intact) versions of each image also used the same ovals, but the ovals were moved to different locations and given new orientations in each. All images were centrally presented on a CRT monitor from a viewing distance of 75 cm, were 768 × 768 pixels in size, and subtended 15 × 15° of visual angle. Details of image creation can be found in the Appendix A.

Test images in Experiment 3 (Fig. 1) were constructed by rotating a cutout object ±45° and placing it at the center of a 768 × 768 uniform gray background.

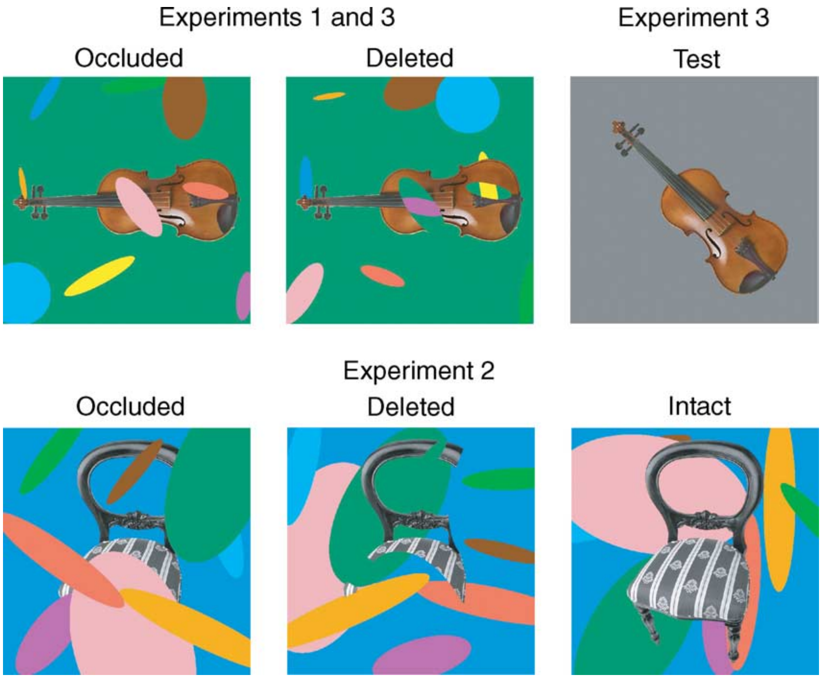

Fig. 1. Sample images. Sample images are shown for the three experiments conducted. The violin pictured here was missing 20% of its pixels due to occlusion/deletion. Images used in the recognition memory test in Experiment 3 consisted of the complete object rotated ±45° and placed against a uniform gray background. Experiment 1 used images with 15%, 30%, 45%, 60%, or 75% missing pixels. Experiment 3 used images with 20%, 40%, or 60% missing pixels. All images from Experiments 1 and 3 had nine ovals. All images from Experiment 2 had eight ovals, and occluded and deleted images were missing 60% of their pixels.

The cutout photographs were drawn from a wide range of objects obtained from a commercial image source (www.hemera.com) and were selected to have a long axis between about 400–700 pixels. Most images were single objects, but some consisted of a group of objects of the same type (e.g., dominoes).

2.2. Experimental procedure

2.2.1. Experiment 1

Cued-target behavioral experiment (Fig. 2A). A total of 40 volunteers participated in this experiment. Before each trial, subjects were presented with an entry-level (Jolicoeur, Gluck, & Kosslyn, 1984; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) word cue informing them of the target object on that trial, to be followed by a test image. The word cue remained on screen until the subjects pressed a button to initiate the presentation of the test image. After pressing the button, the cue word was replaced by a central fixation point for 400–900 ms. The test image then appeared, followed by a mask stimulus. Subjects were instructed to respond yes/no, as quickly as possible, whether the object(s) in the test image corresponded to the target cue. The stimulus onset asynchrony (SOA) of the mask stimulus was one of five possible values: 40, 100, 170, 270, or 400 ms. Mask stimuli were comprised of five successive 768 × 768 images of $1/f$ bandpass noise, ranging from low spatial frequency (2 cycles per image) to high spatial frequency (20 cycles per image). Each individual mask image was presented for 100 ms for a total of 500 ms of mask, and the order of the mask images was randomized on each trial. Following the mask stimulus, there was a 1700 ms delay before the appearance of the subsequent target word.

Test images were presented in either occluded or deleted conditions, with five possible amounts of missing information from each: 15%, 30%, 45%, 60%, and 75% missing pixels. Each source object was assigned to one of the five levels of missing information; a source object seen by one half of the subjects as a deleted object missing 45% of its image pixels would be seen by the other subjects as an occluded image missing the same 45% of its pixels. The presentation of a source object as an occluded or deleted image, and as a target or a nontarget, was counterbalanced across subjects. There were 100 total possible image conditions: 2 (Target/Nontarget) × 2 (Occluded/Deleted) × 5 (Mask SOA) × 5 (% missing pixels). Each of these image conditions was presented 10 times over the course of the experiment for a total of 1000 image presentations per subject. Image presentations were broken into 10 blocks of 100 images each, and the order of image conditions was randomized across the experiment.

2.2.2. Experiment 2

Free recognition EEG experiment (Fig. 2B). A total of 12 volunteers participated in this experiment. Other results from Experiment 2 are discussed separately (Johnson & Olshausen, 2005). During the experiment, the subject was required to identify an image (cue)

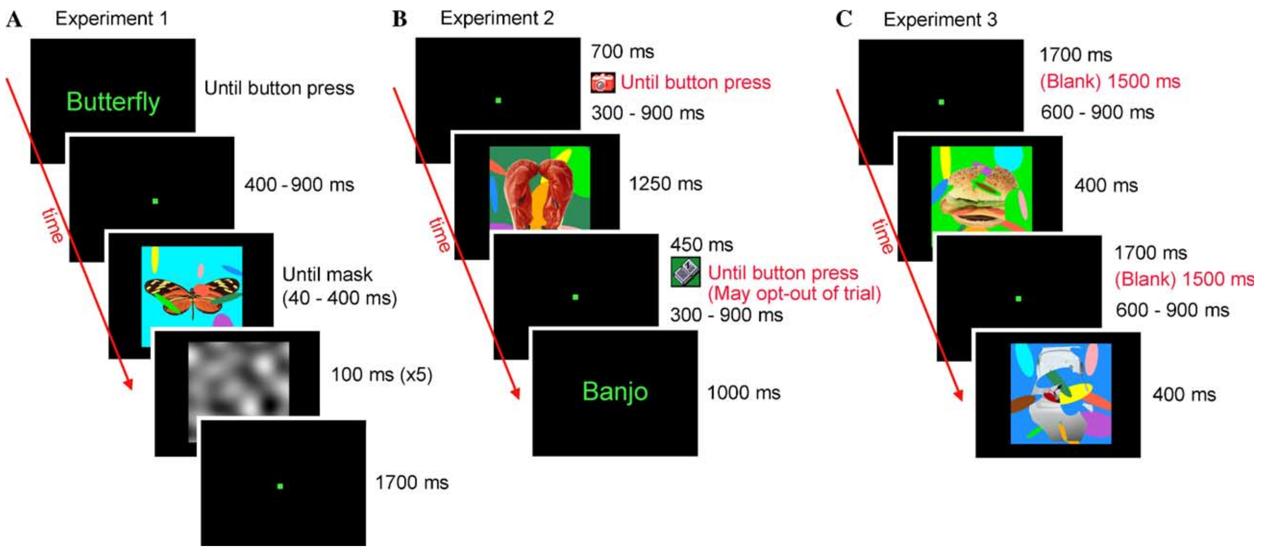

Fig. 2. Experimental sequences. In all panels, red text indicates designated blink periods to help avoid EEG artifact. (A) Experiment 1. The target cue word remained on screen until the subject pressed a button to initiate the trial. The image had a duration of 40, 100, 170, 270, or 400 ms, after which it was followed by a series of five 100 ms masks (only one of which is pictured). (B) Experiment 2. A camera icon (red) replaced the fixation point before each cue image, and remained on screen until a button press. A book icon (green) replaced the fixation point before each test word. Subjects could press one button to view the test word, but were instructed to press a second button to skip to the next camera icon if they were not confident of the identity of the cue image. (C) Experiment 3 was not self-paced. Two trials are depicted.

