

Representing the Forest before the Trees: A Global Advantage Effect in Monkey Inferotemporal Cortex

Arun P. Sripati¹ and Carl R. Olson^{1,2}

¹Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, and ²Department of Neuroscience, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Hierarchical stimuli (large shapes composed of small shapes) have long been used to study how humans perceive the global and the local content of a scene—the forest and the trees. Studies using these stimuli have revealed a global advantage effect: humans consistently report global shape faster than local shape. The neuronal underpinnings of this effect remain unclear. Here we demonstrate a correlate and possible mechanism in monkey inferotemporal cortex (IT). Inferotemporal neurons signal the global content of a hierarchical display ~30 ms before they signal its local content. This is a specific expression of a general principle, related to spatial scale or spatial frequency rather than to hierarchical level, whereby the representation of a large shape develops in IT before that of a small shape. These findings provide support for a coarse-to-fine model of visual scene representation.

Introduction

Perceptual psychologists have made frequent use of hierarchical displays to analyze global and local visual perception in humans. Hierarchical displays are well suited to this purpose because they allow measuring responses to the same shape both at the level of the whole object and at the level of its parts. A key phenomenon discovered by the use of such stimuli is the global advantage effect: humans report global shape faster than local shape (Navon, 1977, 2003; Kimchi, 1992). This effect might depend on hierarchical level, size, or spatial frequency (Kimchi, 1992). Its neuronal underpinnings are not well understood. It is known that the right and left hemispheres of humans contribute disproportionately to the processing of global and local shape, respectively, as indicated by studies of brain activation (Fink et al., 1996; Han et al., 2002), brain injury (Robertson and Lamb, 1991), and performance in the two visual hemifields (Hübner, 1998). However, studies in humans have not revealed any obvious mechanism, such as stronger or earlier neuronal responses to global shape, that could explain the global advantage effect. This issue is not only of basic but also of clinical relevance because tests involving hierarchical figures have been used as probes for visual system dysfunction in a variety of developmental disorders and diseases (Bihrlé et al., 1989; Slavin et al., 2002; Johnson et al., 2005; Behrmann et al., 2006).

We set out to investigate this issue by monitoring neuronal responses to hierarchical stimuli in monkey inferotemporal cortex (IT). Inferotemporal cortex is a natural target for study be-

cause, as the terminus of the ventral stream, a chain of areas dedicated to pattern vision (Ungerleider and Mishkin, 1982), it contains neurons selective for complex images (Tanaka, 1996, 2000). We used an identical set of stimuli in recording from all neurons so as to characterize the population representation of displays that could also be used in human behavioral experiments. We aimed to answer the following specific questions: are IT neurons sensitive to both global and local shape; do global and local signals differ in strength or timing; do different neurons represent global or local image content; and, if so, is it because they are specialized for representing shape at a particular hierarchical level (global or local) or at a particular scale (large or small) or spatial frequency (low or high)?

Materials and Methods

Stimuli

The stimulus set consisted of four hierarchical stimuli and four solid stimuli in the form of circles and diamonds (Fig. 1A). The use of shapes containing detail solely in the visual field periphery (Navon and Norman, 1983) prevents confounds that occur when some local elements are on or near the fovea (Kinchla and Wolfe, 1979). The diamonds and circles forming the local components of the hierarchical stimuli were of equal area. The length of the side of a local diamond was 1.54°, and the diameter of a local circle was 1.74°. A global diamond and a global circle were of the same relative size. The length of the side of a global diamond (the distance between the centers of the small shapes at adjacent corners) was 6.4°, and the diameter of a global circle (the distance between the centers of diametrically opposed small shapes) was 7.2°. The small solid shapes were identical to the local components of the hierarchical stimuli. The large solid shapes (contours with a width of 0.62°) were of the same size as the global shapes. All displays were centered at fixation with the exception of the small solid shapes, which were centered at an eccentricity of 3.6° in the visual field opposite the recording hemisphere. It was necessary to place the small shapes eccentrically so as to equate them to local elements in the hierarchical displays with regard to retinal visual acuity. Each stimulus was white and had a luminance of 30.3 cd/m² against a background of <0.001 cd/m². The viewing distance was 46 cm.

Received Dec. 3, 2008; revised May 7, 2009; accepted May 12, 2009.

Research support was provided by National Institutes of Health (NIH) Grant R01 EY018620 and the Pennsylvania Department of Health through the Commonwealth Universal Research Enhancement Program. Technical support was provided by NIH Grant P30 EY08098. Collection of magnetic resonance images was supported by NIH Grant P41 RR03631. We thank Karen McCracken for technical assistance.

Correspondence should be addressed to Arun P. Sripati, Center for the Neural Basis of Cognition, Carnegie Mellon University, 115 Mellon Institute, 4400 Fifth Avenue, Pittsburgh, PA 15213. E-mail: sparun@cnbc.cmu.edu.

DOI:10.1523/JNEUROSCI.5766-08.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/297788-09\$15.00/0

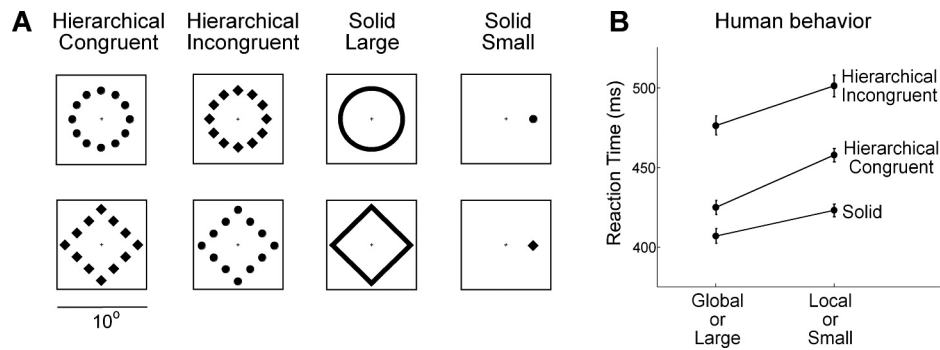


Figure 1. *A*, The visual stimuli. The square frame around each is provided for spatial reference but was not visible to the monkey. At its center, corresponding to the center of the monitor, is a small fixation cross. This was displayed before and after but not during presentation of the stimulus. *B*, Six observers performed a two-choice circle–diamond discrimination task under viewing conditions precisely matched in all respects to those imposed during neuronal data collection (see Materials and Methods). There was a significant global advantage effect as evidenced by the fact that responses to global shape were significantly faster than responses to local shape (global mean = 450 ms, local mean = 479 ms; ANOVA, $p < 0.00001$). In contrast, the global percentage correct (91%) and the local percentage correct (93%) did not differ significantly (χ^2 test, $p = 0.29$). There was interference between the local and global levels as evidenced by the fact that responses to congruent stimuli were significantly faster than responses to incongruent stimuli (congruent mean = 441 ms, incongruent mean = 489 ms; ANOVA, $p < 0.00001$) with interference in the two directions (global–local and local–global) of equal strength as evidenced by the absence of a significant interaction between hierarchical level (global or local) and congruence status (congruent or incongruent) (ANOVA, $p = 0.46$). This effect were mirrored in the percentage correct scores. The percentage correct on congruent trials was significantly greater than on incongruent trials both in the global task (96% vs 88%, χ^2 test, $p = 0.0002$) and in the local task (97% vs 89.5%, χ^2 test, $p = 0.0002$). Finally, responses to large solid stimuli were significantly faster than responses to small solid stimuli (large mean = 407 ms, small mean = 423 ms; ANOVA, $p = 0.009$), whereas the percentage correct scores for large and small shapes (96% and 95%) did not differ significantly (χ^2 test, $p = 0.99$). Each error bar represents \pm SEM for the corresponding condition as based on intrasubject variability with intersubject variability factored out. This was computed as the average of the SD across six subjects divided by $\sqrt{6}$.

