



Visual Grouping in Human Parietal Cortex

Citation

Xu, Yaoda and Marvin M. Chun. 2007. Visual grouping in human parietal cortex. Proceedings of the National Academy of Sciences of the United States of America 104(47): 18766-18771.

Published Version

<http://dx.doi.org/10.1073/pnas.0705618104>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3296249>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Visual grouping in human parietal cortex

Yaoda Xu* and Marvin M. Chun

Department of Psychology, Yale University, New Haven, CT 06510

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved September 27, 2007 (received for review June 15, 2007)

To efficiently extract visual information from complex visual scenes to guide behavior and thought, visual input needs to be organized into discrete units that can be selectively attended and processed. One important such selection unit is visual objects. A crucial factor determining object-based selection is the grouping between visual elements. Although human lesion data have pointed to the importance of the parietal cortex in object-based representations, our understanding of these parietal mechanisms in normal human observers remains largely incomplete. Here we show that grouped shapes elicited lower functional MRI (fMRI) responses than ungrouped shapes in inferior intraparietal sulcus (IPS) even when grouping was task-irrelevant. This relative ease of representing grouped shapes allowed more shape information to be passed onto later stages of visual processing, such as information storage in superior IPS, and may explain why grouped visual elements are easier to perceive than ungrouped ones after parietal brain lesions. These results are discussed within a neural object file framework, which argues for distinctive neural mechanisms supporting object individuation and identification in visual perception.

brain imaging | object processing | visual attention | working memory | fMRI

To comprehend the continuous influx of visual information in everyday life, visual input needs to be organized into discrete units that can be selectively attended and processed. One important selection unit is visual objects, whose formation has been shown to profoundly impact subsequent visual processing (1). An important factor determining object-based selection is the grouping between visual elements by various gestalt principles (2, 3), such as connectedness and closure (4–7). Such grouping cues shape conscious visual perception. For example, after unilateral parietal lesions, observers' ability to perceive the presence of two objects, one on each side of the space, was greatly improved by connecting the two objects with a bar forming one big object with two parts instead of two separated objects (8–10). Likewise, after bilateral parietal lesions that result in Balint's syndrome (11), patients could still perceive a single complex object, but their ability to perceive the presence of multiple visual objects is severely impaired (11–13). Such lesion data point to the importance of the parietal cortex in object-based representations, but our understanding of these parietal object grouping and selection mechanisms in normal observers remains largely incomplete.

Parietal brain responses have been associated with the number of visual objects actively represented in the mind, including those from the inferior intraparietal sulcus (IPS), which participates in attention-related processing (14–16), and the superior IPS, whose response correlates with the number of objects successfully stored in visual short-term memory (VSTM) (17–20, †). For example, when observers retain variable numbers of object shapes in VSTM, inferior IPS functional MRI (fMRI) activations increased linearly with display set size and plateaued at set size four regardless of object complexity. Activations from the superior IPS also increased linearly with display set size, but plateaued at the maximal number of objects held in VSTM as determined by object complexity (20). Thus, a fixed number of objects are first represented and selected by the inferior IPS by their spatial locations; depending on their complexity, a subset of these selected objects are then retained in VSTM with great detail by the superior IPS. Activities in these

parietal mechanisms thus reflect the number of discrete visual objects represented in the mind at different stages of visual processing.

Such neural representations of visual objects provide us with an opportunity to examine the impact of visual grouping on these parietal mechanisms during object-based representations. Specifically, by manipulating visual grouping and measuring the number of discrete objects present in the different parietal areas, we can infer how grouping affects object representations in these areas. In the present study, in two experiments, we used a region of interest (ROI) approach (15, 21) and measured averaged fMRI responses from functionally defined inferior and superior parietal ROIs in a VSTM change-detection task (17, 22). In addition to these parietal ROIs, we also examined grouping-related fMRI responses in the lateral occipital complex (LOC), an area involved in object shape processing (15, 21, 23, 24).

Results

Experiment 1. In the first experiment, we asked observers to retain the identities of either two or three shapes in VSTM (Fig. 1A). We presented four filled black rectangles around the central fixation in the background. When the display contained two shapes, each appeared on a separate black rectangle. When the display contained three shapes, they could either be grouped, appearing on two black rectangles, or ungrouped, appearing on three black rectangles (Fig. 1A). The placement on the black rectangles determined how the target shapes were grouped, although it was irrelevant to the VSTM task. To isolate an object-based grouping representation independent of a space-based one, we matched the spatial distance between the shapes in the grouped and ungrouped displays.

Because inferior IPS response increases linearly with the number of objects represented (up to four) (20), if grouping reduces the number of discrete objects present, its response should be low for two ungrouped shapes, intermediate for three grouped shapes, and high for three ungrouped shapes. Consequently, the relative ease of representing the grouped shapes would allow more shape information to be passed onto later stages of visual processing, resulting in better behavioral VSTM performance and higher responses in the superior IPS that encodes detailed shape features.

