

Global Image Dissimilarity in Macaque Inferotemporal Cortex Predicts Human Visual Search Efficiency

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

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

Finding a target in a visual scene can be easy or difficult depending on the nature of the distractors. Research in humans has suggested that search is more difficult the more similar the target and distractors are to each other. However, it has not yielded an objective definition of similarity. We hypothesized that visual search performance depends on similarity as determined by the degree to which two images elicit overlapping patterns of neuronal activity in visual cortex. To test this idea, we recorded from neurons in monkey inferotemporal cortex (IT) and assessed visual search performance in humans using pairs of images formed from the same local features in different global arrangements. The ability of IT neurons to discriminate between two images was strongly predictive of the ability of humans to discriminate between them during visual search, accounting overall for 90% of the variance in human performance. A simple physical measure of global similarity—the degree of overlap between the coarse footprints of a pair of images—largely explains both the neuronal and the behavioral results. To explain the relation between population activity and search behavior, we propose a model in which the efficiency of global oddball search depends on contrast-enhancing lateral interactions in high-order visual cortex.

Introduction

Finding a target in a visual scene can be easy (like finding a fruit in a tree) or difficult (like finding a face in a crowd). Classic accounts of visual search assumed that search is efficient (occurs speedily through parallel processing of all items in the display) only if the target and distractors differ from each other with regard to local features such as are represented in primary visual cortex (Treisman and Gelade, 1980; Treisman and Gormican, 1988; Treisman and Sato, 1990; Treisman, 2006). Subsequent accounts incorporated the idea that search efficiency may depend in a graded manner on the degree of difference between target and distractors as determined not only by local features represented in primary visual cortex but also by global attributes represented in high-order areas (Duncan and Humphreys, 1989, 1992; Hochstein and Ahissar, 2002; Wolfe and Horowitz, 2004). This idea has been difficult to assess because little is known about the representation of global image attributes in high-order visual cortex. To address this issue, we have performed a study based on the use of images that contain the same local features and differ only at the level of global organization. Such images may be difficult (Fig. 1*A,B*) or easy (Fig. 1*C*) for humans to tell apart during oddball search. We have asked, for such image pairs, whether the ability of humans to tell them apart during visual search is correlated with the ability of neurons in macaque inferotemporal cortex (IT) to discriminate between them.

Neurons in IT, unlike those in low-order visual areas, possess receptive fields large enough to encompass an entire image (Op De Beeck and Vogels, 2000) and are sensitive to the global arrangement of elements within the image (Tanaka et al., 1991; Kobatake and Tanaka, 1994; Messinger et al., 2005; Yamane et al., 2006). Population activity in IT discriminates better between some images than others (Op de Beeck et al., 2001; Allred et al., 2005; Kayaert et al., 2005a,b; De Baene et al., 2007; Kiani et al., 2007; Lehky and Sereno, 2007). Moreover, if a pair of images is well discriminated by population activity in IT then humans tend to characterize them as dissimilar (Allred et al., 2005; Kayaert et al., 2005b) and monkeys are able to tell them apart when comparing them across a delay (Op de Beeck et al., 2001). It might be supposed, in light of these observations, that population activity in IT should necessarily predict human search efficiency. However, this outcome is uncertain for two reasons. First, search for an item in an array may depend on brain mechanisms fundamentally different from those underlying inspection of a single item (Treisman and Gelade, 1980; Treisman and Gormican, 1988; Treisman and Sato, 1990; Treisman, 2006). Previous studies comparing monkey physiology and human behavior were based on behavioral tests involving the inspection of single items. Second, the representation of global image attributes in high-order visual cortex may differ between monkeys and humans. Previous studies comparing monkey physiology and human behavior used images differing with regard to local features and so did not touch on this issue (Allred et al., 2005; Kayaert et al., 2005b).

Materials and Methods

Neuronal experiments

Data collection

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

Received April 22, 2009; revised Oct. 27, 2009; accepted Dec. 3, 2009.

This work was supported by National Institutes of Health (NIH) Grants R01 EY018620 and P50 MH084053, and the Pennsylvania Department of Health through the Commonwealth Universal Research Enhancement Program. MRI was supported by NIH Grant P41 EB001977. 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, Mellon Institute, Room 115, 4400 Fifth Avenue, Pittsburgh, PA 15213. E-mail: sparun@cnbc.cmu.edu.

DOI:10.1523/JNEUROSCI.1908-09.2010

Copyright © 2010 the authors 0270-6474/10/301258-12\$15.00/0

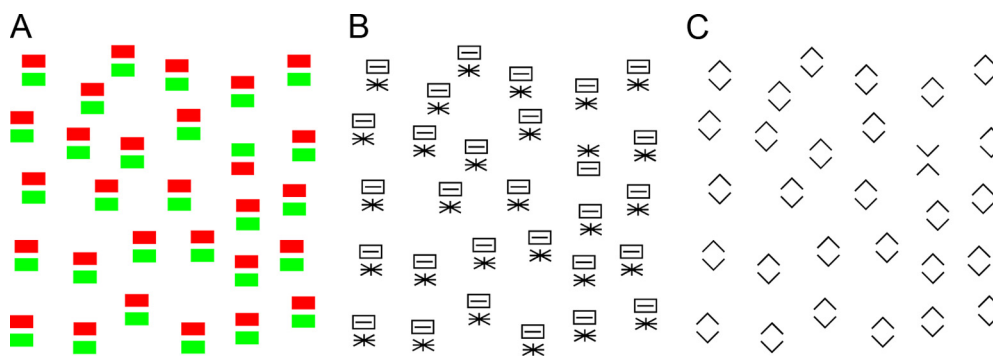


Figure 1. In some cases, but not all, a target containing two parts is easy to detect within a field of distractors containing the same two parts in transposed arrangement. In this example, the green-above-red target does not stand out among red-above-green distractors in **A**, nor does the asterisk-above rectangle target stand out among rectangle-above-asterisk distractors in **B**. In contrast, the X target stands out strongly among diamond distractors in **C**.