without any prior information as to what the object might be, and then to decide whether a subsequently presented test word corresponded to the image. Each trial was initiated with a button press. After a 300–900 ms fixation point the cue image was presented for 1250 ms, followed by a 450 ms fixation point. During the presentation of the cue image, subjects were instructed to maintain fixation at the center of the screen with the goal of reacquiring the fixation point when it reappeared without a saccade. After this second fixation point, the stimulus sequence paused and a screen appeared giving subjects the choice to continue or abandon the trial by a button press. Subjects were instructed to continue the trial if they felt they could name the object that was just presented, and to opt out of the trial if they felt they could not name the object. If the subject chose to continue the trial, there was another 300–900 ms fixation point followed by a 1000 ms test word presentation and a 700 ms fixation point before the next trial initiation screen. If the subject opted out of the trial, the sequence skipped directly to the next trial initiation screen. To avoid EEG artifact, subjects were asked to avoid blinking except while viewing an experimentally paused screen (trial continuation, trial initiation).

Intact, 60% occluded and 60% deleted images were used as cues. Subjects viewed 900 images in randomized order in 9 runs of 100 images each. Each occlusion condition was presented 300 times, and the appearance of a source object as occluded, deleted, or intact was counterbalanced across subjects.

In this study, we only present EEG data arising from the cue image phase and behavioral data from the opt-in/opt-out decision phase of Experiment 2. EEG and behavioral data arising from the test word phase of Experiment 2 is presented in Johnson and Olshausen (2005).

2.2.3. Experiment 3

Relative depth EEG experiment (Fig. 2C). A total of 12 volunteers participated in this experiment. During the first stage of the experiment, subjects were required to decide, as quickly as possible, whether the ovals in the image were in front of the object (i.e., occluded images) or behind the object (i.e., deleted images) without regard to the identity of the object itself, and to indicate their decision with a button press. The stimulus sequence in Experiment 3 was not self paced. Each image was presented for 400 ms, followed by a 1700 ms fixation point, a 1500 ms blank screen, a 600–900 ms fixation point, and then the next image. To avoid EEG artifact, subjects were asked to blink only during the 1500 ms blank screen.

Occluded and deleted images were created at each of three levels of missing information: 20%, 40%, and 60% missing pixels. Subjects viewed 900 images in randomized order in 9 runs of 100 images each. Each image type (e.g., 40% occluded) was presented a total of 150 times to each subject, and appearance of a source object in the occluded or deleted condition was counterbalanced across subjects.

Following this stage of the experiment the EEG equipment was removed and subjects were presented with a surprise memory test of the objects seen in the experiment. Test images were formed of complete cutout objects placed against a uniform gray background, with the objects rotated ±45° to reduce the possibility that the subjects could rely on remembering the positions of the object fragments rather than having encoded the complete object. Subjects were presented with 270 images—90 novel images and 180 repeat images. Thirty repeat images were randomly selected from each type of image (e.g., 20% deleted) used in the experiment. Each image remained on screen until the subject made a button press to indicate, yes/no, if they believed the image had been presented in the experiment. Subjects were not informed of the ratio of old/novel images in the memory test phase.

2.3. EEG recording and data analysis

Subjects were prepared for EEG recording using standard techniques. EEG was recorded at 19 scalp electrodes (FP1, FP2, F7, F3, FZ, F4, F8, T7, C3, CZ, C4, T8, P7, P3, PZ, P4, P8, O1, and O2) chosen from the International 10–20 set of electrode positions (American Electroencephalographic Society, 1994) and was referenced to the right mastoid. Horizontal and vertical electrooculogram were also recorded. Impedances at all electrodes were lowered to below 5 kΩ before beginning recording. The EEG signal was amplified with a high-pass cutoff of 100 Hz and a low-pass cutoff of 0.01 Hz before being digitized and recorded at 256 Hz.

Raw data were normalized, artifact rejected, and analyzed using Matlab software developed in-house. Software for the display of scalp topographies was developed by Scott Makeig (SCCN, UC San Diego). The trial-averaged EEG waveform—known as the event-related potential (ERP)—was computed separately for all conditions of interest. ERP waveforms were combined into grand averages over all subjects. ERP waveforms in Experiment 2 were computed using all cue images on which the subject subsequently made an "opt-in" decision. Those in Experiment 3 were computed using only correct trials. Both were timelocked to the presentation of the image. Before averaging, all data were artifact rejected on a trial-by-trial basis for eyeblink and on a channel-by-channel basis for drift, blocking and excessive alpha wave.

When assessing differences between two ERP waveforms, we have adopted a criterion of at least seven consecutive samples which are different at a $p < 0.10$ level (two sample t test for difference of means). Following the tabled values in Guthrie and Buchwald (1991), a run of 7 such samples results in a corrected p value of less than 0.039 ($N < 15$, autocorrelation estimate for each channel <0.7, 50 data samples [55–250 ms after presentation] assessed for differences).

3. Results

3.1. Varying amounts of missing pixels and mask onset time

In Experiment 1, we presented subjects with an object word and asked them to determine whether the object in