Behavioral testing in humans

Six right-handed adults, five male and one female, completed tests conducted under a protocol approved by the Institutional Review Board of Carnegie Mellon University. All viewing conditions, including screen distance and the configuration, size, eccentricity, and luminance of the stimuli, were identical to those imposed during single neuron recording in monkeys. The sole deviation from practice in the monkey experiments was to present the small stimuli above instead of lateral to fixation. Each participant sat with his/her head resting on a chin rest, facing the screen with the left and right index fingers on two keys. Each trial began with onset of a red cross on which participants were instructed to maintain fixation. After an interval of 750 ms, a stimulus appeared for 150 ms. Participants were instructed to depress the right key for a circle and the left key for a diamond. After a correct response, the fixation cross was flashed several times to provide feedback to the subject. After a practice session, which continued until the participant was comfortable with the task, four blocks of trials ensued.

Global block. The four hierarchical stimuli were presented in random order. The participant was instructed to respond on the basis of the global shape. Testing continued until 25 trials had been correctly completed for each stimulus.

Local block. The four hierarchical stimuli were presented in random order. The participant was instructed to respond on the basis of the local shape. Testing continued until 25 trials had been completed correctly for each stimulus.

Large and small blocks. The two solid stimuli of a given size were presented in random order. Testing continued until 25 trials had been completed correctly for each stimulus.

The order of the global and local blocks and that of the large and small blocks was counterbalanced across participants. The global and local blocks were always presented first.

Analysis of human behavioral data

To assess the dependence of reaction time (RT) on task factors in the global and local blocks, we performed a four-way ANOVA with RT as the dependent variable and with hierarchical level (global or local), congruence status (congruent or incongruent), shape at the attended level (circle or diamond), and subject (six subjects) as factors. To assess the dependence of RT on task factors in the large and small blocks, we performed a three-way ANOVA with RT as the dependent variable and with size (large or small), shape (circle or diamond), and subject (six subjects) as factors. In both analyses, shape was included as a factor to

account for variance that would otherwise have been treated as noise. However, effects involving shape were uninterpretable because shape (circle or diamond) was correlated with responding hand (right or left). Accordingly, we do not comment on these effects in the main text. Cross-condition differences in percentage correct were assessed statistically by applying a χ^2 test to counts of correct and incorrect trials collapsed across participants.

Neuronal recording

Two rhesus macaque monkeys, one male and one female (laboratory designations Je and Ec), were used. All experimental procedures were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee and were in compliance with the guidelines set forth in the United States Public Health Service *Guide for the Care and Use of Laboratory Animals*. All aspects of the behavioral experiment (stimulus presentation, eye position monitoring, and reward delivery) were under control of a computer running Cortex software (NIMH Cortex). Eye position was monitored by means of a scleral search coil system (Riverbend Instruments). At the beginning of each day's session, a varnish-coated tungsten microelectrode with an initial impedance of ~ 1.0 M Ω at 1 kHz (FHC) was introduced into the temporal lobe through a transdural guide tube advanced such that its tip was ~ 10 mm above IT. The electrode was then advanced by use of a micromanipulator until phasic visual responses were observed. Action potentials of single neurons were isolated from the multineuronal trace using a commercially available spike-sorting system (Plexon). All waveforms were recorded during the experiments, and spike sorting was performed off-line using commercially available spike-sorting software (Plexon). Neurons were selected for recording if they appeared to respond to at least one of the eight stimuli. This was the only step of selection. Consequently, the data may be taken as accurately reflecting the population representation of the fixed set of stimuli used in the behavioral phase of the experiment. At the end of the data collection period, recording sites were established by structural magnetic resonance imaging to occupy the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus [at levels ranging from anterior 14 (A14) to A22 mm in Je and A9 to A18 mm in Ec relative to the interaural plane].

We monitored neuronal activity during passive fixation. A trial began with onset of a small red fixation cross and attainment of fixation. After a delay of 200 ms, eight stimuli were presented in succession, each with a duration of 200 ms followed by an interval of 200 ms during which the fixation cross was again visible. After completion of the sequence, the

display vanished and the monkey was rewarded with ~ 0.1 cc of water. Although the fixation window was large (4.2°), we found on *post hoc* analysis that the gaze remained closely centered throughout the duration of each trial. The average across sessions of the SD of horizontal and vertical gaze angle was 0.2° . On each trial, two images were presented in alternation for a total of four presentations each. There were eight trial conditions representing the cross of four stimulus pairs (two congruent hierarchical stimuli, two incongruent hierarchical stimuli, two large stimuli, and two small stimuli) with two orders (stimulus A first and stimulus B first). The sequence in which conditions were imposed was random subject to the constraint that within each block of eight successfully completed trials, there had to be one successfully completed trial conforming to each condition. Over the course of a session, consisting of two such blocks, each of the eight images was presented 16 times.

Statistical assessment of shape selectivity

To characterize the selectivity of each neuron for hierarchical stimuli, we performed an ANOVA with global shape (diamond or circle) and local shape (diamond or circle) as factors and with the firing rate 50–250 ms after stimulus onset as the dependent variable (see Fig. 3A). To characterize selectivity for large solid stimuli (and likewise for small solid stimuli), we performed an ANOVA with shape (diamond or circle) as the single factor and with the firing rate 50–250 ms after stimulus onset as the dependent variable.

Latency of onset of discriminative shape signal

For each neuron, we estimated the onset time of the shape signal as the leading edge of the time window in which the difference between the firing rates elicited by the two shapes was most significant. We adopted this approach because it maximized the signal-to-noise ratio and thus minimized random jitter in the estimate of latency. For each window in a large family ranging in duration from 10 to 200 ms and ranging in start time from 50 to 200 ms (in 1 ms steps), we performed a *t* test comparing the firing rates elicited by a (global or local) circle and diamond across trials. Onset time was taken as the leading edge of the window yielding the lowest *p* value.

We cross-checked results obtained by this method through use of a more orthodox but more noise-prone approach based on finding the earliest window of fixed size within which the firing rates elicited by two shapes were significantly different (*t* test, $\alpha = 0.05$). Over window sizes ranging from 50 to 150 ms, the global onset time was consistently shorter than the local onset time. The effect achieved maximal significance with a window of 90 ms. It is on this window that comments in Results are based.