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. Before the recording period, each monkey was surgically fitted with a cranial implant and scleral search coils. After initial training, a 2 cm-diameter vertically oriented cylindrical recording chamber was implanted over the left hemisphere in both monkeys.

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 to a depth such that its tip was ~ 10 mm above IT. The electrode could be advanced reproducibly along tracks forming a square grid with 1 mm spacing. The action potentials of a single neuron were isolated from the multineuronal trace by means of a commercially available spike-sorting system (Plexon). All waveforms were recorded during the experiments and spike sorting was performed offline using cluster-based methods. Eye position was monitored by means of a scleral search coil system (Riverbend Instruments) and the x and y coordinates of eye position were stored with 1 ms resolution. All aspects of the behavioral experiment, including stimulus presentation, eye position monitoring, and reward delivery, were under control of a computer running Cortex software (NIMH Cortex). Monkeys were trained to fixate for ~ 2 s on a red fixation cross while a series of stimuli appeared in rapid succession (stimulus duration = 200 ms; interstimulus interval = 200 ms) at an eccentricity of 2° (see Fig. 3A). At the end of each trial, they were rewarded for correctly maintaining fixation by a drop of juice. Although the fixation window was large (4.2°), we found on *post hoc* analysis that the gaze remained closely centered on the fixation cross throughout the duration of the trial. The average across sessions of the SD of horizontal and vertical gaze angle was 0.2° . All recording sites were in the left hemisphere. At the end of the data collection period, they were established by 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 14–22 mm anterior to the interaural plane in Je and 9–18 mm anterior in Ec.

Data analysis

Database. All neuronal data analysis was based on the firing rate within a window extending from 50 to 300 ms after stimulus onset. A neuron was included in the database if and only if it was visually responsive as indicated by a significant difference (t test; $p < 0.05$) between its firing rate in this window and its firing rate in a 50 ms window centered on stimulus onset with data collapsed across all stimuli in an experiment.

Statistical comparison between image pairs. The key question, for each pair of images in every experiment, was how well neuronal activity discriminated between the two members of the pair. The index of discriminability was the mean across neurons of $|A - B|$ where A and B were the mean firing rates elicited by the two stimuli. This is the Neuron Index given in Table 1. To compare the discriminability of two image pairs, we applied a paired t test to distributions of index values obtained across the population of tested neurons. The p values reported in Table 1 reflect the outcome of this test.

Visual search experiments

Data collection

We collected and analyzed data from six adults, four male and two female, each of whom completed the entire battery of visual search experiments described below under a protocol approved by the Institutional Review Board of Carnegie Mellon University. In each experiment, the observer sat facing the screen with the right and left index fingers on two keys. The observer was instructed to maintain fixation on a central cross throughout each trial. The experiment consisted in responding with a key press to each of a succession of displays. On each trial, six stimuli appeared simultaneously at an eccentricity of 5° in a hexagonal array centered on fixation and arranged so that three items were to the right of fixation and three to the left at symmetric locations (see Fig. 3B). The observer was instructed to press the key on the same side as the oddball as quickly as possible without guessing. The display was presented continuously until a response had occurred or until 5 s had elapsed. At 5.3 s, if no response had occurred, the trial was aborted and the next trial began after 0.5 s.

Data analysis

Statistical comparison between image pairs. The key question was whether the search reaction time differed significantly between two pairs of images. Given a pair of images, A and B, we collapsed the data across cases in which the target was A and the distractor was B and cases in which the reverse was true. We did likewise for the second pair of images, C and D. The decision to collapse was justified by preliminary analysis indicating that there were no significant reaction time (RT) asymmetries. To assess the significance of the difference in RT between the two cases, we then performed a two-way ANOVA with RT as the dependent variable and with subject (six subjects) and image pair (AB or CD) as factors. The p values reported in Table 1 reflect the outcome of this test.

Behavioral discrimination index. The raw measure of discriminability between two images (the mean visual search RT) was smaller when the two images were more discriminable. For a more direct comparison to the neuronal data, we converted this to an index that increased with discriminability: $I_{\text{behav}} = (RT_{\text{oddball}} - RT_{\text{baseline}})^{-1}$, where RT_{oddball} was the mean across participants of the reaction time to report the side of the oddball and RT_{baseline} was the mean across participants of the reaction time to report the location of a single 2° diameter white disk presented 5° to the left or right of fixation in a separate block of trials. This is the Search Index given in Table 1. Computing this index is tantamount to computing the strength of the input fed to an integrator that triggers a behavioral response when its output crosses a fixed threshold (see Neural network model of visual search).