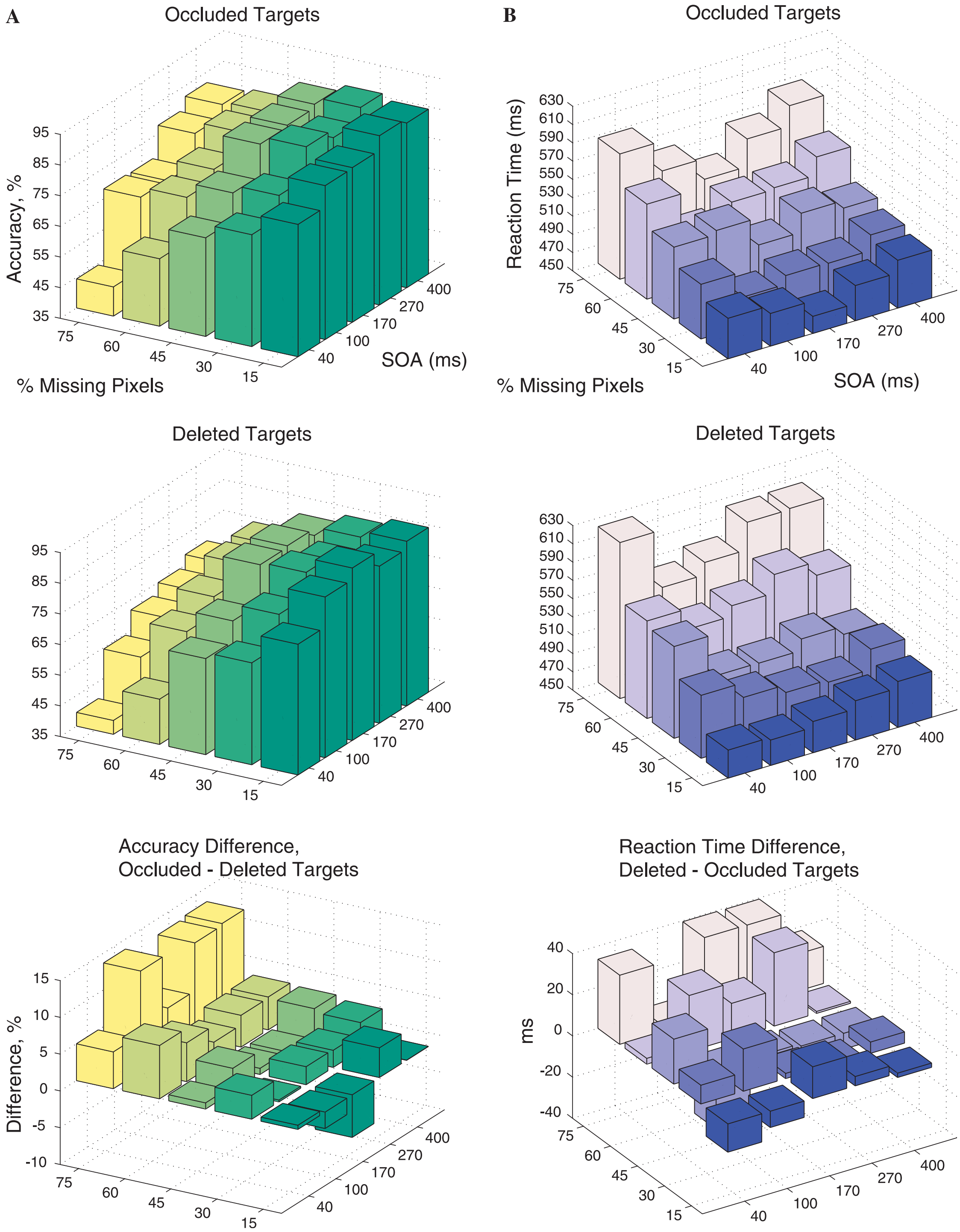

Fig. 3. Accuracy and reaction times in Experiment 1. (A) Average accuracy is shown at each level of missing pixel information and mask SOA for occluded images (top) and deleted images (middle). The difference between average accuracy (occluded minus deleted) is shown at bottom. All accuracies are calculated from target presentations only. (B) Average reaction time on correct target presentations are shown for occluded and deleted images as in (A). The difference between average reaction times (deleted minus occluded) is shown at bottom.

a subsequent picture was an exemplar of that word. Objects could be occluded or deleted, and were presented with various levels of missing information (15%, 30%, 45%, 60%, or 75% missing pixels, "PCT") and with varying mask stimulus onset asynchronies (40, 100, 170, 270, or 400 ms, "SOA"). Fig. 3A plots the overall accuracy (on target images only) for occluded images and deleted images at each level of PCT and SOA. On both occluded and deleted targets, accuracy decreases as the amount of missing pixels is increased, and as the duration of the image presentation is decreased. Occluded images appear to have a recognition advantage compared to their deleted counterparts (Fig. 3A, bottom). Differences between occluded and deleted targets are most pronounced in images missing 75% of their original pixels and are essentially absent in images missing only 15% of their original pixels.

To determine more precisely which of the image manipulations affected the subjects' accuracy, we ran a three-factor within-subjects analysis of variance (PCT, SOA, and occlusion type) on the data, the results of which are shown in Table 1. The ANOVA indicates that there are significant main effects of PCT ($F = 165.4$, df = 4, $p \ll 0.0001$), SOA ($F = 108.8$, df = 4, $p \ll 0.0001$), and image type ($F = 43.8$, df = 1, $p \ll 0.0001$) on accuracy as suggested by Fig. 3A. There are also two interaction effects. The first is an interaction between PCT and SOA ($F = 7.6$, df = 16, $p \ll 0.0001$) which can be seen in Fig. 3A. As the percentage of missing pixels increases and the mask SOA decreases, the subjects' accuracy, regardless of image type, decreased more than would be expected by a linear combination of the two. The second interaction, between PCT and image type ($F = 10.0$, df = 4, $p \ll 0.0001$), is replotted in Fig. 4. Here, data are collapsed across SOA, with occluded targets in blue and deleted targets in red. Error bars indicate the standard error. Accuracy on occluded and deleted images does not differ at 15% missing pixels, but begins to diverge as the amount of missing pixels increases, with occluded images showing a strong recognition advantage at 60 and 75% missing pixels. Overall, subjects correctly identified 78.2% of occluded targets and 74.5% of deleted targets across all PCT and SOA levels in Experiment 1 (Fig. 4, "mean"); the difference between these two means is statistically significant (main effect of image type as above).

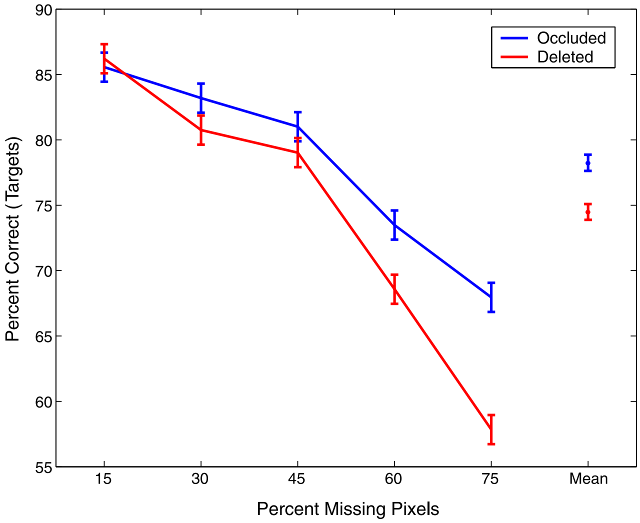

Fig. 4. Experiment 1 accuracy, averaged across SOA. Accuracy at each percent level of missing pixels is plotted in blue for occluded images and red for deleted images. Error bars indicate the standard error. Occluded images exhibit an advantage in recognition accuracy at high levels of missing information, but not at low levels. Overall mean accuracy for the experiment is plotted at right. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

Notably, there does not appear to be any interaction between mask SOA and image type ($F = 1.1$, df = 4, $p = 0.36$), which suggests that the occluded image recognition advantage is not increased when the image is presented for longer than 40 ms.