To cross-check the method further, we applied it to simulated data in which the signal had a known latency. We generated spikes according to a Poisson process in which, 100 ms after stimulus onset, the firing rate shifted from baseline to a steady level corresponding to the peak of the measured population response (see Fig. 5A, B). We generated two sets of spikes with 32 trials each, corresponding to the responses to the preferred and nonpreferred shapes. We then measured the latency of the discriminative shape signal. The estimated latency (110 ms) was only slightly offset from the preset value of 100 ms. This was true regardless of whether signal strength was matched to the observed mean for global shapes or the observed mean for local shapes.

Latency of response onset

To estimate the latency of the onset of the visual response—as distinct from the onset of discriminative activity—we used an approach identical to the one described above with the exception that we used a *t* test to compare the firing rate within a given window to the firing rate 0–50 ms after stimulus onset.

Population histograms

To create population histograms, we calculated the firing rate in successive 15 ms bins beginning with onset of the stimulus. We identified the preferred shape using one half of all trials, and constructed the population histogram using the other half. We repeated this procedure twice, identifying the preferred shape first on the basis of odd-numbered trials and then on the basis of even-numbered trials, after which we averaged

the results. This procedure eliminates the spurious signal that would arise if the histogram were based on the same trials used to identify the preferred shape.

Partial correlation analysis of selectivity for global, local, large, and small shape

For each neuron and for each condition (global, local, large, and small), we computed an index of shape selectivity: $\text{sign}(C - D) \times Z$, where *C* and *D* were the firing rates 50–250 ms after onset of a circle or diamond, and *Z* was the *z*-score corresponding to the *p* value of the main effect of shape in the corresponding ANOVA. The sign of the index reflected which stimulus the neuron preferred, and its magnitude reflected the statistical significance of the preference. We adopted this approach to emphasize the contribution of significantly selective neurons. Measures of partial correlation were generated with the Matlab *partialcorr* function. Each measure reflected the degree to which shape selectivity under a given condition (say global) was correlated with shape selectivity under another condition (say large) after factoring out of their common correlations with shape selectivity under other conditions (in this example, local and small).

Results

Representation of global and local shape

We recorded the responses of 123 neurons in the left inferotemporal cortex of two monkeys (Ec: *n* = 70; Je: *n* = 53) while they passively viewed hierarchical stimuli from a set (Fig. 1A) for which human observers exhibited a global advantage effect (Fig. 1B). Some IT neurons showed clear signs of sensitivity to global or local image content (Fig. 2). Cell 1 was selective for local content, responding most strongly to hierarchical stimuli composed of circles. Cell 2 was selective for global content, responding more strongly to circular than to rhomboidal arrangements of elements. To determine how frequent such effects were, we performed an ANOVA ($\alpha = 0.05$) on the data from each neuron with firing rate 50–250 ms after stimulus onset as the dependent variable and with global and local shape as factors (see Materials and Methods). This revealed a significant main effect of global shape in 17% of neurons and of local shape in 24% of neurons (Fig. 3A). The two rates of incidence were not significantly different (χ^2 test, $p = 0.27$). To compare the strengths of global and local signals, we computed for each neuron an index representing how well it differentiated between a circle and a diamond at each level: $(C - D)/(C + D)$, where *C* and *D* were the responses to the circle and diamond as measured 50–250 ms after stimulus onset. The distributions of the indices are shown in Figure 3, B and C. The median magnitude of the global index was less than the median magnitude of the local index, and the difference was significant (global median = 0.25, local median = 0.32; Wilcoxon rank-sum test on distributions of absolute values, $p = 0.028$). Thus, although approximately comparable numbers of IT neurons were selective for global and local shape, the strength of the signal was somewhat greater for local shape.

IT neurons respond more weakly to a stimulus when it is repeated than on the first presentation (Baylis and Rolls, 1987; Miller et al., 1991; McMahan and Olson, 2007; Liu et al., 2009). Therefore it seemed possible that the global and local shape signals might vary in strength across the four successive presentations of a stimulus that occurred in each trial. To assess this possibility, we separated neuronal responses into four groups according to whether the presentation was the first, second, third, or fourth in the trial. The mean firing rate 50–350 ms after stimulus onset (13.8, 11.6, 11.9, and 11.8 Hz during phases 1–4, respectively) was, as expected, significantly greater during phase 1 than during later phases (paired *t* tests, $p < 0.00003$). However, the strength of the global shape signal (6.4, 6.3, 6.8, and 5.8 Hz for

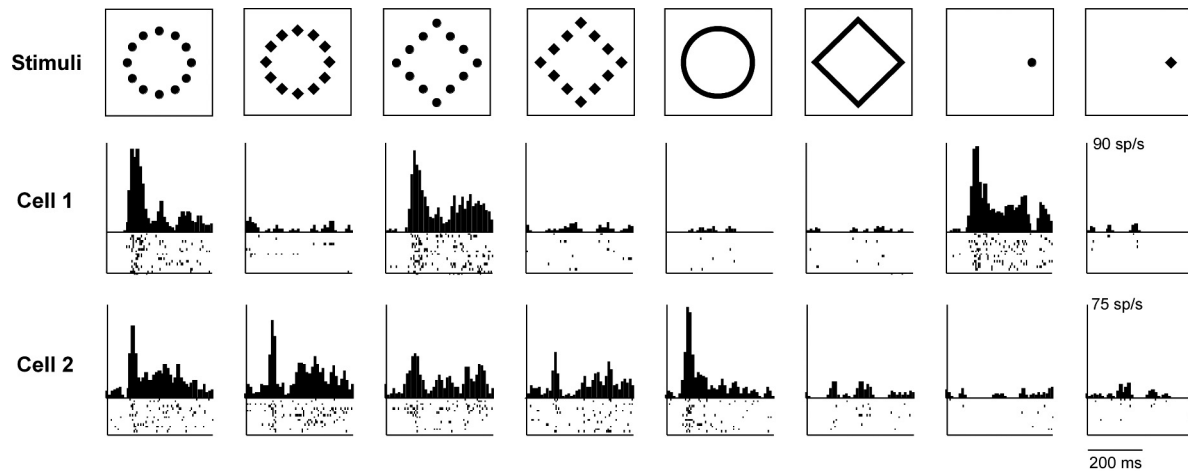


Figure 2. Responses of two IT neurons to hierarchical and solid stimuli. Cell 1 (ec099a_ch1) responded selectively to hierarchical stimuli containing local circles and to a local circle in isolation. Cell 2 (je180b_ch1) responded most strongly to hierarchical stimuli with the global form of a circle and to a large solid circle. Ticks represent action potentials (1 row for each trial). Histograms represent mean firing rate in successive 25 ms windows.