Set 1: Variable part identity

Stimuli. Each individual part (red or green rectangle, asterisk or slotted rectangle, or upward or downward pointing chevron) was 1° wide and 0.5° tall. In a compound display, the distance from the base of the top part to the apex of the bottom part (the measure of interpart distance used

Table 1. Summary of neuronal and behavioral results

Set	Neuron count	Condition	RT (ms)	Search index (s^{-1})	Neuron index (s^{-1})	Stat test	Search p value	Neuron p value
1	174	A. Color	1310	1.0	2.8	A vs B	—	—
		B. Pattern	1325	1.0	3.2	B vs C	***	***
		C. Chevron	614	3.5	4.8	A vs C	***	***
2	122	A. Far	831	2.0	3.8	A vs B	***	—
		B. Middle	638	3.2	4.1	B vs C	*	—
		C. Near	563	4.3	4.7	A vs C	***	*
3	114	A. Small	1045	1.4	3.1	A vs B	***	*
		B. Medium	654	3.1	3.9	B vs C	***	**
		C. Large	553	4.5	5.5	A vs C	***	***
4	128	A. Far	1299	1.0	3.0	A vs B	***	—
		B. Middle	1011	1.5	3.0	B vs C	***	*
		C. Near	706	2.7	3.6	A vs C	***	*
5	78	A. Tall	1554	0.8	2.6	A vs B	***	—
		B. Square	1072	1.3	2.6	B vs C	***	***
		C. Wide	685	2.8	4.4	A vs C	***	***
6	78	A. Tall	1544	0.8	2.6	A vs B	***	—
		B. Square	1072	1.3	2.6	B vs C	***	***
		C. Wide	679	2.9	4.3	A vs C	***	***
		Baseline	328					

Set and Condition: See Fig. 2. Neuron count: Number of neurons tested with the stimulus set. RT: Average across six subjects of the reaction time to report the side on which an oddball or the baseline stimulus (a single salient target) appeared. Search index: $1/(RT - B)$, where B is the baseline reaction time (328 ms)—increases with greater behavioral discriminability. Neuron index: The mean across neurons of the absolute difference between the firing rates elicited by the two members of the pair—increases with greater neuronal discriminability. Stat test: The two conditions being compared. Search p value: Statistical significance of the difference between the RTs for the two conditions (ANOVA with subject and image pair as factors). Neuron p value: Statistical significance of the difference between the neuronal discrimination indices for the two conditions (paired t test on the distribution of the two indices across neurons). For all significant pair-wise comparisons, the level of significance is indicated as: * $p \leq 0.05$, ** $p \leq 0.005$, or *** $p \leq 0.0005$.

throughout this paper) was 0.25° . The compound stimuli are shown in the top row of Figure 2.

Neuronal recording. The stimuli were presented at locations such that the center of a compound stimulus was 2° contralateral to fixation. For each of three cases (color, pattern, and chevron), there were six stimuli: the two compound stimuli with transposed arrangement, the two individual parts at the upper location and the two individual parts at the lower location. We presented individual parts as well as compound stimuli so as to determine whether there was a systematic relation between the responses to the two as discussed in supplemental Section 4, available at www.jneurosci.org as supplemental material. On a given trial, all six compound stimuli, all six parts at the upper location or all six parts at the lower location were presented. Each was presented for 200 ms followed by a 200 ms interstimulus interval. The order of the stimuli within the trial was random. During recording from a given neuron, each compound stimulus was presented 12 times, and each part was presented six times at the upper location and six times at the lower location.

Visual search. Observers completed a separate block of trials for each pair of stimuli. For each of three stimulus sets (color, pattern, and chevron), there was a block of trials. In each block, on randomly interleaved trials, each member of the pair appeared as an oddball at each location four times.

Set 2: Variable inter-chevron distance

Stimuli. The stimuli are shown in the second row of Figure 2. The individual chevrons were 1° wide. The distance between the chevrons (from the base of the top part to the apex of the bottom part) was 1.0° (far), 0.5° (middle), or 0.25° (near).

Neuronal recording. The stimuli were centered 2° contralateral to fixation. For each of three stimulus sets (far, middle, and near), there were two stimuli with transposed arrangement. On a given trial, all six stimuli were presented. Each was presented for 200 ms followed by a 200 ms interstimulus interval. The order of the stimuli within the trial was random. During recording from a given neuron, each stimulus was presented 12 times.

Visual search. Observers completed a separate block of trials for each pair of compound stimuli (far, middle, and near). In each block, on randomly interleaved trials, each member of the pair appeared as an oddball at each location four times.

Set 3: Variable chevron size

Stimuli. The stimuli are shown in the third row of Figure 2. The chevrons were 0.5° wide (small), 1° wide (medium), and 2° wide (large). The distance from the base of the top part to the apex of the bottom part was 0.5° in all cases. Selection of these dimensions ensured that the small, medium and large stimuli in this set were scaled versions of the far, middle, and near stimuli in Set 2.

Neuronal recording. With the exception that the stimuli differed in size rather than in distance, all procedures were the same as in experiment 2.

Visual search. With the exception that the stimuli differed in size rather than in distance, all procedures were the same as in experiment 2.

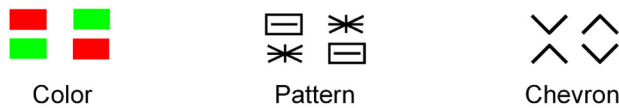
Set 4: Variable inter-contour distance

Stimuli. Representative stimuli are shown in the fourth row of Figure 2. Each part was constructed by stacking vertically two semicircles and a quarter-circle. The height of each part was 2.25° and the width 0.5° . Compound stimuli were constructed by juxtaposing the concave faces of two parts with a distance between the inner edges of 1° (far), 0.5° (middle), or 0.17° (near). At each distance, there were four stimuli: two symmetric bug configurations (quarter-circles at top or at bottom) and two asymmetric worm configurations (quarter-circles at top left and bottom right or bottom left and top right).