In addition to accuracy, we also recorded reaction times (RTs) for correct responses on target stimuli in Experiment 1, which are shown in Fig. 3B for occluded images and deleted images. Reaction times increase with increasing percentage of missing pixels, but exhibit a U-shaped function with respect to increasing mask SOA, with longer RTs at short and long SOAs and shorter RTs at intermediate SOAs. A three-factor within-subjects analysis of variance (PCT, SOA, and occlusion type) was also performed on the RT data and is

Table 1
Three-factor ANOVA, Experiment 1 accuracy

Source	SS	df	MS	F	Prob > F
Pct Miss. Info	138078.3	4	345189.5	165.4	<0.0001
SOA	126477.8	4	31619.5	108.8	<0.0001
Type	7065.4	1	7065.4	43.8	<0.0001
Pct × SOA	17255.2	16	1078.5	7.6	<0.0001
Pct × Type	6599.4	4	1649.8	10.0	<0.0001
SOA × Type	897.4	4	224.3	1.1	0.3625
Pct × SOA × Type	2123.1	16	132.7	0.77	0.7260
Error	484444.9	1950	5081.6		
Total	782941.5	1999			

Table 2
Three-factor ANOVA, Experiment 1 reaction times

Source	SS	df	MS	F	Prob > F
Pct Miss. Info	2226200.3	4	556550.1	86.2	<0.0001
SOA	296840.1	4	74210.0	10.9	<0.0001
Type	35005.5	1	35005.5	7.8	0.0081
Pct × SOA	136874.4	16	8554.7	2.1	0.0059
Pct × Type	57688.1	4	14417.0	3.3	0.0117
SOA × Type	10144.5	4	2536.1	0.56	0.6911
Pct × SOA × Type	80957.9	16	5059.9	1.3	0.2152
Error	25312734.0	1900	479851.6		
Total	28156444.8	1949			

shown in Table 2. In this ANOVA, the total degrees of freedom have been reduced because of one subject's failure to record any correct responses for a particular image type (60% Deleted, 40 ms SOA, 0/10 accuracy) leaving no basis for correct reaction time at that data point. The ANOVA indicates that there is a significant main effect of PCT ($F = 86.2$, df = 4, $p \ll 0.0001$) and SOA ($F = 10.9$, df = 4, $p \ll 0.0001$) as described above. The analysis indicates a much smaller main effect of occlusion type ($F = 7.8$, df = 1, $p = 0.008$). However, despite the effect of occlusion type, average reaction times for correct targets only differed by 7 ms (occluded = 524 ms, deleted = 531 ms).

Two interaction effects were also significant. One is an interaction between PCT and image type ($F = 3.3$, df = 4, $p = 0.01$) which indicates that the reaction time advantage for occluded images is stronger at high levels of missing information. The other is an interaction between PCT and SOA ($F = 2.1$, df = 16, $p = 0.006$).

As a closer look into the RT distributions, Fig. 5 shows a histogram of reaction time values for correct responses on target images for occluded (blue) and deleted (red) images in 20 ms wide bins, normalized to the total number of correct responses. The reaction time distributions for occluded and deleted images are virtually identical, suggesting that the type of image has only a small effect on reaction time, rather than producing two dissimilar distributions with coincidentally similar mean values. The finding of little difference between occluded and deleted reaction times is consistent with a similar finding in Johnson and Olshausen (2005) measured using 60% missing pixels and no masking.

Overall, Experiment 1 suggests that when information is missing from a scene, we are more accurate at identifying the objects in that scene when there is an occluder present than when there is not, and that this occlusion advantage becomes more significant as more information is removed from the scene. This occlusion advantage does not, however, depend upon the duration of presentation, even though mask SOAs varied from 40 to 400 ms. In contrast to accuracy results, Experiment 1 suggests that there is little difference in the time it takes to successfully identify a partially visible object based on the presence or absence of the occluder.

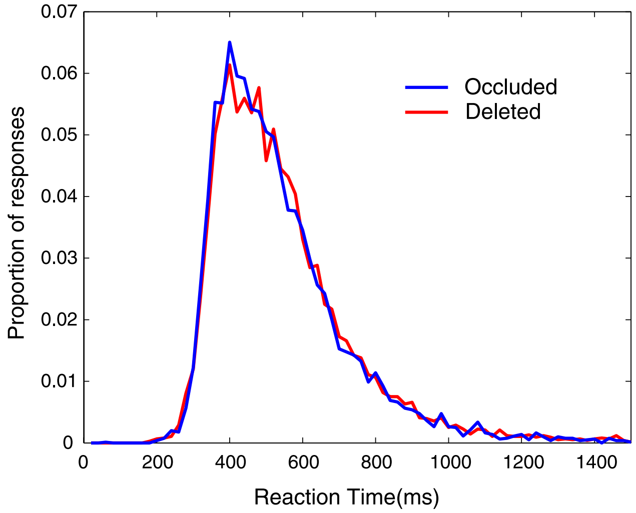

Fig. 5. Experiment 1 reaction time histograms. The distribution of reaction times in 20 ms wide bins is plotted in blue for occluded images and in red for deleted images. Reaction times over 1500 ms are excluded. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

3.2. Recognition without a target cue

In Experiment 1, subjects were first given a target cue which informed them with 50% probability of the identity of the upcoming object. Because of this cue, subjects may have been able to recognize some objects on the basis of features that would not normally be diagnostic for that object. For instance, if the target cue was 'pumpkin', the subjects might correctly respond 'yes' when viewing an otherwise unrecognizable object which they identified as being orange, even though without a target cue they might not be able to decide whether it was a pumpkin, a basketball, a tangerine, or something else entirely. Thus, in Experiment 2, subjects were asked to recognize intact, occluded, and deleted objects (the latter two at 60% missing pixels) without the advantage of a preceding target cue. Following the image presentation, subjects were instructed to decide whether they could identify the object or not. If they could ("opt-in"), they proceeded to view a test word and to respond whether it matched the object; if not ("opt-out"), they skipped the test word and moved on to the next trial.

The "opt-in/opt-out" phase of Experiment 2 allowed us to use the subject's own report to determine how well they were able to recognize the object in the image. Fig. 6 plots the opt-in rate for intact, occluded, and deleted images in Experiment 2, with error bars indicating the 95% confidence interval. Subjects chose to continue the trial on 99.1% of intact image presentations, compared to 92.0% of occluded image presentations and 80.0% of deleted image presentations. All pairwise comparisons are highly significant ($p < 10^{-16}$, z test for two independent proportions). Here, the subjects' own report suggests that recognition of uncued occluded objects is much better than recognition of uncued deleted objects.

3.3. Recognition memory of occluded and deleted objects

In Experiment 3, subjects performed a task (identifying the depth of the ovals relative to the object) which did not require the identification of the partially visible objects. At the conclusion of the experiment, subjects were given a surprise memory test in which they were asked to decide which objects had been used in the

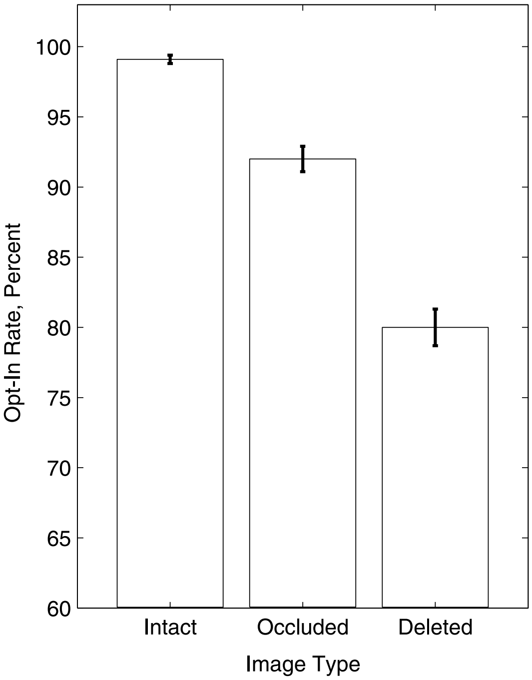

Fig. 6. Opt-in rates in Experiment 2. Subjects in Experiment 2 were asked to discontinue trials in which they did not confidently identify the object in the image cue. The percentage of trials on which the subject decided that they identified the object is shown for all three types of images, intact, occluded, and deleted. Occluded and deleted images were both missing 60% of their pixel information. Error bars indicate the 95% confidence interval of the mean of the opt-in rate.