As predicted, grouping allowed more shapes to be stored in VSTM, resulting in an object-based benefit in behavioral performance as reported previously (6, 7, 22) (Fig. 2A). We transformed behavioral response accuracies to VSTM capacity estimate K (25), and the results are plotted in Fig. 2A. Behavioral reaction times (RTs) were 679 (SE = 34) ms, 749 (SE = 30) ms, and 740 (SE =

Author contributions: Y.X. and M.M.C. designed research; Y.X. performed research; Y.X. analyzed data; and Y.X. and M.M.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: fMRI, functional MRI; IPS, intraparietal sulcus; fMRI, functional MRI; LOC, lateral occipital complex; ROI, region of interest; RT, reaction time; VSTM, visual short-term memory.

*To whom correspondence should be addressed. E-mail: yaoda.xu@yale.edu.

†Although the IPS region reported by Todd and Marois (17, 18) encompassed both the inferior and superior IPS, the mean Talairach coordinates reported for this brain region were located at the superior IPS. Moreover, when Xu and Chun (20) manipulated object complexity, only the superior IPS response correlated with VSTM capacity.

© 2007 by The National Academy of Sciences of the USA

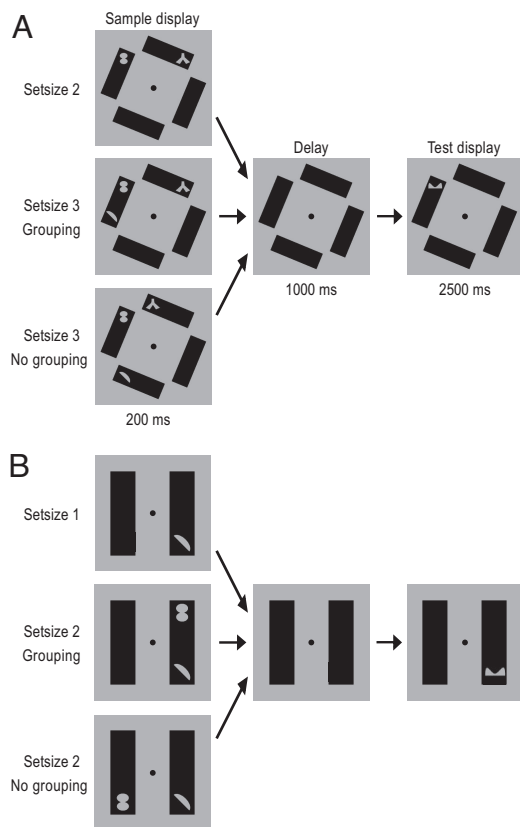


Fig. 1. Example trials used in the two experiments. (A) Example trials from Experiment 1. Observers viewed either two or three gray shapes appearing briefly on the background black rectangles and, after a brief delay, detected a shape change at the probed location. (B) Example trials from Experiment 2. The displays contained either one or two gray shapes appearing on the background black rectangles and observers performed a similar change detection task as in A. The display background was always present throughout a given experiment and determined grouping between the shapes, although it was task-irrelevant. In both experiments, the orientation of the illustrated display background was rotated 90° in half of the trials to control for the differences associated with a particular display background orientation.

36) ms, respectively, for two-shape, three grouped-shape, and three ungrouped-shape displays. Overall, K was lower and RT was faster for the two-shape than for either of the three-shape displays [$F(1, 11) = 329.86, P < 0.001$ for K difference; $F(1, 11) = 22.02, P < 0.01$ for RT difference between two and three grouped shapes; $F(1, 11) = 31.23, P < 0.001$ for K difference; $F(1, 11) = 17.04, P < 0.01$ for RT difference between two and three ungrouped shapes]. Critically, K was lower for the three ungrouped than for the three grouped shapes [$F(1, 11) = 6.69, P < 0.05$], although RT did not differ between the two ($F < 1$).

Grouping also significantly modulated fMRI responses within independently defined inferior and superior IPS ROIs. fMRI activations in both ROIs were lower for the two-shape than for either of the three-shape displays (Fig. 2A) [$F(1, 11) = 19.52, P < 0.01$ in the inferior IPS; $F(1, 11) = 23.94, P < 0.001$ in the superior IPS for the difference between two and three grouped shapes; $F(1, 11) = 41.14, P < 0.001$ in the inferior IPS; $F(1, 11) = 22.61, P < 0.01$ in the superior IPS for the difference between two and three ungrouped shapes]. Critically, for three-shape displays, mean activation was lower for the grouped than for the ungrouped shapes in the inferior IPS [$F(1, 11) = 5.64, P < 0.05$], but reversed direction in the superior IPS [$F(1, 11) = 17.13, P < 0.01$]. The interaction between the two three-shape displays and the two parietal ROIs was significant [$F(1, 11) = 14.11, P < 0.01$]. These results supported

our prediction that grouping reduces the number of discrete objects represented in inferior IPS and increases information storage in superior IPS. Although the object-selective ventral visual area LOC was sensitive to the number of shapes present [$F(1, 11) = 19.07, P < 0.01$ for the difference between two and three grouped shapes; $F(1, 11) = 8.76, P < 0.05$ for the difference between two and three ungrouped shapes], its response was not modulated by grouping ($F < 1$ for the difference between the two three-shape displays).