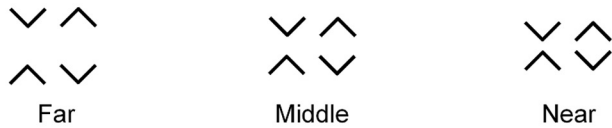
Neuronal recording. The stimuli were centered 2° contralateral to fixation. On a given trial, the four stimuli with the same interpart distance (far, middle, or near) appeared two times each in random order, subject to the constraint that no stimulus appeared twice in succession. Each stimulus was presented for 200 ms followed by a 200 ms interstimulus interval. Recording continued until each stimulus had been presented 12 times.

Visual search. At each distance (far, middle, or near), there were two symmetric bug-like variants (quarter-circles at top or at bottom) and two asymmetric worm-like variants (quarter-circles at top left and bottom right or at bottom left and top right). This made for eight target-distractor combinations (four possible targets crossed with two possible variants of the distractor). Participants completed a separate block of trials for each condition (far, middle, or near). The 48 randomly sequenced trials in each block conformed to the 48 cases obtained by crossing two vertical mirror-image variants of bug, two lateral mirror-image variants of worm, six target locations, and two

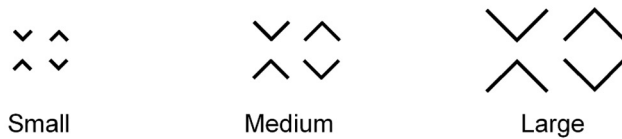
Set 1: Variable Part Identity



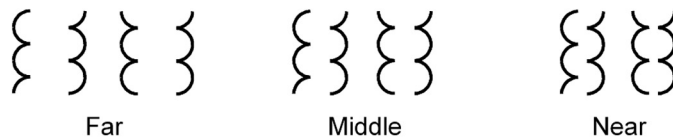
Set 2: Variable Inter-Chevron Distance



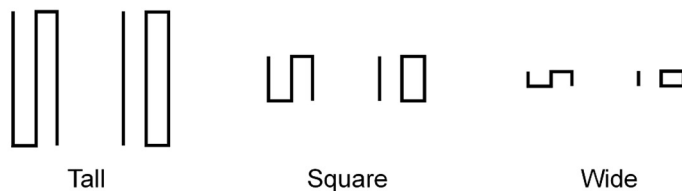
Set 3: Variable Chevron Size



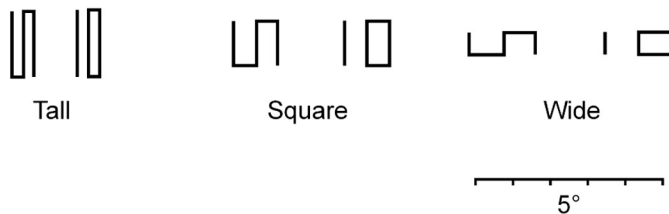
Set 4: Variable Inter-Contour Distance



Set 5: Variable Aspect Ratio (Width Constant)



Set 6: Variable Aspect Ratio (Perimeter Constant)



target identities (bug or worm). Thus each form (bug or worm) was presented as target four times at each location.

Sets 5–6: Variable aspect ratio

Stimuli. The stimuli were based on a design introduced by Julesz (1981). As noted by that author, the forms with S and IO topology represent different global arrangements of the same local elements. For example, each form contains two line terminators and four corners. As noted by subsequent authors, the discriminability of the two forms is affected by their aspect ratio (Enns, 1986; Tomonaga, 1999). The stimuli in the present experiment fell into five size groups across which the aspect ratio varied systematically (Table 2). Each size group contained four stimuli: two mirror images with S topology and two mirror images with IO topology. For data analysis, the five size groups were sorted into two sets (fifth and sixth rows of Fig. 2): set 5: tall, square, wide with width held constant (A, C, and E in Table 2); and set 6: tall, square, wide with perimeter held constant (B, C, and D in Table 2). Perimeter here refers to the rectangular envelope ($= 2 \times \text{height} + 2 \times \text{width}$).

Neuronal recording

The stimuli were centered 2° contralateral to fixation. On a given trial, a set of four stimuli from the same size group appeared two times each in random order, subject to the constraint that no stimulus appeared twice in succession. Each stimulus was presented for 200 ms followed by a 200 ms interstimulus interval. Recording continued until each stimulus had been presented 12 times.

Visual search

For each size group (A–E), there were two mirror-image variants with S topology and two mirror-image variants with IO topology. This made for eight target-distractor combinations (four possible targets crossed with two possible variants of the distractor for the given target). Participants completed a separate block of trials for each size group (A–E). The 48 randomly sequenced trials in each block conformed to the 48 cases obtained by crossing two mirror-image variants of the form with S topology, two mirror-image variants of the form with IO topology, six target locations and two target identities (S or IO). Thus each form (S or IO) was presented as target four times at each location.

Neural network model of visual search

To explore the possible nature of the causal linkage between visually selective neuronal ac-

Figure 2. Six sets of image pairs used for neuronal testing in macaques and behavioral testing in humans. Except for the chromatic pair, all were presented as white forms against a black background. The scale bar at the base applies to all images.