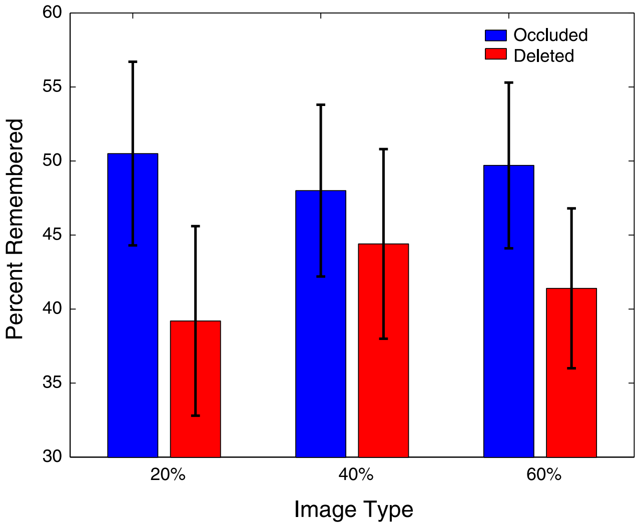

Fig. 7. Recognition memory of occluded and deleted images. Accuracy results for identification of previously presented images are plotted by type of image (occluded, blue; deleted, red) and by amount of pixels missing. Error bars indicate the standard error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

experiment. Fig. 7 shows the performance of the subjects on this task. Overall, subjects appear to show a recognition memory advantage for occluded images. Rather than exhibiting a systematic shift in performance based on the amount of missing pixels, performance across the 20%, 40%, and 60% levels remains relatively constant. For occluded images, accuracy was 50.6 (20%), 48.1 (40%), and 49.7 (60%), while for deleted images, accuracy was 39.2 (20%), 44.4 (40%), and 41.4 (60%). To make a statistical assessment of the recognition effect, we performed a two-factor within-subjects analysis of variance (Image type and PCT) on the data, shown in Table 3. The analysis indicates that there is a main effect of image type ($F = 51.5$, df = 1, $p \ll 0.0001$), confirming that occluded images were better remembered than deleted images. The analysis also indicates that there is no main effect of PCT ($F = 0.15$, df = 2, $p = 0.86$), suggesting that varying the amount of missing pixels between 20% and 60% does not affect subsequent recognition memory.

Table 3
Two-factor ANOVA (repeated measures), Experiment 3

Source	SS	df	MS	F	Prob > F
Type	0.109	1	0.109	51.5	<0.0001
Pct Miss. Info	0.002	2	0.001	0.15	0.8648
Type × Pct	0.018	2	0.009	0.59	0.5652
Error	2.83	55	0.234		
Total	2.959	71			

3.4. Early EEG differences between occluded and deleted images

Behavioral evidence from the three experiments described above suggests that given a particular set of object fragments, recognition is better when the missing portions of an object are replaced by an occluder rather than simply removed. However, our behavioral measures do not give a good indication as to how early in the visual processing stream the two types of images are treated differently. Thus, we recorded EEG during the presentation of the occluded and deleted images in an attempt to identify electrophysiological correlates of the behavioral differences.

Event-related potentials (ERPs) on seven occipital and parietal electrodes (P3, PZ, P4, P7, P8, O1, and O2) from Experiment 2 are shown in Fig. 8A, beginning from the presentation of the image. Here, the subject's task was to determine the name of the object; trials where the subject "opted-out" (claimed they did not recognize the object well enough to name it) are excluded from the analysis. ERPs evoked by 60% occluded images are in blue and ERPs from 60% deleted images are in red. For comparison, ERPs from intact images are shown in a lighter black line. The earliest differences between the occluded and deleted images arise over posterior parietal cortex (electrode P3) 129 ms after presentation ($p < 0.10$, seven consecutive samples, two sample t test for difference of means,

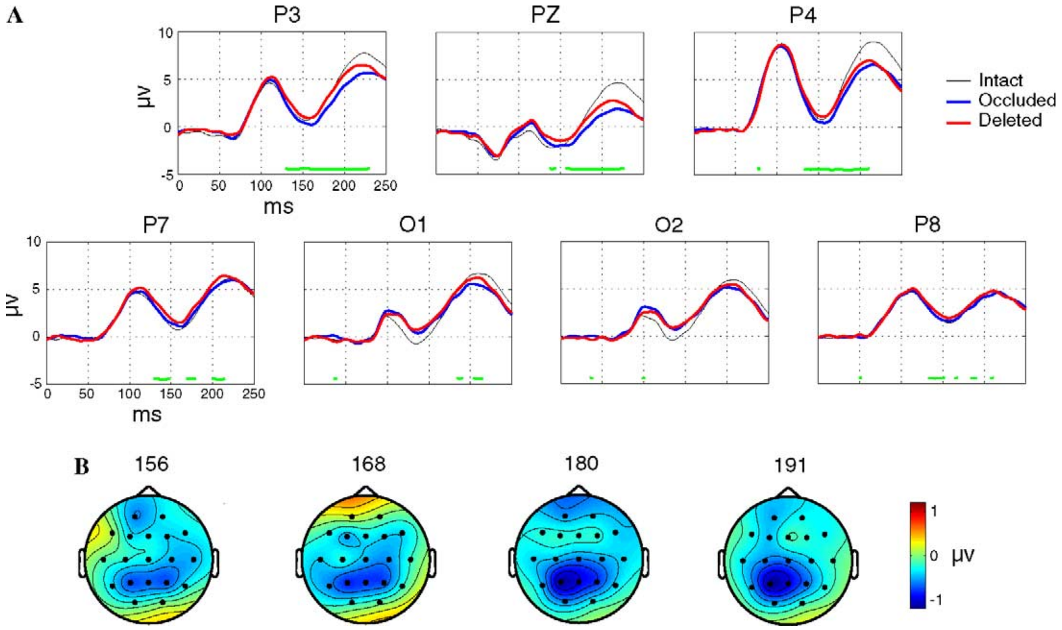

Fig. 8. Event-related potentials from Experiment 2. (A) ERPs recorded at seven occipital and parietal electrodes in Experiment 2 for occluded (blue), deleted (red), and intact (light black) images. Green line at bottom indicates timepoints which result in a statistically significant difference between occluded and deleted images ($p < 0.10$, two sample t test). (B) ERP scalp topographies for the occluded minus deleted difference at the four timepoints listed above the plots in milliseconds.

significant points plotted below in green). The statistical criterion is also reached on electrodes P4 (133 ms) and PZ (156 ms). The remaining electrodes do not exhibit differences that reach the statistical criterion. The overall structure of the intact waveform is not drastically different from the occluded or deleted wave-