These results indicate that grouping changed the number of discrete objects represented in the inferior IPS, such that grouped shapes elicited a lower response than ungrouped shapes. This relative ease of representing and selecting grouped shapes in the inferior IPS allowed more shape information to be passed onto later stages of visual processing. As a result, more shape information is stored in VSTM as reflected in both the behavioral change detection accuracies and superior IPS responses, which correlate with VSTM capacity (17, 20).

Behavioral studies reported that grouping could influence visual performance even when observers are unaware of the presence of such grouping (26, 28). Consistent with these observations, in our study, many of the observers reported after the experiment that they were unaware of the different kinds of groupings present among the shapes and that they simply ignored the background bars during the change-detection task. This finding suggests that grouping between visual elements may be represented in the inferior IPS even when observers are not explicitly aware of it. However, further systematic investigations are needed to verify this observation.

Experiment 2. Behavioral performance in Experiment 1 was lower for the three ungrouped than for the three grouped shapes. This result indicated that the ungrouped shapes were harder to encode than the grouped ones, although observers might not have been explicitly aware of it. Thus, it was possible that observers might have used more effort to encode the ungrouped shapes, resulting in greater inferior IPS responses unrelated to the representation of grouping. Although the opposite response patterns in the superior IPS and the absence of grouping-related responses in the LOC rule out a simplistic effort account, to further rule out this explanation, we conducted a second experiment that repeated the design of Experiment 1, but reduced the display set size by one (Fig. 1B). Because maximal VSTM capacity for these shapes was about three (20), we could match task demands between the grouped and ungrouped displays by asking observers to remember only two shapes, which were easily within their performance capacity.

Specifically, we asked observers to retain the identities of either one or two shapes in VSTM (Fig. 1B). We presented two filled black rectangles on either side of the central fixation, similar to those used in a previous behavioral study (4). When the display contained two shapes, they could either be grouped, appearing on two ends of the same rectangle, or ungrouped, appearing on the ends of separate rectangles the same distance apart (Fig. 1B). If the inferior IPS indeed represents visual grouping independent of task difficulty, its response should be low for one shape, intermediate for two grouped shapes, and high for three ungrouped shapes. Meanwhile, because task demand in this experiment was low and within observers' performance capacity, behavioral change-detection accuracies and superior IPS responses should not differ between the grouped and ungrouped shapes.

Grouping did not affect behavior when the task demand was below capacity. Behavioral K measures are plotted in Fig. 2B. RTs were 627 (SE = 40) ms, 697 (SE = 38) ms, and 688 (SE = 44) ms, respectively, for one-shape, two grouped-shape, and two ungrouped-shape displays. There was no difference in either K or RT between grouped and ungrouped two-shape displays (F s < 1), although K was lower and RT was faster for the one- than for either of the two-shape displays [$F(1, 7) = 25.16, P < 0.01$ for K difference; $F(1, 7) = 31.82, P < 0.01$ for RT difference between one and two grouped shapes; $F(1, 7) = 36.74, P < 0.01$ for K difference;