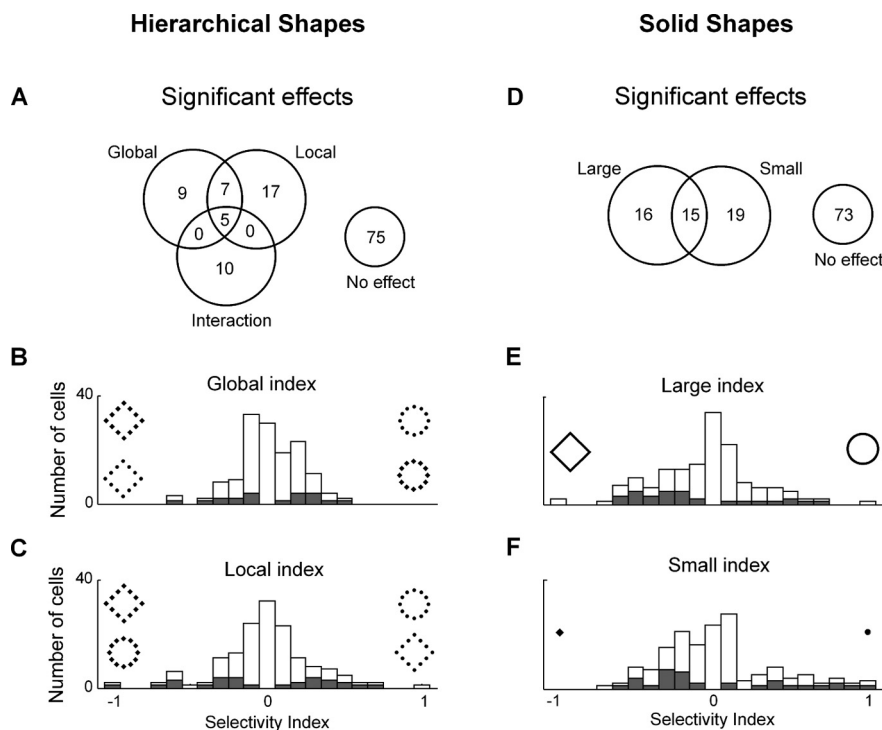


Figure 3. IT neurons were selective for shape at the local and global level in hierarchical stimuli and for the shapes of large and small stimuli. **A**, Counts of statistically significant effects in an ANOVA with global shape and local shape as factors. **B**, Distribution of an index representing the strength of selectivity for global shape in hierarchical displays. **C**, Distribution of an index representing the strength of selectivity for local shape in hierarchical displays. **D**, Counts of statistically significant effects in an ANOVA, performed independently for large and small solid stimuli, with shape as the sole factor. **E**, Distribution of an index representing the strength of selectivity for the shape of a large solid stimulus. **F**, Distribution of an index representing the strength of selectivity for the shape of a small solid stimulus. **B**, **C**, **E**, and **F** are based on data from all 123 neurons. Shaded bars in each panel represent neurons in which the represented form of selectivity was statistically significant.

21 globally selective neurons) and the strength of the local shape signal (9.3, 8.8, 9.6, and 8.8 Hz for 29 locally selective neurons) were not significantly different between the first and later phases (paired *t* tests, $p > 0.6$).

We wondered whether global and local signals developed at different times following stimulus onset. To investigate this possibility, we identified, for each globally or locally selective neuron, the leading edge of the time window in which the difference

between the firing rates elicited by the two shapes was maximally significant (Fig. 4A). The onset time of the global shape signal was significantly shorter than the onset time of the local shape signal (global mean = 99 ms, local mean = 127 ms; difference = 28 ms; *t* test, $p = 0.01$). An alternative method based on finding the time at which the difference in firing rates first achieved significance yielded shorter estimates of latency but the same difference (global mean = 74 ms, local mean = 102 ms; difference = 28 ms; *t* test, $p = 0.04$). The global signal developed at a time indistinguishable from the time of onset of the visual response itself (response mean = 95 ms, global mean = 99 ms; *t* test, $p = 0.86$), whereas the local signal was significantly delayed relative to visual response onset (response mean = 94 ms, local mean = 127 ms; *t* test, $p = 0.0009$). The trend was present in both monkeys and achieved significance in one (Ec, *t* test, $p = 0.001$). We confirmed it in an independent analysis demonstrating that the mean global signal (3.2 Hz) was significantly stronger than the mean local signal (0.3 Hz) during an early phase of the response 50–80 ms after stimulus onset (*t* test, $p = 0.02$). Finally, the effect was visible in histograms representing population activity (Fig. 5A, B) and remained so upon separate consideration of responses to congruent and incongruent stimuli (see Fig. 7).

These differences in timing might have been specific to global and local signals themselves or to the neurons carrying them. To distinguish between these possibilities, we repeated the analysis on a subset of 12 neurons exhibiting significant main effects of both global and local shape (Fig. 3A). The difference in latency was still present (global mean = 97 ms, local mean = 139 ms; *t* test, $p = 0.006$). We performed an identical analysis on 15 neurons exhibiting interaction effects between

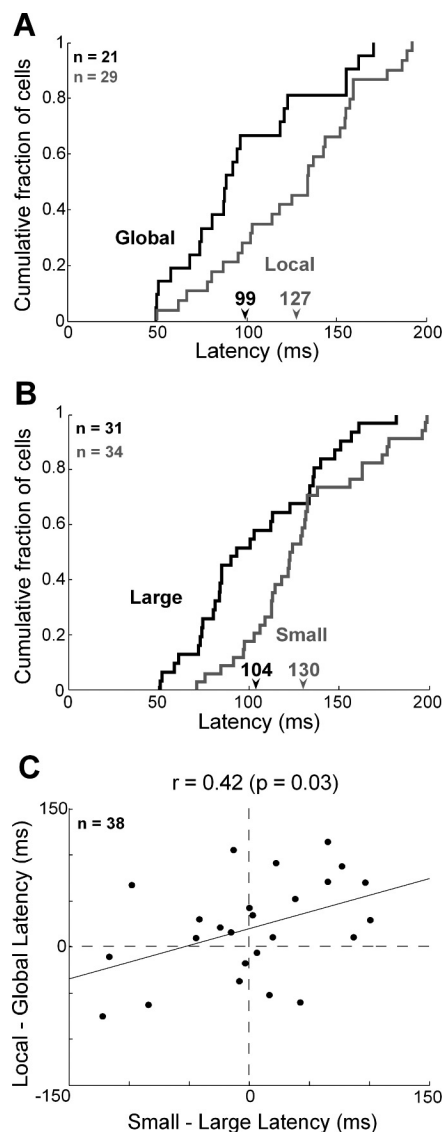


Figure 4. Global and large shape signals have a shorter latency than local and small shape signals. **A**, Cumulative distributions of the latency of the discriminative signal for global shape carried by 21 globally selective neurons (black) and the signal for local shape carried by 29 locally selective neurons (gray). Numbers under the arrows indicate the mean global and local latencies (in milliseconds) for the corresponding groups of cells. **B**, Cumulative distributions of the latency of the discriminative signal for large shape carried by 31 large-shape selective neurons (black) and the signal for small shape carried by 34 small-shape selective neurons (gray). **C**, Global signal lead time (local latency minus global latency) was positively and significantly correlated with large-signal lead time (small latency minus large latency) across 38 neurons selected on the basis of exhibiting significant global or local selectivity.

global and local shape. Again, the mean latency of global signals (121 ms) was less than the mean latency of local signals (151 ms), and the effect approached significance (t test, $p = 0.07$). Thus the difference between global and local latencies was a property of the signals rather than of the neurons.