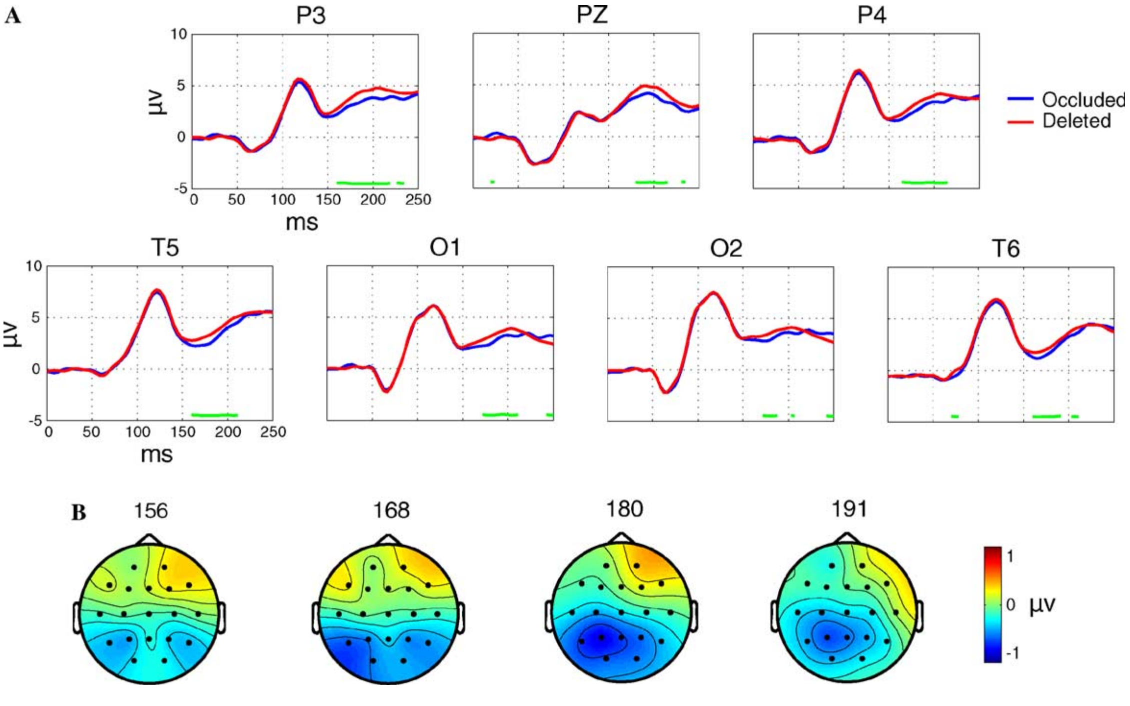

Fig. 9. Event-related potentials from Experiment 3. All scales are identical to those in Fig. 8. (A) ERPs recorded at seven occipital and parietal electrodes in Experiment 3 for occluded (blue) and deleted (red) images. Green line at bottom indicates timepoints which result in a statistically significant difference between occluded and deleted images ($p < 0.10$, two sample t test). (B) ERP scalp topographies for the occluded minus deleted difference at the four timepoints listed above the plots in milliseconds.

forms, showing that even a large manipulation of the object surface does not cause a drastic change in the visual processing of the scene.

Fig. 8B plots the scalp topography of the occluded/deleted difference at four timepoints. During the early portion of the difference, it is distributed mostly across three parietal electrodes (P3, PZ, and P4) without lateralization. By about 180 ms after presentation, the difference has become somewhat lateralized, being stronger over left hemisphere locations than right hemisphere locations.

ERPs were also recorded in Experiment 3, where subjects were presented with 20%, 40% or 60% occluded and deleted images and asked to perform a different task—to identify whether the ovals were in front of or behind the object without regard to the identity of the object itself. Results from these recordings are shown in Fig. 9A for the same electrodes as the data from Experiment 2. The waveforms for occluded and deleted images are collapsed across all percentages of missing information. The earliest occluded/deleted differences in Experiment 3 are about 30 ms later than the earliest corresponding differences in Experiment 2. Electrodes P3, P7, and P8 reach criterion 160 ms after presentation, with P4 (164 ms), O1 (172 ms), and PZ (180 ms) following. Differences on O2 beginning at 172 ms do not reach our statistical criterion of seven consecutive samples.

Fig. 9B plots the scalp topographies for the occluded/deleted difference in Experiment 3 at the same timepoints shown in Fig. 8B. Similar to Experiment 2, as the difference arises it is evenly distributed across hemispheres, but by about 180 ms after presentation it has become lateralized to the left. Although the late ($\geqslant$180 ms) portion of the difference has a similar topography in both experiments, this is not true for the early portions of the difference. In Experiment 3, instead of being focused mostly on P3, PZ, and P4, the difference is stronger at the inferior electrodes and appears to be somewhat weaker at the superior PZ. Coupled with a difference in the time required for a significant difference to arise, this change of scalp distribution suggests that different processes may be driving the early portion of the occluded/deleted differences in the two tasks.

Because there are potential low-level feature differences, particularly in the number of T-junctions, between occluded and deleted images, we further performed a manual count of the number of T-junctions in a subset of our images (50 occluded/deleted counterparts from each level of missing pixels from Experiment 3, selected randomly). The average number of T-junctions per image does differ slightly for 20% images (occluded = 13.3, deleted = 16.8) but does not differ for 40% (occluded = 19.8, deleted = 19.6) or 60% (occluded = deleted = 24.8) images, suggesting that differential responses of T-junction detectors are not a likely source of ERP differences between occluded and deleted images, especially for Experiment 2, where there would appear to be no difference in the number of T-junctions.

4. Discussion

Here we report three experiments which suggest that the recognition of partially visible natural objects is facilitated when the information missing from those objects is replaced by an occluder which explains its absence. We find behavioral differences between occluded and deleted images in three different tasks: a cued-target task, a free recognition task, and a recognition memory task. ERP differences between occluded images and their deleted counterparts are also visible relatively early—beginning as early as 130 ms after presentation in one task and between 160 and 180 ms in another. Because the salient difference between the two types of images lies in the inferred depth of the scene (which is not specified by the object fragments), these studies suggest that inferred depth across the scene plays a critical role in the recognition of partially visible objects, perhaps by guiding amodal completion processes prior to recognition.

4.1. Behavioral differences between occluded and deleted images

Perhaps the most convincing evidence that the visual system treats partially visible objects differently in the presence and absence of occluders comes from the free recognition task of Experiment 2, where subjects reported sufficient confidence of the object's identity to continue with the trial on 92% of occluded trials, compared to only 80% of deleted trials. This difference is seen despite the fact that the subjects had a relatively long 1250 ms, unmasked interval in which to view the images, suggesting that relatively prolonged processing (within a single fixation) is not of particular use in bringing performance on deleted objects into register with performance on occluded objects. One potential concern with this experiment is that subjects may exhibit a bias in their decisions to continue the trial, with a lower opt-in rate on deleted images. However, behavioral measures of the follow-up presentation of the target word (Johnson & Olshausen, 2005) suggest the opposite: subjects were less accurate following deleted cue images than occluded cue images even though the test words were identical in either case, suggesting that subjects' opt-in criteria were less stringent, or their identifications less accurate, for deleted cue images than for occluded cue images.

One complicating factor in any study of the recognition of partially visible objects is the phenomenon of recognition from partial information (RPI), which is proposed to occur separately from amodal completion processes (Kellman et al., 2001). There seem to be two routes for recognition from partial information. The first is through diagnostic features—the presence of a tire and a bumper are sufficient to alert us to the location of a car, even if the remainder of the vehicle is occluded. The other route is contextual, where otherwise ambiguous partial information may be combined with Bayesian priors to "recognize" an object on the basis of features that would otherwise not be sufficient for such recognition, as in the example of an ambiguous orange-colored patch being identified as a "target" pumpkin. In a free report experiment, only diagnostic features can be reasonably expected to help in recognition, but in a cued-target experiment (like Experiment 1), the subject has a strong Bayesian prior for the presence of a particular target and contextual knowledge may also play a role. In either case, RPI does not require that the boundaries or surfaces of objects are interpolated, so effects of RPI should be equivalent for either occluded or deleted images. In the cued-target task, we can expect some number of source objects to be recognized based on the context of the target cue which would otherwise have been more easily recognizable in the occluded case than in the deleted case. As such, we would expect that the advantage conferred to occluded partially visible objects would be greater in the free report task than in the cued-target task of Experiment 1. Indeed, this is what we see. In comparison to a 12% opt-in advantage on 60% images in the free report task, we see less than a 5% accuracy advantage in the 60% images from the cued-target task. Nonetheless, the difference between occluded and deleted image recognition is statistically significant in the cued-target experiment as well.