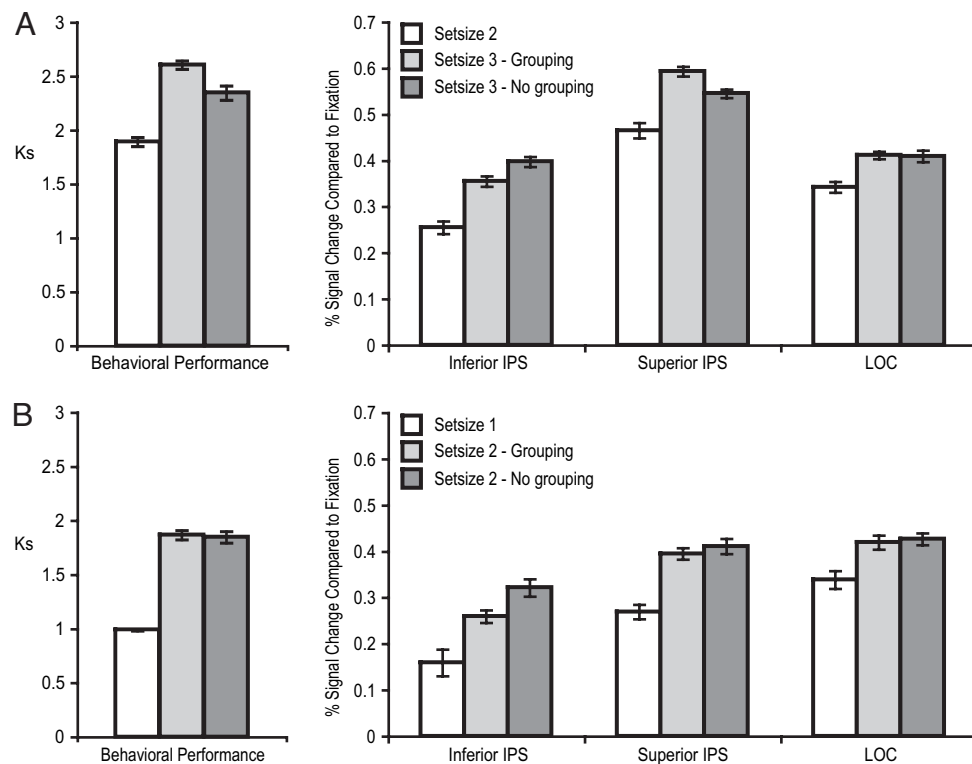


Fig. 2. Behavioral and fMRI results from both experiments. (A) In Experiment 1, for the two three-shape displays, behavioral VSTM capacity estimate K was higher for the grouped than for the ungrouped shapes. The mean fMRI response in the inferior IPS was lower for the grouped than for the ungrouped shapes. This grouping effect reversed direction in the superior IPS and was absent in the LOC. (B) In Experiment 2, behavioral K values did not differ between the two two-shape displays. The mean fMRI response in the inferior IPS was lower for the grouped than for the ungrouped shapes. This grouping effect was absent in both the superior IPS and the LOC.

$F(1, 7) = 41.18, P < 0.001$ for RT difference between one and two ungrouped shapes] (Fig. 2B).

Nevertheless, grouping still significantly modulated responses in the inferior IPS. Mean response was lower for the two grouped than for the two ungrouped shapes in the inferior IPS [$F(1, 7) = 14.80, P < 0.01$] even when behavioral performance was matched (Fig. 2B). However, responses between the two grouped and two ungrouped shapes did not differ in the superior IPS ($F < 1$), mirroring the behavioral results. The interaction between the two two-shape displays and the two parietal ROIs was significant [$F(1, 7) = 6.93, P < 0.05$]. As in Experiment 1, overall fMRI activations in both parietal ROIs were lower for the one- than for either of the two-shape displays [$F(1, 7) = 5.96, P < 0.05$ in inferior IPS; $F(1, 7) = 30.43, P < 0.01$ in superior IPS for the differences between one and two grouped shapes; $F(1, 7) = 12.03, P < 0.05$ in inferior IPS; $F(1, 7) = 22.61, P < 0.01$ in superior IPS for the differences between one and two ungrouped shapes]. LOC was again sensitive to the number of shapes present [$F(1, 7) = 6.19, P < 0.05$ for the difference between one and two grouped shapes; $F(1, 7) = 9.26, P < 0.05$ for the difference between one and two ungrouped shapes], but not to the presence of grouping ($F < 1$ for the difference between the two two-shape displays).

Experiment 2 thus not only replicated Experiment 1, but it additionally showed that grouping-related responses in the inferior IPS are independent of task difficulty. Moreover, because grouping was irrelevant to the main task and did not contribute to task performance, the inferior IPS response likely reflects an obligatory encoding of grouping among visual elements, rather than observers' strategic allocation of attention to improve task performance.

Findings from both experiments also ruled out the possibility that inferior IPS responses simply reflect the number of background bars that were attended. This is because inferior IPS responses

tracked the number of shapes to encode even while attending to the same numbers of background bars (comparing two-shape and the three grouped-shape displays in Experiment 1 and comparing the one-shape and the two grouped-shape displays in Experiment 2).

General Discussion

Together the two experiments demonstrate the role of the inferior IPS and the interplay between the inferior and superior IPS in visual grouping. The inferior IPS represents perceptually grouped items even when such grouping is task-irrelevant. This relative ease of selecting and representing grouped shapes then allowed more shape information to be encoded and stored in the superior IPS. This finding may explain why grouped visual elements are easier to perceive than ungrouped ones after parietal brain lesions (8–13). Unlike patient studies, which usually involve extensive parietal lesions, our findings provide more well defined and fine-grained analyses of these fundamental perceptual mechanisms in the normally functioning parietal lobe.

Previous fMRI studies have shown that attentional control signals are influenced by object-based representations during shifts of visual attention in both the superior parietal lobule and the primary visual cortex (29, 30). For such attentional shifts to occur, however, object-based representations must be formed first. Thus, our study provides a critical missing link by showing grouping effects in the inferior IPS even when the grouping cues were task-irrelevant. The visually parsed representations in the inferior IPS may subsequently modulate neural responses in the superior parietal lobule and the primary visual cortex when a shift of visual attention is required.

Although visual grouping is clearly represented in the inferior IPS and could affect subsequent visual processing in other brain areas, further research is needed to understand how grouping is

initially computed, whether in lower visual areas, the LOC, the inferior IPS, or through the interaction of occipital and parietal mechanisms.

A Neural Object File Framework. Previous behavioral research and theories have argued that there are two distinctive stages of visual processing when multiple objects are encoded. Namely, there exists an object-individuation stage, where objects are selected based on their spatial/temporal information, and an object-identification stage, where the full object representation becomes available. For example, Sagi and Julesz (31) noticed that observers were faster in detecting targets among distractors in a visual search task, but were considerably slower in identifying the targets. Kahneman, Treisman, and Gibbs (32) proposed in their well known object file theory that spatial and temporal information allows an object file to be created or assigned (corresponding to object individuation), which can then be filled with object features to allow objects to be identified (corresponding to object identification). Similarly, Pylyshyn argued in his FINST theory (33, 34) that there is a preattentive stage of visual processing where a fixed number of four objects are indexed by their spatial locations. Featural information is only available at a later attentive stage of processing. In infant research, the distinction between object individuation and identification also has been noted, and it is argued that the development of object individuation precedes object identification by a few months (35, 36).