Table 2. Summary of stimuli in Sets 5 and 6

Size group	Aspect ratio	Height	Width
A	1:3 (tall)	3.6°	1.2°
B	1:3 (tall)	1.8°	0.6°
C	1:1 (square)	1.2°	1.2°
D	3:1 (wide)	0.6°	1.8°
E	3:1 (wide)	0.4°	1.2°

tivity and visual search reaction time, we developed a six-unit neural network (see Fig. 8). The six units had receptive fields centered at the locations of the six hexagonally arrayed visual-search stimuli. The level of activation of each unit was the linear sum of visually driven bottom-up excitatory input and lateral inhibitory input from the five other units. All units were selective for the same stimulus. On any given simulated trial, one unit had a target in its receptive field and five had distractors. Due to symmetry, the five units with distractors exhibited the same level of activation. Thus the network could be described by two equations, one

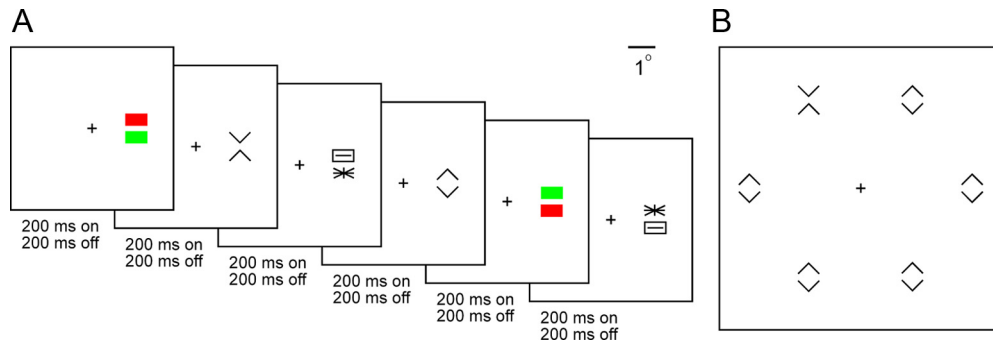


Figure 3. *A*, Neuronal recording experiments involved rapid serial presentation of randomly interleaved images centered 2° contralateral to fixation. Responses to the six image sets shown in Figure 1 were collected in separate blocks of trials. *B*, On each trial of a visual search session, a hexagonal array of images at 5° eccentricity was presented. One of the images was an oddball. The oddball and the distractors belonged to one of the pairs shown in Figure 2. The observer was required to press the right or left key according to whether the oddball was on the right or left. The difference in eccentricity between images in the neuronal and visual search experiments arose from the need to optimize the placement of the images in the neuron's receptive field during recording and to optimize the spacing of the items in the array during visual search.

for the activation of the unit with the target in its receptive field (A_1), and one for the activation of each of the five units with distractors in their receptive fields (A_2):

$$\begin{aligned} A_1 &= M + 0.5d - 5kA_2 \\ A_2 &= M - 0.5d - k(A_1 + 4A_2). \end{aligned} \quad (1)$$

The first term in each equation represents the average of the activations elicited by the target and the distractor. We set M to 13.7 spikes/s (the average strength of the visual response across all IT neurons and all stimuli). The second term represents bottom-up stimulus-selective visual excitation. If simulating the response of a network of X-selective neurons to a display in which the target was an X and the distractors were diamonds, we set d to 4.8 spikes/s (the average across all neurons of the measured difference in the firing rates elicited by an X and a diamond). If simulating the response of a network of diamond-selective neurons to the same display, we set d to -4.8 spikes/s. For other simulated conditions, we did likewise, always basing d on the average discriminative signal as measured in IT. The third term represents lateral inhibition. We set k to 0.1. The particular choice of the value of k was not critical to the qualitative pattern of results.

We fed the output of each unit to an integrator. The RT was taken as the time following stimulus onset at which the first integrator—the one driven by the most active unit in the network—crossed threshold. In the case with which we were concerned—the case in which the target was the stimulus preferred by the units in the network—the reaction time was given by the equation:

$$RT = B + q/(A_1 + c), \quad (2)$$

where B was the baseline response time, q was the threshold, and c was a heuristic constant. By transposing terms and using the definition of the behavioral discrimination index used in the visual search experiments [$I_{\text{behav}} = 1/(RT - B)$], we obtain the relation:

$$I_{\text{behav}} = (A_1 + c)/q. \quad (3)$$

Equations 1 and 3 together define the behavioral discrimination index, I_{behav} , as a function of the neuronal discrimination index, d , with two free parameters, q and c . We adjusted the free parameters (*lsqcurvefit* function, MatLab) to obtain the best fit among 17 measured values of I_{behav} and the values of I_{behav} generated by the model when given as input the 17 corresponding neuronal discrimination indices (see Fig. 5). The best fit was obtained with $q = 0.54$ and $c = -10.5$ spikes/s.

Fourier power difference index

For each of the two members of an image pair, we first computed an orientation and spatial frequency power spectrum. This yielded a value for power at each point in a rectangular grid spanning two-dimensional Fourier space. To these points we applied a Cartesian-to-polar transformation. Then we computed by interpolation the power at each point on

a rectangular grid in the transformed space. This grid spanned 64 spatial frequencies from one cycle per frame to 64 cycles per frame in equal steps of 0.09 octave and 61 orientations from -90° to 90° in equal steps of 3° . Next, we blurred the power values to simulate the bandpass characteristics of V4 neurons (David et al., 2006). The blurring function was a Gaussian with a bandwidth (full width at half height) of 1.2 octaves in the spatial frequency domain and 7.7° in the orientation domain. This step eliminated fine-grained patterns, present in the spectra of simple geometric figures, that vary in a nonmonotonic manner under continuous variation of properties such as interpart distance. We then normalized the spectrum by scaling power at each point to the average across all points. This step, by ensuring that the volume under each surface had a value of 1, eliminated accidental effects due to factors such as contrast and brightness. We proceeded to subtract one spectrum from the other. Rectifying and integrating over the resulting surface provided a scalar measure of the difference between the Fourier spectra. The value could range from zero (for identical images) to two (for images—such as orthogonal gratings—with no overlap in Fourier power space).