Results from the recognition memory test are intriguing because they demonstrate that a recognition advantage exists for occluded versions of images even when the identity of the partial objects is task-irrelevant and is presumably largely ignored at the time of presentation. This result suggests that the processes which distinguish the two types of images are an automatic part of visual processing rather than an optional step performed for the sake of object recognition (though EEG results suggest they may be modified by recognition effort).

4.2. EEG differences between occluded and deleted images

In addition to the behavioral results, we also find early EEG differences between occluded and deleted images at parietal and occipital electrodes. The occipitoparietal EEG difference, which becomes evident between 130 and 160 ms after presentation, is seen in both tasks on which we recorded EEG. The early portion of this signal has a different time of onset and scalp distribution in the two tasks (free recognition and oval depth), suggesting that the disparate demands of the tasks have top-down effects on early visual processing. One possible source of these differences is attention to the object form itself—in the oval depth task, subjects are not required to attend to object form, but in the recognition memory task the subject must explicitly determine the identity of the object, and object form is crucial. It may be that attention to object form facilitates an early, depth-related visual process which is not facilitated during passive viewing. By 180 ms after presentation the difference signals become quite similar, which may indicate that there are multiple processes involved, the latter of which is independent of task demands.

ERP studies investigating the timecourse of object closure using deleted line drawings (Doniger et al., 2000) have demonstrated an occipitotemporal signal, called the "negativity of closure", which is believed to correlate with neural activity arising from the formation of a unified percept of the line object. This signal onsets about 230 ms after presentation. Although the negativity of closure could be related to amodal completion, the ERP signal that we find occurs somewhat earlier, beginning before 150 ms. Even so, an earlier result is not necessarily surprising considering that the visual system is specialized to deal in natural images rather than line drawings and other simple stimuli. For example, in one recent dual-task study, it was shown that performance on the secondary (peripheral) discrimination task was over 80% when the task was performed on natural images and under 60% when the task was performed on letters (Li, VanRullen, Koch, & Perona, 2002). In the light of evidence that the recognition of natural objects may come about as quickly as 150 ms after presentation (Johnson & Olshausen, 2003; VanRullen & Thorpe, 2001) the speed of line drawing processing may simply lag behind that of natural images.

On the other hand, it could be that the differences we see between occluded and deleted images, which arise nearly 100 ms before the negativity of closure, are more consistent in temporal onset with known EEG effects of collinearity which arise about 100 ms after presentation (Khoe, Freeman, Woldorff, & Mangun, 2004; Norcia, Sampath, Hou, & Pettet, 2005). Certainly the connection of collinear line segments should be found among the early steps of any completion process, and such activity could underlie the ERP differences we find. However, because our occluded and deleted stimuli both contain the same collinear segments at the edge of the object, any change in the effort or efficacy of connecting these collinear segments would seem to reflect not only a bottom-up analysis of the object edges but an analysis of the remainder of the image as well.

4.3. Alternative explanations to amodal completion

We have suggested that the differences that we see between the recognition of occluded and deleted objects may be driven by amodal completion processes that are differentially active in the two cases, but there are potentially other image features that could bring about the same result. For instance, the location of a point of occlusion is generally specified by the presence of a T-junction. One possible explanation of the ERP differences we find between occluded and deleted images is that we are seeing a reflection of the presence of these local cues in the differential activity of simple T-junction detectors. However, because the object can occlude the ovals as well as the ovals can occlude the object, there does not seem to be any difference between the number of T-junctions found in occluded and deleted images (although in each case the T-junctions are in different locations) except for a small difference (about 20% of the total) between images with small amounts of missing pixels. This would suggest that the activity of putative T-junction detectors themselves would be unlikely to be the source of our ERP differences.

The two types of images are also likely to differ in the presence of a similar low-level image feature, the L-junction. An L-junction occurs when there is a change in the surface curvature of a foreground object, and it follows the veridical boundary of that object. Unlike the T-junction, the definition of an L-junction is somewhat arbitrary, so we have not attempted to quantify the number of L-junctions in our images. It is clear, however, that the number of L-junctions is greater in the deleted images, because deletion is the only manipulation we have performed that can produce them. Thus, it is possible that our ERP differences may be driven by differential activity of L-junction detectors rather than amodal completion processes.

Because L-junctions are a strong indicator that amodal completion should not occur, and the presence of collinear T-junctions could be a simple signal that amodal completion is appropriate, both types of junction could act as initial inputs to completion processes without the benefit of a complex representation of object depth. While some have argued that in natural images features such as L- and T-junctions may be too ambiguous at the local level to be accurately detected without access to a more global representation of the scene (McDermott, 2004; Olshausen & Field, 2005), further experiments are necessary to determine whether local structure or global structure is responsible for the differential analysis, and differential ERP signature, of occluded and deleted images.

In addition to the question of how low-level features might contribute, we cannot directly discount the possibility that the effects we see, both behavioral and ERP, are the result of neural competition rather than a reflection of amodal processing. For example, if the visual system explicitly represents the "invisible" ovals which are "occluding" the deleted objects, the larger number of oval representations could result in an object competition effect reducing the amount of resources devoted to identifying the target object.

4.4. Implications for models of recognition

On top of the differences we have seen in this study between the recognition of occluded and deleted images, it is interesting to note two differences which we do not see. For one, mask SOAs as fast as 40 ms have no significant effect on the relative recognition of occluded and deleted images. This suggests a few possibilities. One is that recognition of these objects proceeds in a feedforward fashion, but that there is some mechanism within the processing pathway that accounts for depth relations of the scene before the process of object recognition. Another is that there may be a very fast mechanism which establishes, via feedback, a surface-based representation of the scene at the very earliest levels of the visual system. Although neural correlates of amodal completion can be seen as low as V1 in the macaque (Lee, 2003), it is unclear that they develop quickly enough to avoid disruption by masking.

Second, we do not see a large difference in target reaction times between occluded and deleted images, despite previous evidence for both a deletion advantage (Brown & Koch, 1993) and for an occlusion advantage (Gerbino & Salmaso, 1987) in RT in other tasks. Our data here agree with RT data for a previous occlusion/deletion cued-target task (Johnson & Olshausen, 2005)—apparently in a cued-target task using natural images, RT differences, if present, are minor. Because reaction times are similar for both occluded and deleted images, it would suggest that if feedback is being used to establish a surface-based representation, this feedback is obligatory; that is, because there is no delay for occluded images (in fact there is a slight delay for deleted images), we should assume that the same feedback processes occur for all types of images, and must be accomplished before objects in the scene can be recognized at higher levels. A feedforward mechanism, on the other hand, would be reasonably expected to result in similar reaction times for occluded and deleted images.

Perhaps most importantly, these results suggest that any complete model of the visual processing leading to object recognition must necessarily account for early amodal completion effects, presumably involving the determination of border ownership at the points of discontinuity of partially visible objects. How this could be accomplished at the neural level remains to be seen. Local features (e.g., T-junctions) can be used to some benefit, but in general, border ownership, which underlies a global representation of surfaces, is a difficult property to incorporate into feedforward models of visual processing (Riesenhuber & Poggio, 2002). Current models that establish a full-blown surface-based representation of the scene and assign border ownership and establish depth relations in the early stages of processing do so with feedback.