To determine whether the relative timing of global and local signals was consistent across the four successive presentations of a stimulus that occurred in each trial, we divided each neuron's responses into four groups according to whether the presentation was the first, second, third, or fourth in the trial. Due to a lack of power arising from the fact that each stimulus was presented only four times at a given trial phase, the method for measuring latency that we had applied to the combined data yielded uninter-

pretable results when applied to the subsets of data. Accordingly, we adopted a cruder measure. For each neuron, and for each pair of stimuli, we identified the preferred stimulus as the one eliciting greater activity during the period 100–350 ms after stimulus onset. Then we measured the strength of the discriminative signal (difference in mean firing rates evoked by preferred and nonpreferred stimuli) during the period 0–100 ms after stimulus onset. The mean across neurons of the global signal was 3.4, 1.6, 1.4, and 3.8 Hz during phases 1–4, respectively. The corresponding values for local signal were 0.3, -0.4 , 1.5, and 2.9 Hz. During no phase was the difference between the global and local signals significant (paired t tests, $p > 0.1$). Nevertheless the difference was generally positive (3.1, 2.0, -0.1 , and 0.9 Hz during phases 1–4, respectively), in accordance with the idea that the global signal was stronger than the local signal very early in the response regardless of phase.

Representation of large and small shapes

Neurons exhibiting selectivity for shape at a given hierarchical level might have done so either because they processed shape at that level (global or local) or because they processed shape at a particular scale (large or small). If scale were the determining factor, then we would expect neurons selective for shape at a given level (global or local) in a hierarchical display to exhibit concordant shape selectivity at the corresponding scale (large or small) in a solid display. This was the case for the cells shown in Figure 2. Cell 1 preferred hierarchical displays made up of local circles and responded selectively to a small solid circle. Cell 2 preferred hierarchical displays in the form of a global circle and responded selectively to a large solid circle. To assess the degree to which shape selectivity in one context (global or local) was correlated with shape selectivity in the other context (large or small), we performed a partial correlation analysis. This revealed that the global shape index was significantly correlated with the large (but not the small) shape index, whereas the local shape index was significantly correlated with the small (but not the large) shape index (Fig. 6). We conclude that selectivity for shape at the global or local level was at least in part explained by selectivity for shape at a large or small scale.

To determine whether the two distinctive traits of global signals—their lower strength and their earlier onset—were related to the scale at which global shape was defined, we analyzed the responses of the same neurons to large and small solid stimuli. Upon comparing indices of selectivity for large and small solid shapes (Fig. 3*E,F*), we observed an insignificant trend favoring large shapes (large median = 0.38, small median = 0.31; Wilcoxon rank-sum test on distributions of absolute values, $p = 0.95$). Thus the relative weakness of global shape signals cannot have derived from weakness in signals representing large shape. It may have arisen instead from the poor definition (due to fragmentation) of global shape in a hierarchical display. Upon computing, for each neuron selective for large or small shape, the time at which the difference in firing rate between the preferred and the nonpreferred shape became significant (Fig. 4*B*), we found that the latency of the large shape signal was significantly shorter than the latency of the small shape signal (large mean = 104 ms, small mean = 130 ms; t test, $p = 0.003$). The large-shape signal developed at a time indistinguishable from the time of onset of the visual response itself (response mean = 100 ms, large-shape mean = 104 ms; t test, $p = 0.86$), whereas the small-shape signal was significantly delayed relative to the onset of the visual response (response mean = 113 ms, small-shape mean = 130 ms; t test, $p = 0.01$). These patterns are evident even at the level of

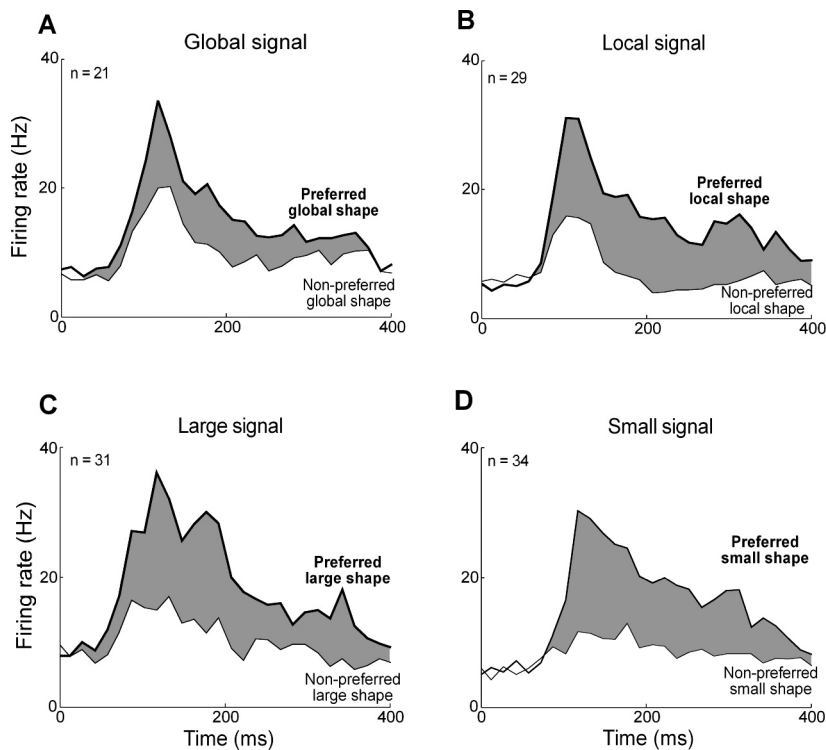


Figure 5. Time course of population responses. **A**, Average response of 21 globally selective neurons to preferred global shape (thick lines) and nonpreferred global shape (thin lines). **B**, Corresponding plot for 29 locally selective neurons. **C**, Corresponding plot for 31 large-solid-shape-selective neurons. **D**, Corresponding plot for 34 small-solid-shape-selective neurons. The width of the gray ribbon in each plot corresponds to the strength of the shape signal.

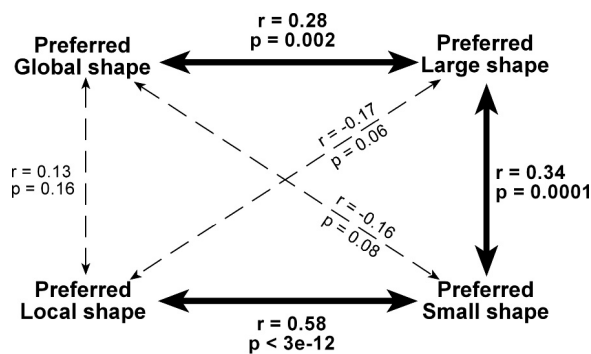


Figure 6. Partial correlation, across neurons, of four indices reflecting selectivity for global, local, large, and small shape. The preferred shape at the global level in hierarchical displays was significantly correlated with the preferred large (but not small) solid shape. The preferred shape at the local level in hierarchical displays was significantly correlated with the preferred small (but not large) solid shape. The preferred large solid shape was significantly correlated with the preferred small solid shape. Each partial correlation coefficient reflected the degree to which shape selectivity under a given condition (say global) was correlated with shape selectivity under another condition (say large) after factoring out their common correlations with shape selectivity under other conditions (in this example, local and small).

population activity (Fig. 5C,D). The large–small latency difference was closely related to the global–local latency difference as indicated by the fact that the two measures were significantly and positively correlated across the population of 38 neurons exhibiting a significant main effect of global or local shape (Fig. 4C). We conclude that the global signal developed earlier because shapes at a larger scale elicit discriminative neuronal activity earlier.