Coarse footprint difference index

To calculate the coarse footprint difference index, we first low-pass filtered each image of a pair, using a Gaussian blur function with a SD 0.08 times the extent of the longer dimension. This reduced spectral power by $>80\%$ beyond a low-pass cutoff of 3 cycles per object. Next we normalized the volume under each image to a value of 1. Finally, superimposing the images with their centers of mass in alignment, we created a difference image by pixel-wise subtraction. Rectification and integration of the pixel values in the difference image yielded a scalar index of the difference between the images that could in principle range from 0 to 2.

Results

Correlated neuronal and behavioral measures of image discriminability

We performed parallel single-neuron recording experiments in monkeys and visual search experiments in humans using six stimulus sets each of which consisted of three pairs of images (Fig. 2). The images in each pair contained an identical collection of local elements but differed in the arrangement of those elements. Thus discrimination between the members of the pair depended on registering their global organization. Within each set, the pairs differed with regard to some variable likely to affect the discriminability of the two members of the pair, for example, in set 1, the identity of the parts and, in set 2, the interpart distance (Fig. 2).

In microelectrode recording experiments, we performed testing with images from each set in a different block of trials (Fig. 3A). Over the course of a given block, each image in the set was

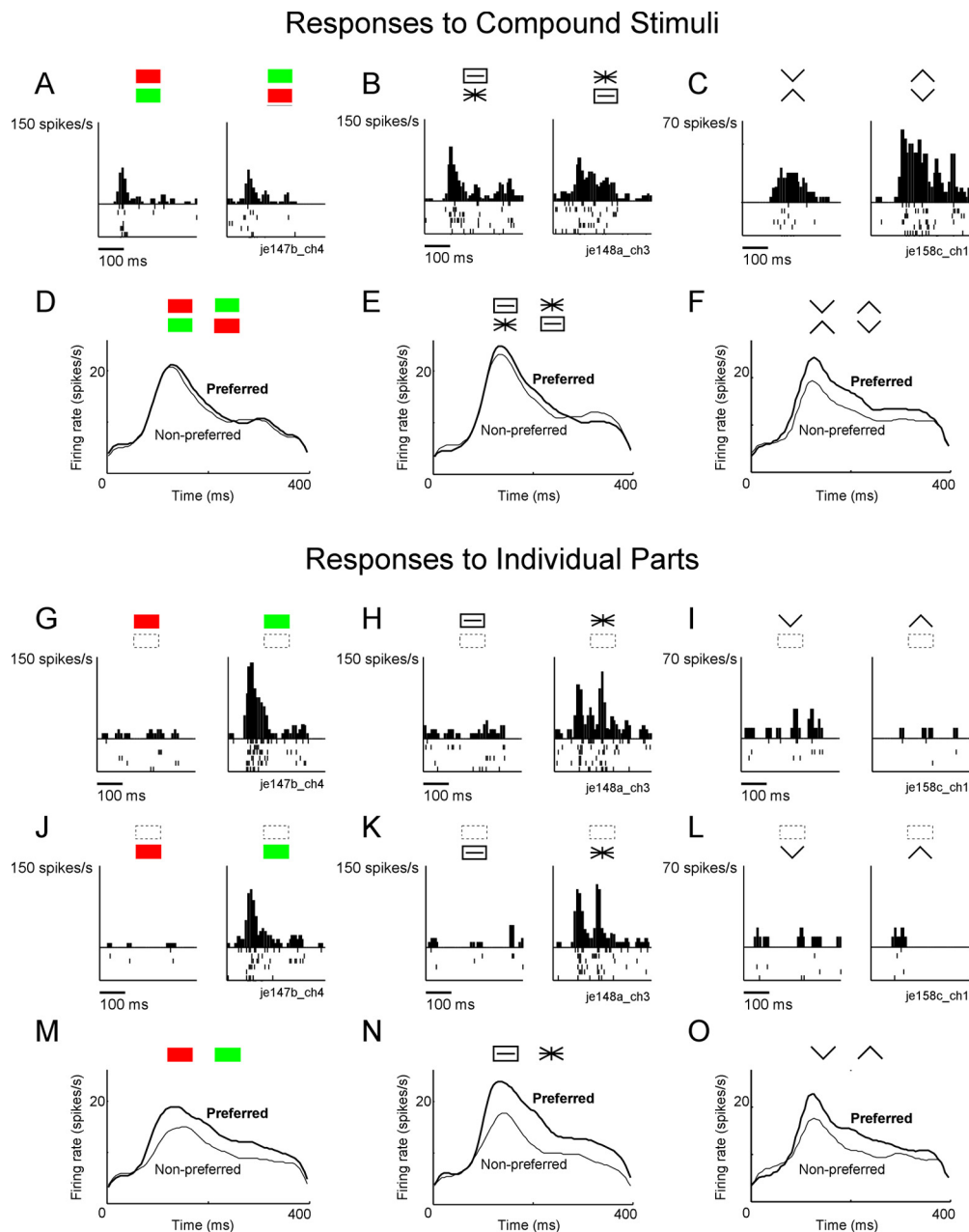


Figure 4. Examples of neuronal responses to the three pairs of compound images from set 1 and to the individual parts from which they were formed. **A–C**, Representative responses to the three pairs of compound images. Data are from different neurons. **D–F**, Average activity elicited among a population of 174 neurons, by the preferred and nonpreferred members of each pair. Note that the red–green compounds and the asterisk–rectangle compounds are poorly discriminated relative to chevron compounds. **G–I**, Representative responses to the three pairs of parts when presented at the upper location (**G–I**) and the lower location (**J–L**). Data are from the same neurons as in **A–C**. **M–O**, Average activity elicited among a population of 174 neurons by the preferred and nonpreferred parts. Note that the red and green rectangles and slotted rectangle are at least as well discriminated as the upward and downward pointing chevrons. Preferred and nonpreferred stimuli were identified on the basis of one half of trials and the population plot was based on the other half. Thus, if neuronal responses were unselective, the preferred and nonpreferred curves would overlap. Stimulus onset coincides with the left edge of each panel.