Surface-based models, with their early depth representations, are well-suited to account for an amodal completion related recognition advantage but are hard pressed to explain similar reaction times for occluded and deleted images. Amodal completion in a feedback system should take time to implement; unless this feedback delay is obligatory, objects which do not get completed should be available in their final form to recognition processes sooner. Furthermore, the failure of masks with onsets as short as 40 ms to disrupt the occlusion advantage also casts doubt on the involvement of feedback processes. Feedforward models on the other hand, especially those which suggest minimal neuronal integration times (e.g., VanRullen & Thorpe, 2002), are better equipped to explain identical reaction time distributions and mask immunity, but have difficulty establishing the depth relations that allow otherwise target-irrelevant visual information at the site of occluders/deleters to affect identification accuracy. While it appears clear that depth relations and amodal completion play an important role in the pre-recognition visual processing of images, future work is necessary to determine exactly how the visual system implements these processes while parsing and identifying objects in cluttered natural scenes.

Acknowledgments

We thank Dr. Cindy Erickson for advice on statistical methods. This work was supported by NIH Grant #MH57921.

Appendix A. Image creation details

This section describes the creation of the occluded, deleted, and intact versions of our stimuli in detail. The occluded, deleted, and intact (Experiment 2 only) images corresponding to a single cutout object were made concomitantly, and we refer to this set of images as an image series.

Because the visible portion of the occluded object was also used in the corresponding deleted image, occluded images were created first. A 768 × 768 pixel background was divided into nine 256 × 256 pixel regions. The nine centers of these regions served as starting points to anchor the centers of the ovals. In the case of Experiment 2 where there were only 8 ovals, the central region was not used. Each oval anchor was jittered from its original location by a randomized number of pixels in the x and y directions, independently. The size of the random jitter was equal to a constant multiplied by the output of the "randn" function in Matlab, which draws a random number from a distribution with a mean of zero and a standard deviation and variance of 1. In Experiments 1 and 3, the jitter constant was 768/9 pixels, in Experiment 2, the jitter constant was 768/11 pixels. Each *anchor* was also assigned a random orientation (evenly distributed between 0 and 180 deg, continuous) which served as the orientation of the long axis of the oval which would be placed at that anchor.

The background and each of the ovals were assigned a solid color from a list of 10 perceptually distinct colors (randomly, without replacement for each image series). Each oval was randomly assigned an aspect ratio (between 1.0 and 6.0, continuous) and an initial long axis length between 30 and 80 pixels. The total pixel area of the ovals was calculated and each long axis was multiplied by the same scaling factor to bring the total area of the ovals to approximately PCT (where PCT is the intended percentage of occluded or deleted pixels) of the pixel area of the 768 × 768 background. Each oval was randomly assigned to an anchor, which also determined its orientation. (Because the ovals could extend outside the border of the background, the total area covered by the ovals at this point would usually be somewhat less than PCT.)

Before placing the ovals on the background, the cutout photograph was placed in the center of the background. The ovals were now placed on top of the cutout + background image, with the depth order of the ovals randomized. The percentage of the cutout object which was covered by ovals was calculated, and an iterative process was used to find a single scaling factor for long axis length which would allow PCT ±1% of the pixel area of the cutout to be covered by ovals. The resulting image was the occluded image.

The intact and deleted images were made next, separately. Anchors were returned to the centers of the nine (or eight) regions and re-jittered according to the same process described above. A new random orientation was selected for each anchor point, but this orientation was constrained to not lie within 10 deg of the orientation of the same anchor in the occluded image (intact and deleted orientations were not constrained with respect to each other). The ovals used in the final occluded image were also used in the intact and deleted images with the same color, aspect ratio, and size (major axis length). Each oval was randomly assigned to a new anchor, constrained to be a different anchor from the one it was assigned to in the occluded image (again, intact and deleted anchor assignments were not constrained with respect to each other) and the ovals were placed, with random depth assignments, atop the background. For the intact images, the complete cutout object was now placed on top of the background/oval image. For the deleted images, the *unoccluded* portion of the cutout object from the occluded image was extracted and placed on top of the background/oval image so that the same object information was present in both the occluded and deleted images.

References

American Electroencephalographic Society (1994). Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 11*, 111–113.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*, 115–147.

Bregman, A. L. (1981). Asking the "what for" question in auditory perception. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual Organization* (pp. 99–118). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, J. M., & Koch, C. (2000). Influences of occlusion, color, and luminance on the perception of fragmented pictures. *Perceptual and Motor Skills, 90*, 1033–1044.

Brown, J. M., & Koch, C. (1993). Influences of closure, occlusion, and size on the perception of fragmented pictures. *Perception & Psychophysics, 53*, 436–442.

Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., et al. (2000). Activation timecourse of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes. *Journal of Cognitive Neuroscience, 12*, 615–621.

Fukushima, K. (2005). Restoring partly occluded patterns: A neural network model. *Neural Networks, 18*, 33–43.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36*, 193–202.

Gerbino, W., & Salmaso, D. (1987). The effect of amodal completion on visual matching. *Acta Psychologica, 65*, 25–46.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28*, 240–244.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision, 5*, 299–312.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision, 3*, 499–512.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognitive Psychology, 16*, 243–275.

Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception.* New York: Praeger.

Kellman, P. J. (2003). Interpolation processes in the visual perception of objects. *Neural Networks, 16*, 915–923.

Kellman, P. J., Guttman, S. E., & Wickens, T. D. (2001). Geometric and neural models of object perception. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 183–245). Oxford, UK: Elsevier Science Publishers.

Kellman, P. J., Yin, C., & Shipley, T. F. (1998). A common mechanism for illusory and occluded object completion. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 859–869.

Khoe, W., Freeman, E., Woldorff, M. G., & Mangun, G. R. (2004). Electrophysiological correlates of lateral interactions in human visual cortex. *Vision Research, 44*, 1659–1673.

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology, 97*, 121–139.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20*, 1434–1448.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences USA, 99*, 9596–9601.

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception, 33*, 1101–1127.

Mel, B. W. (1997). SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation, 9*, 777–804.

Michotte, A., Thinès, G., & Crabbé, G. (1964). Amodal completion of perceptual structures. In: G. Thinès, A. Costall, & G. Butterworth (Eds.), Michotte's experimental phenomenology of perception (pp. 140–167, published 1991). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: a critical link between lower-level and higher level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *Vision: An invitation to cognitive science* (pp. 1–70). Cambridge, MA: MIT Press.

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.

Norcia, A. M., Sampath, V., Hou, C., & Pettet, M. W. (2005). Experience-expectant development of contour integration mechanisms in human visual cortex. *Journal of Vision, 5*, 116–130.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding V1? *Neural Computation, 17*, 1665–1699.

Peterson, M. A., & Gibson, B. S. (1994). Must shape recognition follow figure-ground organization? An assumption in peril. *Psychological Science, 5*, 253–259.

Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research, 38*, 2489–2505.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology, 12*, 162–168.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Singh, M. (2004). Modal and amodal completion generate different shapes. *Psychological Science, 15*, 454–459.

Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature, 401*, 269–272.

Ullman, S., & Bart, E. (2004). Recognition invariance obtained by extended and invariant features. *Neural Networks, 17*, 833–848.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*, 682–687.

VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research, 42*, 2593–2615.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13*, 454–461.

van Tuijl, H. F. J. M. (1975). A new visual illusion: Neonlike color spreading and complementary color induction between subjective contours. *Acta Psychologica, 39*, 441–445.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260–1262.