Representation of congruent and incongruent stimuli

Humans processing hierarchical stimuli respond more slowly to incongruent than to congruent displays (Navon, 1977; Kimchi, 1992). In some cases, global-to-local interference is stronger than local-to-global interference (Navon, 1977; Kimchi, 1992), but this is not always the case (Navon and Norman, 1983; Kimchi, 1992; Poirel et al., 2008). In the present study, global-to-local and local-to-global interference both occurred and were equivalent in strength (Fig. 1B). Interference effects are commonly thought to arise from response conflict (Miller and Navon, 2002). By this account, when an incongruent stimulus is displayed, global and local signals representing opposite shapes are relayed from visual cortex to frontal cortex, where they interfere with each other during the period when shape evidence is being accumulated. Signals representing shape at the relevant level prevail because they are stronger (due to enhancement by attention), but they take longer to prevail because signals representing the opposite shape at the irrelevant level are still present (due to incomplete attenuation by attention). To test this account of interference in the monkey would require monitoring neuronal activity during task performance

rather than during passive fixation as in the present study.

Putting aside consideration of this issue, it is still possible to pose the question, do IT neurons, during passive fixation, distinguish better between congruent than between incongruent stimuli? To answer this question, we inspected population plots representing the responses of globally and locally selective neurons to their preferred and nonpreferred global shapes under conditions of congruence (Fig. 7A,B) and incongruence (Fig. 7C,D). The strength of the shape signal (represented by the width of the gray ribbon) indeed appeared to be reduced for incongruent stimuli. To assess the significance of the reduction, we computed, for each globally selective neuron and for each locally selective neuron, a measure of signal attenuation on incongruent trials, $[(P_c - N_c) - (P_i - N_i)] / (P_c + N_c + P_i + N_i)$, where P_c was the firing rate 50–350 ms after onset of the neuron's preferred shape in a congruent setting, N_i was the firing rate after onset of the nonpreferred shape in an incongruent setting, and so on. The incongruence index was greater than zero, an effect that achieved significance among globally selective neurons (mean index = 0.11; t test, $p = 0.03$) but not among locally selective neurons (mean index = 0.01; t test, $p = 0.62$). The effect presumably arose from the tendency, present but not significant in the sampled population, for neurons to prefer the same shape at both the global and the local level (Fig. 6).

In a subpopulation of neurons ($n = 15$), global and local signals combined nonlinearly, as indicated by the presence of a global–local interaction effect (Fig. 3A). The interaction effect by its nature took the form of a stronger response either to congruent or to incongruent stimuli. To assess whether one trend was predominant, we computed the mean across the 15 neurons of the firing rates elicited by congruent and incongruent stimuli.

The congruent response (21.9 Hz) was slightly greater than the incongruent response (20.1 Hz), but the effect did not achieve significance (t test, $p = 0.28$).

Discussion

We have characterized the responses of a population of IT neurons to a fixed set of hierarchical stimuli. The essential findings are as follows: that neurons in macaque IT represent both the global and the local shape of a hierarchical stimulus; that the global signal develops earlier than the local signal; and that this effect is specific to the size of the shape as distinct from its hierarchical level. We will first discuss these findings in relation to previous observations on form selectivity in monkey IT. We will then consider whether the timing of neuronal activity in monkey IT constitutes a plausible mechanism for the global advantage effect in humans.

We found that 17% of neurons are selective for the global configuration of the image. This should be taken as a lower bound since the images were not tailored to the requirements of individual neurons as determined by either receptive field location or shape selectivity. The demonstration of selectivity for global configuration goes beyond the common observation of shape selectivity (Schwartz et al., 1983; Kobatake et al., 1998; Brincat and Connor, 2004; Kayaert et al., 2005; Lehky and Sereno, 2007). Tested with solid shapes, as in most previous studies, a neuron might exhibit selectivity either because it is sensitive to local detail (e.g., the sharp corners of a diamond) or because it is sensitive to global configuration (e.g., the overall form of a diamond). That shape selectivity depends on local details has been suggested by studies demonstrating sensitivity to changes in a restricted part of an image (Baker et al., 2002; Brincat and Connor, 2004). That it depends as well on overall form has been suggested by the observation that IT neurons remain selective for faces even after substantial blurring (Rolls et al., 1985, 1987). Using stimuli that allow global and local content to be varied independently, we have obtained firm new evidence for sensitivity to overall form as distinct from local detail.

We found that at least 24% of IT neurons are sensitive to the identity of local elements in a hierarchical display. These neurons do not simply represent texture (Sáry et al., 1993; Kovács et al., 2003; Köteles et al., 2008) but rather represent the shape of the local elements. This is indicated by the significant positive correlation across neurons between the preferred local element in a hierarchical figure and the preferred unitary small solid shape (Fig. 6). It is an interesting question whether, as the elements in a hierarchical display become progressively smaller and more numerous, there is a limit beyond which IT neurons still differentiate between hierarchical figures containing different local elements but represent local content in a texture-based rather than a shape-based code.

We found that neurons preferring a given large solid shape tend to prefer the corresponding small solid shape (Fig. 6). Thus there is a tendency for shape selectivity to be scale invariant.

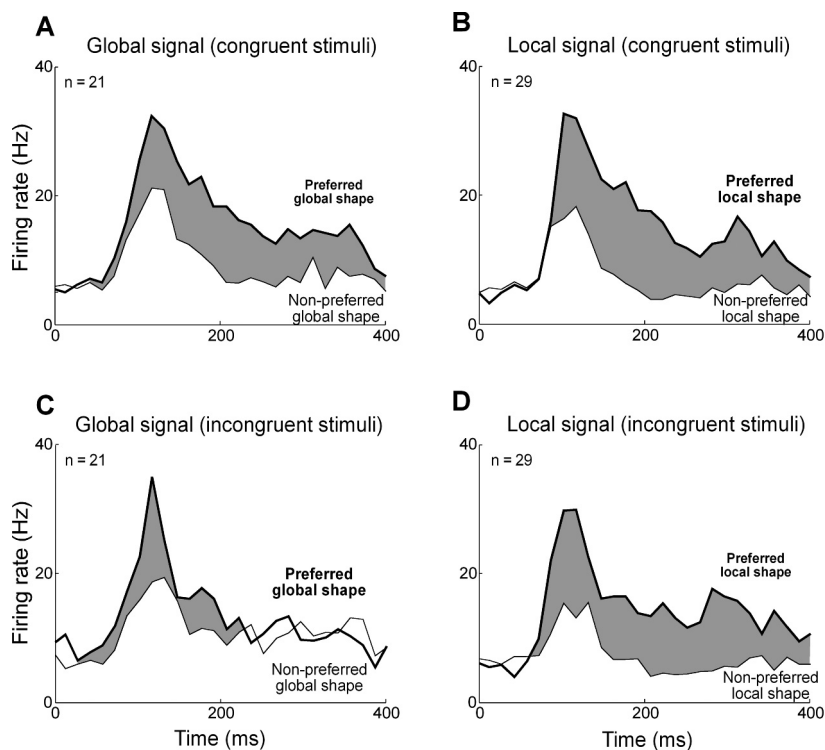


Figure 7. Time course of population responses to congruent and incongruent hierarchical displays. **A**, Average response of 21 globally selective neurons to their preferred global shape (thick lines) and nonpreferred global shape (thin lines) for congruent displays. **B**, Corresponding plot for responses of 29 locally selective neurons. **C**, Average responses of 21 globally selective neurons to their preferred and nonpreferred global shapes for incongruent displays. **D**, Corresponding plot for 29 locally selective neurons. The width of the gray ribbon in each plot corresponds to the strength of the shape signal.