presented 12 times. The number of neurons tested with each set is indicated in Table 1 (Neuron count). The number varied from set to set because some neurons could not be held long enough for testing with all sets. Neurons in IT responded differentially to stimuli consisting of the same local elements in different global arrangements (Fig. 4). The selective responses were genuinely based on global arrangement as indicated by the fact that responses to parts could not predict responses to wholes (supplemental Section 1, available at www.jneurosci.org as supplemental material). Furthermore, they were determined by relatively abstract global properties as indicated by the fact that a given ar-

angement was preferred consistently over substantial changes in the size of the parts and the distance between them (supplemental Section 2, available at www.jneurosci.org as supplemental material). The strength of neuronal selectivity varied across image pairs. For example, in set 1, neurons differentiated poorly between compound images formed from colored parts (Fig. 4A, D) and patterned parts (Fig. 4B, E) but discriminated well between compound images formed from chevrons (Fig. 4C, F). As a measure of the ability of population activity in IT to discriminate between the images in each pair, we computed the average, across all visually responsive neurons in both monkeys, of the absolute

difference between the mean firing rates elicited by the two images. The resulting neuronal discrimination index is provided in Table 1 (Neuron index). All major trends demonstrated by this approach were confirmed by comparing counts of neurons showing individual significant effects (supplemental Section 3, available at www.jneurosci.org as supplemental material) and were found to be present in data from each monkey considered individually (supplemental Section 4, available at www.jneurosci.org as supplemental material). Population histograms demonstrated qualitatively the same trends as seen in the quantitative analysis (supplemental Section 5, available at www.jneurosci.org as supplemental material).

In visual search experiments involving six human participants, we measured the mean reaction time to indicate the location (left or right visual field) of an oddball presented among distractors, where the oddball and the distractors were the two members of a pair (Fig. 3B). We collected data in a separate block of trials for each pair of images. In the course of a block, each member of the pair appeared as an oddball four times at each location. Oddball search was easy for some pairs of images and difficult for others. For example, in set 1, the images formed from chevrons (Fig. 1C) popped out from each other whereas the images formed from colored and patterned parts (Fig. 1A,B) did not (for a demonstration that X and diamond stimuli popped out from each other according to the classical criterion for popout, see supplemental Section 6, available at www.jneurosci.org as supplemental material). We also collected data in a baseline block of trials that required reporting the location (right or left visual field) of a single salient stimulus. The mean oddball reaction time for each pair and the mean baseline reaction time are presented in Table 1 (RT). For each image pair, we computed a behavioral discrimination index that became larger the more discriminable two images were and that was consequently directly comparable to the neuronal discrimination index. This had the form $1/(RT - B)$ where RT was the mean oddball reaction time for a given image pair and B was the mean baseline reaction time. This index is equal to the strength of the difference signal that would have to be fed to an integrator starting at time 0 in order for its output to cross a fixed threshold and trigger a behavioral response at time RT (see Materials and Methods). The behavioral discrimination index for each pair is reported in Table 1 (Search index).

It is evident from comparison of the neuronal and behavioral discrimination indices (Table 1) that they tended to vary in parallel across image pairs. To quantify this effect, we computed the correlation between the two measures across the entire set of

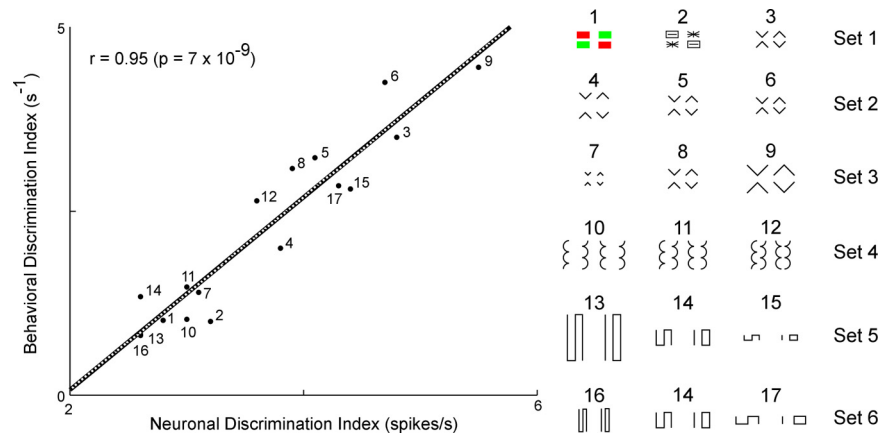


Figure 5. The behavioral discrimination index is plotted against the neuronal discrimination index for all 17 tested image pairs. The point corresponding to each pair is indicated by a number keyed to the display of images at the right. The best fit line is indicated in black. The output of a best-fit model incorporating contrast-enhancing lateral inhibition (Fig. 8) is indicated by the superimposed dotted white curve.