Our fMRI studies indicate possible neural mechanisms that may support object individuation and identification. In a previous study (20), by varying the number as well as the complexity of objects encoded in VSTM, we found that, although representations in the inferior IPS are fixed to four objects/locations regardless of object complexity, those in the superior IPS and lateral LOC are variable, tracking the number of objects held in VSTM, and representing fewer than four objects as their complexity increases. These results support the existence of two distinctive stages of neural processing when multiple visual objects are encoded. Corresponding to object individuation, the inferior IPS selects a fixed number of four objects by their spatial locations; corresponding to object identification, the superior IPS and LOC encode the shape features of a subset of the selected objects into great detail.

Because object individuation mainly concerns object location and not identity, and because object identification mainly concerns object features such as shape, in a recently completed study, we found that when identical objects were presented simultaneously at different spatial locations, the brain area involved in object individuation (the inferior IPS) treated them as multiple entries to the system. Thus, the inferior IPS represented four identical objects as four different objects. In contrast, brain areas involved in representing detailed object features and identity information (the superior IPS and LOC for object shapes) treated multiple identical objects as a single unique object because the demand to represent the features of multiple identical objects was the same as that of a single unique object. In this case, the superior IPS and LOC represented four identical objects as one unique object.

If the inferior IPS is indeed involved in object individuation, then grouping between visual elements should significantly modulate how visual objects are represented and selected in this brain area. Confirming this prediction, the present results showed that grouping reduced the number of discrete objects represented in the inferior IPS, such that two grouped shapes were treated as more than one shape, but less than two ungrouped shapes (Experiment 2; see also similar results in Experiment 1). This relative ease of representation and selection for grouped shapes consequently allowed more shape information to be represented at the object-identification stage in the superior IPS when the task became more demanding (Experiment 1). Because visual representations were modulated by perceptual grouping cues in the inferior IPS even when such representations did not contribute to task performance,

the present results further suggest that the steps involved in object individuation may be obligatory whenever there are multiple objects. Thus, the present study provides support for our framework, which distinguishes the neural mechanisms associated with object individuation and identification.

Traditionally, the neural mechanisms involved in localizing objects (where) and for identifying what they are have been mapped onto distinct dorsal and ventral visual processing streams, respectively (37). Our neural object file model is consistent with such what and where distinctions, but it additionally shows how these two types of processing interact during visual object individuation and identification. Together with our previous findings (20), the present results also reveal the involvement of the parietal cortex in not only the where, but also the what processing of visual objects (i.e., object identification in superior IPS), thereby enriching our understanding of parietal cortex function in visual cognition.

Although grouping affected object identification in the superior IPS, it did not seem to modulate object identification in the LOC, although this brain area was sensitive to the number of objects encoded. It is possible that fMRI signal-to-noise ratio was greater in the superior IPS than in the LOC (20). Thus, with more power, a grouping effect also may be observed in the LOC. However, it is also possible that the LOC and superior IPS may play somewhat different roles in object identification, and that grouping may differentially affect processing in these brain areas. Further studies are needed to understand how distinct parietal and occipital mechanisms may cooperate to support object individuation and identification in visual perception.

Objects and Groups. In this study, to understand object-based representation in the brain, we examined the impact of grouping on both behavioral VSTM performance and parietal brain responses. One may wonder, what is the connection between groups and objects? Do grouped visual elements become an object? The answer to this question, however, depends on what one considers an object to be. Marr (38) once commented, “What . . . is an object, and what makes it so special that it should be recoverable as a region in an image? Is a nose an object? Is a head one? Is it still one if it is attached to a body? What about a man on horseback? [A]ll these things can be an object if you want to think of them that way, or they can be a part of a larger object” (p. 270). Thus, although grouping between visual elements may be obligatory and governed by both bottom-up Gestalt principles and top-down knowledge and experience, what exactly a visual object is, or the current unit of information selection, may largely depend on an observer’s current focus of visual attention, intention, and goal. For example, if you are attending to a man on horseback running across a field, both the man and the horse may be parts of a single moving object; however, if you are trying to figure out who is on horseback or which horse the man is riding, then, by attending to each separately, the man and the horse will be considered as two separate objects. Hence, grouped visual elements form the basis of objecthood, but a given group may not always be the current object of attention and information-processing unit (6, 7).

In our study, because observers’ main goal was to select and encode the individual shapes in VSTM and the black rectangles were task-irrelevant, the level of attentional selection was at the individual shapes. According to the logic presented earlier, these shapes may thus be considered as the units of visual information processing or objects for this task. Nevertheless, the groupings between these shapes were still encoded by the inferior IPS, suggesting that this brain area may represent the overall hierarchical structure present in a visual display, independent of the level of attentional selection. This may explain why two grouped shapes were treated as more than a single shape, but less than two ungrouped shapes.