However, neurons selective for a shape at a given hierarchical level (global or local) tend to exhibit matching selectivity at the corresponding size (large or small) more than at the other size (small or large) (Fig. 6). These observations are compatible with previous reports, which have emphasized that shape selectivity is scale invariant over a range of 1° to 12° in size but have also noted that the pattern can break down at extremes of the range (Sato et al., 1980; Lueschow et al., 1994; Ito et al., 1995; Hikosaka, 1999; Brincat and Connor, 2004; Zoccolan et al., 2005). The global ($\sim 8^\circ$) and local ($\sim 1^\circ$) shapes in our study are near these extremes.

The key finding that global shape signals, although weaker than local signals, develop at shorter latency is surprising. Latency, if it varies across stimulus categories, generally varies inversely with response strength. For example, extrastriate neurons respond more strongly and earlier to primate versus nonprimate faces (Kiani et al., 2005), to upright versus inverted faces (Tsao et al., 2006), and to high-contrast versus low-contrast images (Lee et al., 2007). Likewise, signals discriminating the category of a face (human or monkey) are stronger and develop earlier than signals discriminating between individuals in a category (Sugase et al., 1999). In each of these cases, the response might develop earlier simply because, being stronger, it crosses the neuronal activation threshold earlier. The early onset of global signals requires another form of explanation. One possibility is that low-spatial-frequency information (sufficient to define global but not local shape) arrives in IT earlier than high-spatial-frequency information. Neuronal visual response latency in primary visual cortex has been reported to decline by ~ 10 ms per octave as spatial frequency increases (Frazor et al., 2004). This observation is com-

patible with our finding that global and local shapes defined at scales differing by around three octaves elicit discriminative responses offset in time by ~ 30 ms. It also fits with the observation that high-pass filtering abolishes the global advantage effect in humans (Badcock et al., 1990; LaGasse, 1993). However, we cannot rule out a role for image attributes distinct from spatial frequency but correlated with it in our stimulus set. These include the size of the shape (large and small for shapes in which low and high spatial frequencies predominate) and the location of the shape's center (foveal and peripheral for shapes in which low and high spatial frequencies predominate). To assess the impact of these factors would require the use of images in which they vary independently of spatial frequency.

There is striking agreement between the lead time of the global signal in monkey IT (28 ms) and the global advantage effect in humans processing the same stimuli (29 ms). However, to conclude that monkey physiology explains human performance would require answering three questions in the affirmative.

Do monkeys exhibit a global advantage effect?

Nonhuman primates have been reported to base their responses to hierarchical stimuli preferentially on local content (Fagot and Deruelle, 1997; Fagot and Tomonaga, 1999; Spinozzi et al., 2004, 2006; De Lillo et al., 2005). However, the reports are based on tasks that place no premium on response speed. There has been one behavioral study of nonhuman primates (macaque monkeys) performing a speeded response task closely modeled on the classic (Navon, 1977) human paradigm. This task did reveal faster responses to global shape (Tanaka and Fujita, 2000).

Is discriminative signal latency the same during active as during passive processing?

At least two factors absent during passive viewing could affect neuronal visual responses during active processing of hierarchical stimuli. (1) Shape discrimination training is known to induce a subtle improvement in the ability of neurons to differentiate between stimuli (Kobatake et al., 1998; Baker et al., 2002; Rainer et al., 2004; Freedman et al., 2006; Cox and DiCarlo, 2008). However, it has never been observed to affect the latency of visual responses or of discriminative signals. (2) Attention—in this case to the global or local level of the display—might affect the neuronal firing rate (Chelazzi et al., 1993; Desimone, 1996). However, attention does not affect visual response latency (Lee et al., 2007). It is reasonable therefore to expect latency to be unaffected by active processing.

Do monkey and human brains process hierarchical stimuli in the same way?

The representation of hierarchical stimuli is lateralized to some degree in the human brain, with global and local processing preferentially represented in the right and left hemispheres (Robertson and Lamb, 1991; Fink et al., 1996; Hübner, 1998; Han et al., 2002). Monkeys do exhibit some signs of functional lateralization (Hamilton and Vermeire, 1988; Hauser, 1993; Vallortigara and Rogers, 2005) but do not exhibit right hemisphere global specialization (Tanaka et al., 2001). That there is a species difference in hemispheric specialization does not, however, necessarily argue for a difference in the relative timing of global and local signals.

Coarse-to-fine processing

It has been proposed that object recognition depends on a system of coarse-to-fine processing in which the coarse organization of an image determines a subset of possible interpretations and fine

details then provide disambiguating information. This scheme was originally put forward as a means for computing depth from disparity (Menz and Freeman, 2003), but has since been applied to object recognition in general (Morrison and Schyns, 2001; Bar et al., 2006). Supporting evidence includes the observation that humans process visual images more efficiently when the content at different spatial frequencies is presented in low-to-high than in high-to-low order (Schyns and Oliva, 1994; Parker et al., 1997). The key finding of the present study—that neurons in IT carry a coarse representation of shape before they become sensitive to finer detail—is compatible with the coarse-to-fine model and indeed constitutes support for it on the assumption that such a temporal offset would be unlikely to have developed in primate evolution without contributing to visual perception.

References

- Badcock JC, Whitworth FA, Badcock DR, Lovegrove WJ (1990) Low-frequency filtering and the processing of local-global stimuli. *Perception* 19:617–629.
- Baker CI, Behrmann M, Olson CR (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci* 5:1210–1216.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A* 103:449–454.
- Baylis GC, Rolls ET (1987) Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Exp Brain Res* 65:614–622.
- Behrmann M, Avidan G, Leonard GL, Kimchi R, Luna B, Humphreys K, Minshew N (2006) Configural processing in autism and its relationship to face processing. *Neuropsychologia* 44:110–129.
- Bihle AM, Bellugi U, Delis D, Marks S (1989) Seeing either the forest or the trees: dissociation in visuospatial processing. *Brain Cogn* 11:37–49.
- Brincat SL, Connor CE (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7:880–886.
- Chelazzi L, Miller EK, Duncan J, Desimone R (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363:345–347.
- Cox DD, DiCarlo JJ (2008) Does learned shape selectivity in inferior temporal cortex automatically generalize across retinal position? *J Neurosci* 28:10045–10055.
- De Lillo C, Spinozzi G, Truppa V, Naylor DM (2005) A comparative analysis of global and local processing of hierarchical visual stimuli in young children (*Homo sapiens*) and monkeys (*Cebus apella*). *J Comp Psychol* 119:155–165.
- Desimone R (1996) Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci U S A* 93:13494–13499.
- Fagot J, Deruelle C (1997) Processing of global and local visual information and hemispheric specialization in humans (*Homo sapiens*) and baboons (*Papio papio*). *J Exp Psychol Hum Percept Perform* 23:429–442.
- Fagot J, Tomonaga M (1999) Global and local processing in humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*): use of a visual search task with compound stimuli. *J Comp Psychol* 113:3–12.
- Fink GR, Halligan PW, Marshall JC, Frith CD, Frackowiak RS, Dolan RJ (1996) Where in the brain does visual attention select the forest and the trees? *Nature* 382:626–628.
- Frazor RA, Albrecht DG, Geisler WS, Crane AM (2004) Visual cortex neurons of monkeys and cats: temporal dynamics of the spatial frequency response function. *J Neurophysiol* 91:2607–2627.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2006) Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex* 16:1631–1644.
- Hamilton CR, Vermeire BA (1988) Complementary hemispheric specialization in monkeys. *Science* 242:1691–1694.
- Han S, Weaver JA, Murray SO, Kang X, Yund EW, Woods DL (2002) Hemispheric asymmetry in global/local processing: effects of stimulus position and spatial frequency. *Neuroimage* 17:1290–1299.
- Hauser MD (1993) Right hemisphere dominance for the production of facial expression in monkeys. *Science* 261:475–477.
- Hikosaka K (1999) Tolerances of responses to visual patterns in neurons of