Our results have additional interesting implications for understanding the cognitive and neural mechanisms underlying visual

selection. To guide the shift of visual attention across the different levels of the visual hierarchy and to select objects at the appropriate level, we would like to argue that two processing systems are needed during visual perception: one tracking the overall hierarchical structure of the visual display, and the other processing the current objects of attentional selection. Our results suggest that the inferior IPS carries such a hierarchical representation of the visual display, with the LOC and superior IPS possibly representing what is most relevant to the current goal of visual processing. Work by Yantis and colleagues (39) indicates that the control of the attentional shift signal may originate from the superior parietal lobule, which is involved in the shift of visual attention among objects, visual features, spatial locations, and even different sensory modalities. We argue that the interactions among these different cognitive and neural mechanisms enable us to perceive both an individual tree and the entire forest.

Methods

Participants. Thirteen paid observers (six females) participated in Experiment 1, and nine observers (four females) participated in Experiment 2. One observer (female) participated in both Experiments 1 and 2 \approx 6 months apart. All observers were recruited from the Yale University campus. All were right-handed, had normal or corrected to normal vision, and normal color vision. Informed consent was obtained from all observers, and the study was approved by the Yale committee on the use of humans as experimental subjects. Data from one observer (male) in Experiment 1 was excluded from further analysis because of extremely low behavioral performance in the color VSTM experiment (used to localize the superior IPS ROI). Data from one observer (male) in Experiment 2 was excluded because of a lack of time-locked fMRI signal.

Design and Procedure. Experiment 1. We used a VSTM change-detection paradigm (17, 22). To manipulate item grouping, our display background contained four filled black rectangles around the central fixation and was present throughout the experiment. Observers viewed either two or three gray shapes appearing on the rectangles in a sample display and, after a brief delay, detected a shape change (50% probability) in a test display at the probed location (Fig. 1A). For change trials, the probe consisted of a new shape not present in the sample display.

When the display contained two shapes, each appeared on a separate adjacent black rectangle. When the display contained three shapes, they could be either grouped, with two shapes appearing on two ends of the same black rectangle and the third shape appearing on a separate adjacent rectangle, or ungrouped, with all three shapes appearing on separate rectangles. The presence of the black rectangle thus contributed to how shapes were grouped, although it was irrelevant to the VSTM task and could be completely ignored. To isolate an object-based grouping representation independent of a space-based one, we matched the spatial distance between the shapes in the grouped and ungrouped displays. To increase task difficulty and ensure observers' attention on the display, we also included displays that contained four shapes to be retained in VSTM. These trials served as filler trials and were not analyzed for the purpose of the experiment.

Our displays subtended $13.7^\circ \times 13.7^\circ$ and were presented on a light gray background. Seven different object shapes were used as in Xu and Chun (20, Experiment 4) (see Figs. 1 and 2), and they subtended maximally $3.1^\circ \times 3.1^\circ$. Each trial lasted 6 sec and consisted of fixation (1,000 ms), sample display (200 ms), blank delay (1,000 ms), test display/response period (2,500 ms), and response feedback at fixation (1,300 ms) as either a happy face for a correct response or a sad face for an incorrect response presented. To control for the differences associated with a particular display orientation, the illustrated display orientation (Fig. 1) was rotated 90° in half of the trials. Besides

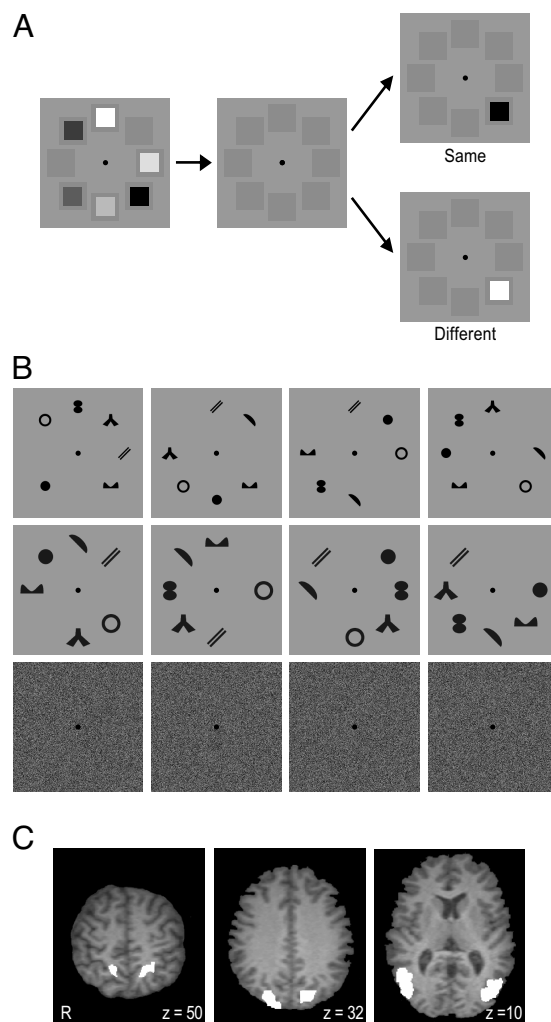


Fig. 3. Example stimuli for the localizer scans and ROIs from an example observer. (A) An example trial of the color VSTM task used to define the superior IPS ROI. Using the same design as the main experiment, observers remembered the colors of each square in the sample display and judged whether the probe color in the test display matched the corresponding color in the sample display. (B) Examples of object images (Top, Experiment 1; Middle, Experiment 2) and noise images (Bottom) used to define the LOC and inferior IPS ROIs in the main experiments. Observers attended these displays and monitored for an occasional spatial jitter occurring randomly to the displays. (C) Superior IPS (Left), inferior IPS (Center), and LOC (Right) ROIs from an example observer.

trials containing shapes, there also were blank fixation trials, in which only a fixation dot was present throughout the 6-sec trial duration. These trials provided the baseline to calculate the percentage of fMRI signal change for the shape trials. The presentation order of the different trial types was pseudorandom and balanced in a run (15, 17, 20). Each observer was tested with two runs, one with the display orientation shown in Fig. 1 and the other with the rotated display orientation. The presentation order of the two runs was balanced among observers. Each run contained 15 trials for each display condition.

Experiment 2. In this experiment, the display background contained two filled black rectangles on either side of the central fixation, similar to those used in a previous behavioral study (4). As in Experiment 1, these rectangles were present throughout the experiment. Observers retained in VSTM either one or two gray shapes appearing on the rectangles (Fig. 1B). When two shapes appeared on the black rectangles in the sample display, they could either

appear on the two ends of the same black rectangle, forming one perceptual group, or on the ends of separate rectangles the same distance apart, forming two separate perceptual groups. Shapes used in this experiment subtended maximally $2.1^\circ \times 2.1^\circ$. As in Experiment 1, displays with four shapes also were included as filler trials to increase task difficulty and were not analyzed for the purpose of the experiment. In addition, to control for the differences associated with a particular display orientation, the illustrated display orientation (Fig. 1B) was rotated 90° in half of the trials. Other aspects of the experiment were identical to those of Experiment 1.

Localizer Scans. As in our previous study (20), to define the superior IPS ROI, a VSTM color experiment with a design similar to that of the main experiments was conducted. A given sample display contained either two to four or six colored squares around the center fixation. The probe color in the test display either matched a color at the same location in the sample display for no-change trials (Fig. 3A Top), or it was a color present elsewhere in the sample displays for change trials (Fig. 2A Bottom). Seven colors (red, green, blue, cyan, yellow, white, and magenta) were used. As in the main experiments, all displays subtended $13.7^\circ \times 13.7^\circ$ and were presented on a light gray background. Each colored square subtended $2.0^\circ \times 2.0^\circ$. Each observer was tested with two runs, each containing 12 trials for each display set size and lasting 5 min and 12 sec.

We defined the LOC and inferior IPS as in our previous study (20). Observers viewed blocks of object images and blocks of noise images (Fig. 3B). Each object image contained six black shapes created by the same algorithm used to generate the displays in the main experiments (Fig. 3B Top and Middle for Experiments 1 and 2, respectively). This step ensured that we only selected brain regions involved in processing the types of visual objects used in our VSTM experiment. Each image lasted 750 ms and was followed by a 50-ms blank interval before the next image appeared. Observers fixated at the center and detected a slight spatial jitter, occurring randomly in 1 of every 10 images. This task helped ensure attention to the displays. Each observer was tested with two runs, each containing 160 object images and 160 noise images. Displays used in this localizer scan had the same spatial extent as those in the main experiments. Examples of the ROIs are shown in Fig. 3C.

fMRI Methods. Observers lied on their back inside a Siemens Trio 3T scanner and viewed through a mirror the displays projected onto

a screen by an LCD projector (Siemens, New York, NY). Stimulus presentation and behavioral response collection were controlled by an Apple Powerbook G4 running Matlab with Psychtoolbox extensions (40, 41). Standard protocols were followed to acquire the anatomical images. A gradient echo pulse sequence (echo time, 25 ms; flip angle 90° ; matrix 64×64) with a repetition time of 2.0 sec was used in the blocked object and noise image runs and TR of 1.5 sec in the event-related VSTM runs. Twenty-four 5-mm-thick (3.75×3.75 -mm in-plane, 0-mm skip) slices parallel to the anterior commissure–posterior commissure line were collected.

Data Analysis. fMRI data collected were analyzed by using BrainVoyager QX (www.brainvoyager.com). Data preprocessing included slice acquisition time correction, 3D motion correction, linear trend removal, and Talairach space transformation (42).

A multiple regression analysis was performed separately on each observer on the data acquired in the color VSTM task. The regression coefficient for each set size was weighted by the corresponding behavioral K estimate from that observer for that set size, as in ref. 17. The superior IPS ROI was defined as the voxels that showed a significant activation in the regression analysis (false discovery rate $q < 0.05$, corrected for serial correlation) and whose Talairach coordinates matched those reported previously (17). When extensive activations were observed in the superior IPS, only 20 ($1 \times 1 \times 1$ mm³) voxels around the reported Talairach coordinates were chosen. The LOC and inferior IPS ROIs were defined as regions in the ventral and lateral occipital cortex and in the inferior IPS, respectively, whose activations were higher for objects than for noise images (false discovery rate $q < 0.05$, corrected for serial correlation).