- the posterior inferotemporal cortex in the macaque against changing stimulus size and orientation, and deleting patterns. *Behav Brain Res* 100:67–76.
- Hübner R (1998) Hemispheric differences in global/local processing revealed by same-different judgements. *Vis Cogn* 5:457–478.
- Ito M, Tamura H, Fujita I, Tanaka K (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73:218–226.
- Johnson SC, Lowery N, Kohler C, Turetsky BI (2005) Global-local visual processing in schizophrenia: evidence for an early visual processing deficit. *Biol Psychiatry* 58:937–946.
- Kayaert G, Biederman I, Vogels R (2005) Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cereb Cortex* 15:1308–1321.
- Kiani R, Esteky H, Tanaka K (2005) Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J Neurophysiol* 94:1587–1596.
- Kimchi R (1992) Primacy of wholistic processing and global/local paradigm: a critical review. *Psychol Bull* 112:24–38.
- Kinchla RA, Wolfe JM (1979) The order of visual processing: “top-down,” “bottom-up,” or “middle-out.” *Percept Psychophys* 25:225–231.
- Kobatake E, Wang G, Tanaka K (1998) Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80:324–330.
- Köteles K, De Mazière PA, Van Hulle M, Orban GA, Vogels R (2008) Coding of images of materials by macaque inferior temporal cortical neurons. *Eur J Neurosci* 27:466–482.
- Kovács G, Sáros G, Köteles K, Chadaide Z, Tompa T, Vogels R, Benedek G (2003) Effects of surface cues on macaque inferior temporal cortical responses. *Cereb Cortex* 13:178–188.
- LaGasse LL (1993) Effects of good form and spatial frequency on global precedence. *Percept Psychophys* 53:89–105.
- Lee J, Williford T, Maunsell JHR (2007) Spatial attention and the latency of neuronal responses in macaque area V4. *J Neurosci* 27:9632–9637.
- Lehky SR, Sereno AB (2007) Comparison of shape encoding in primate dorsal and ventral visual pathways. *J Neurophysiol* 97:307–319.
- Liu Y, Murray SO, Jagadeesh B (2009) Time course and stimulus dependence of repetition-induced response suppression in inferotemporal cortex. *J Neurophysiol* 101:418–436.
- Lueschow A, Miller EK, Desimone R (1994) Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex* 4:523–531.
- McMahon DBT, Olson CR (2007) Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. *J Neurophysiol* 97:3532–3543.
- Menz MD, Freeman RD (2003) Stereoscopic depth processing in the visual cortex: a coarse-to-fine mechanism. *Nat Neurosci* 6:59–65.
- Miller EK, Li L, Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254:1377–1379.
- Miller J, Navon D (2002) Global precedence and response activation: evidence from LRPs. *Quart J Exp Psychol A* 55:289–310.
- Morrison DJ, Schyns PG (2001) Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychon Bull Rev* 8:454–469.
- Navon D (1977) Forest before trees: the precedence of global features in visual perception. *Cogn Psychol* 9:353–383.
- Navon D (2003) What does a compound letter tell the psychologist’s mind? *Acta Psychol (Amst)* 114:273–309.
- Navon D, Norman J (1983) Does global precedence really depend on visual angle? *J Exp Psychol Hum Percept Perform* 9:955–965.
- Parker DM, Lishman JR, Hughes J (1997) Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception* 26:1169–1180.
- Poirel N, Pineau A, Mellet E (2008) What does the nature of the stimuli tell us about the global precedence effect? *Acta Psychol (Amst)* 127:1–11.
- Rainer G, Lee H, Logothetis NK (2004) The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biol* 2:E44.
- Robertson LC, Lamb MR (1991) Neuropsychological contributions to theories of part/whole organization. *Cogn Psychol* 23:299–330.
- Rolls ET, Baylis GC, Leonard CM (1985) Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. *Vision Res* 25:1021–1035.
- Rolls ET, Baylis GC, Hasselmo ME (1987) The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces. *Vision Res* 27:311–326.
- Sáros G, Vogels R, Orban GA (1993) Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* 260:995–997.
- Sato T, Kawamura T, Iwai E (1980) Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discrimination. *Exp Brain Res* 38:313–319.
- Schwartz EL, Desimone R, Albright TD, Gross CG (1983) Shape recognition and inferior temporal neurons. *Proc Natl Acad Sci U S A* 80:5776–5778.
- Schyns PG, Oliva A (1994) From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychol Sci* 5:195–200.
- Slavin MJ, Mattingley JB, Bradshaw JL, Storey E (2002) Local-global processing in Alzheimer’s disease: an examination of interference, inhibition and priming. *Neuropsychologia* 40:1173–1186.
- Spinozzi G, De Lillo C, Castelli S (2004) Detection of grouped and ungrouped parts in visual patterns by tufted capuchin monkeys (*Cebus apella*) and humans (*Homo sapiens*). *J Comp Psychol* 118:297–308.
- Spinozzi G, De Lillo C, Salvi V (2006) Local advantage in the visual processing of hierarchical stimuli following manipulations of stimulus size and element numerosity in monkeys (*Cebus apella*). *Behav Brain Res* 166:45–54.
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–873.
- Tanaka H, Fujita I (2000) Global and local processing of visual patterns in macaque monkeys. *Neuroreport* 11:2881–2884.
- Tanaka H, Onoe H, Tsukada H, Fujita I (2001) Attentional modulation of neural activity in the macaque inferior temporal cortex during global and local processing. *Neurosci Res* 39:469–472.
- Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109–139.
- Tanaka K (2000) Mechanisms of visual object recognition studied in monkeys. *Spat Vis* 13:147–163.
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674.
- Ungerleider LG and Mishkin M (1982) Two cortical visual systems. In: *Analysis of visual behavior* (Ingle DJ, Goodale MA, Mansfield RJW, ed) pp 549–586, Cambridge: MIT.
- Vallortigara G, Rogers LJ (2005) Survival with an asymmetrical brain: advantages and disadvantages of cerebral lateralization. *Behav Brain Sci* 28:575–589; discussion 589–633.
- Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci* 25:8150–8164.