We overlaid the ROIs onto the data from our shape VSTM experiment and extracted time courses from each observer. As in previous studies (15, 17), these time courses were converted to percent signal change for each stimulus condition by subtracting the corresponding value for the fixation trials and then dividing by that value. Following prior convention (17, 20), the peak responses were derived by collapsing the time courses of all the conditions and determining the time point of greatest signal amplitude in the averaged response. This procedure was done separately for each observer in each ROI. The resulting peak responses were then averaged across observers.

We thank Jenika Beck for assistance in fMRI subject recruiting. This work was supported by National Science Foundation Grant 0518138 and 0719975 (to Y.X.).

- Scholl BJ (2001) *Cognition* 80:1–46.
- Wertheimer M (1924) *Gestalt Theory. Address Before the Kant Society, Berlin, 7th December, 1924*; reprinted in Wertheimer M (1950) in *A Sourcebook of Gestalt Psychology*, ed Ellis WD (Humanities, New York), pp 1–11.
- Palmer SE (1999) *Vision Science: Photons to Phenomenology* (MIT Press, Cambridge, MA).
- Egley R, Driver J, Rafal R (1994) *J Exp Psychol Gen* 123:161–177.
- Watson S, Kramer A (1999) *Percept Psychophys* 61:31–49.
- Xu Y (2002) *Percept Psychophys* 64:1260–1280.
- Xu Y (2006) *Percept Psychophys* 68:815–828.
- Mattingley JB, Davis G, Driver J (1997) *Science* 275:671–674.
- Gilchrist I, Humphreys GW, Riddoch MJ (1996) *Cogn Neuropsychol* 13:1223–1256.
- Ward R, Goodrich S, Driver J (1994) *Vis Cogn* 1:101–130.
- Balint R (1909) *Monatschr Psych Neurol* 25:5–81.
- Coslett HB, Saffran E (1991) *Brain* 114:1523–1545.
- Friedman-Hill SR, Robertson LC, Treisman A (1995) *Science* 269:853–855.
- Wojculik E, Kanwisher N (1999) *Neuron* 23:747–764.
- Kourtzi Z, Kanwisher N (2000) *J Neurosci* 20:3310–3318.
- Culham J, Cavanagh P, Kanwisher N (2001) *Neuron* 32:737–745.
- Todd JJ, Marois R (2004) *Nature* 428:751–754.
- Todd JJ, Marois R (2005) *Cogn Affect Behav Ne* 6:144–155.
- Vogel EK, Machizawa MG (2004) *Nature* 428:748–751.
- Xu Y, Chun MM (2006) *Nature* 440:91–95.
- Kourtzi Z, Kanwisher N (2001) *Science* 293:1506–1509.
- Luck SJ, Vogel EK (1997) *Nature* 390:279–281.
- Malach R, Reppas JB, Benson R, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RBH (1995) *Proc Natl Acad Sci USA* 92:8135–8139.
- Grill-Spector K, Kushner T, Edelman S, Yitzchak Y, Malach R (1998) *Neuron* 21:191–202.
- Cowan N (2001) *Behav Brain Sci* 24:87–114.
- Moore CM, Egeth H (1997) *Perception* 23:339–352.
- Driver J, Davis G, Russell C, Turatto M, Freeman E (2001) *Cognition* 80:61–95.
- Chan WY, Chua FK (2003) *Psychon B Rev* 10:932–938.
- Müller NG, Kleinschmidt A (2003) *J Neurosci* 23:9812–9816.
- Shomstein S, Behrmann M (2006) *Proc Natl Acad Sci USA* 103:11387–11392.
- Sagi D, Julesz B (1984) *Perception* 14:619–628.
- Kahneman D, Treisman A, Gibbs BJ (1992) *Cogn Psychol* 24:175–219.
- Pylyshyn ZW (1989) *Cognition* 32:65–97.
- Pylyshyn ZW (1994) *Cognition* 50:363–384.
- Leslie AM, Xu F, Tremoulet PD, Scholl BJ (1998) *Trends Cogn Sci* 2:10–18.
- Leslie AM, Kaldy Z (2007) in *Short- and Long-Term Memory in Infancy and Early Childhood: Taking the First Steps Toward Remembering*, eds Oakes LM, Bauer PJ (Oxford Univ Press, Oxford), pp 103–125.
- Ungerleider LG, Mishkin M (1982) in *Analysis of Visual Behavior*, eds Ingle DJ, Goodale MA, Mansfield RJW (MIT Press, Cambridge, MA), pp 549–586.
- Marr D (1982) *Vision* (Freeman, New York).
- Serences J, Liu T, Yantis S (2005) in *Neurobiology of Attention*, eds Itti L, Rees G, Tsotsos J (Academic, New York), pp 35–41.
- Brainard DH (1997) *Spatial Vision* 10:433–436.
- Pelli DG (1997) *Spatial Vision* 10:437–442.
- Talairach J, Tournoux P (1988) *Co-Planar Stereotaxic Atlas of the Human Brain* (Thieme Medical, New York).