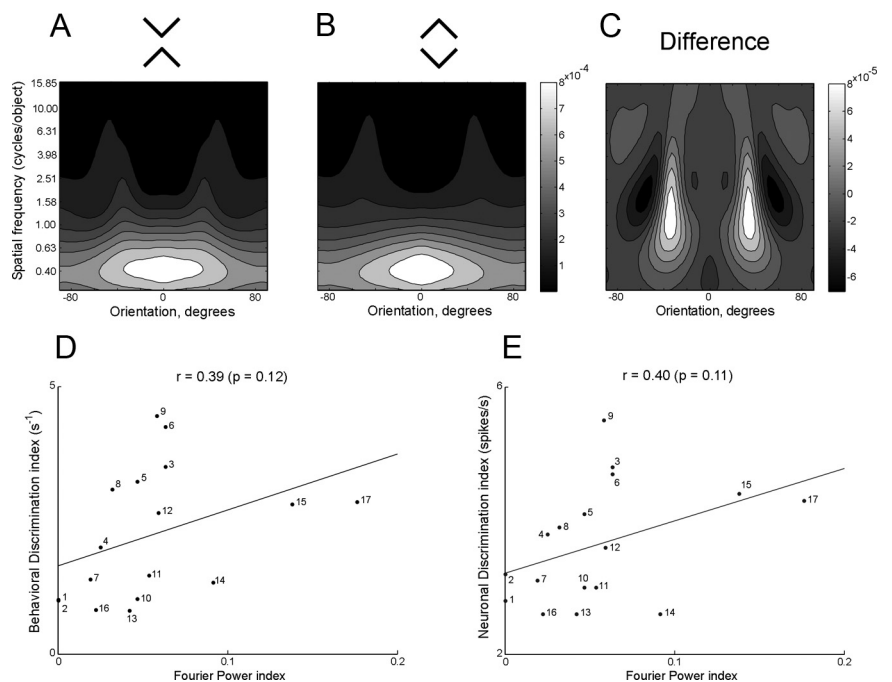


Figure 6. A–C, Blurred normalized Fourier power spectra for X and diamond images and the difference spectrum obtained by subtracting the diamond spectrum from the X spectrum. Note that the brightness scales are not the same. D, E, The behavioral and neuronal discrimination indices as measured experimentally are plotted against the Fourier power difference index for all 17 image pairs. The numbers appended to the points have the same significance as in Figure 5. Note that for the color and pattern pairs from set 1 (points 1 and 2) are related by vertical mirror reflection and therefore have zero Fourier power difference.

image pairs (Fig. 5). The correlation was strongly positive ($r = 0.95$) and highly significant ($p < 0.00000001$). The precision of the correspondence is especially striking because (1) the physiological results for different image sets were obtained from populations of neurons that did not fully overlap, (2) data for each set, even when collected from the same neurons, were collected in different blocks of trials, which might conceivably have produced contextual effects, and (3) there was no prior reason to suppose that the relation between the indices would be linear. This outcome was robust across changes in the epoch during which the firing rate was computed, changes in the properties of the neurons selected for study and changes in the metric used to characterize

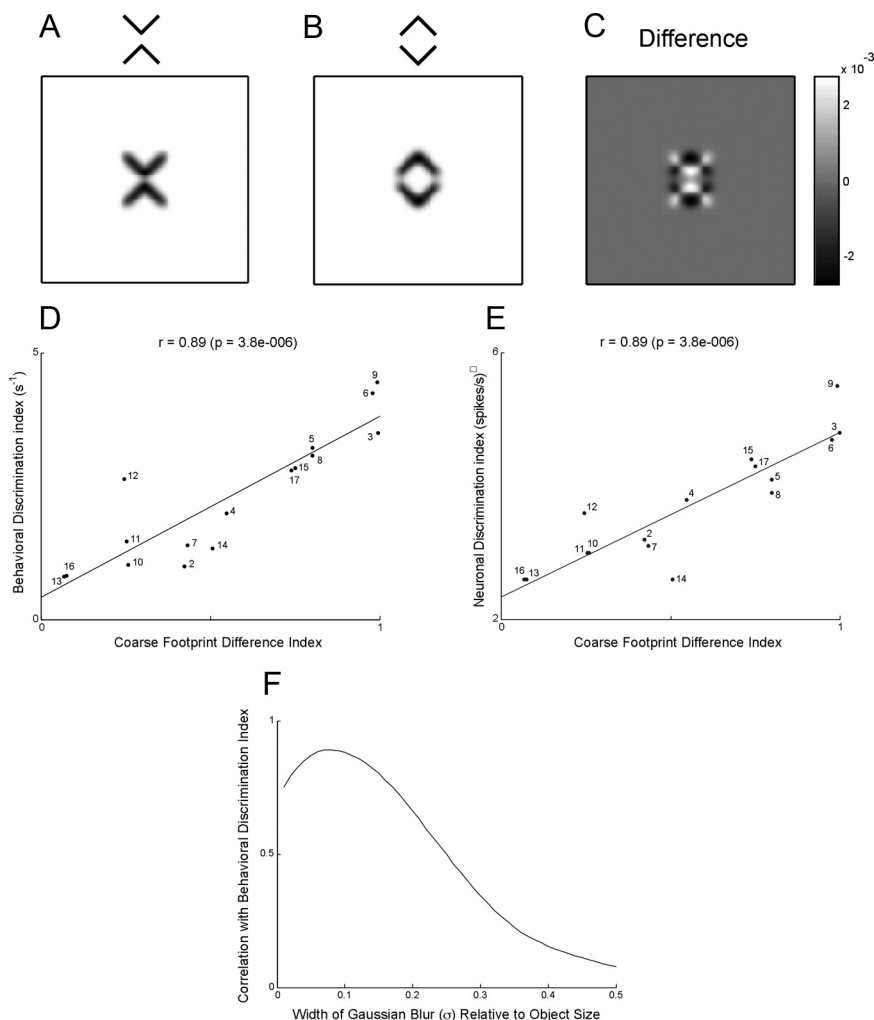


Figure 7. *A–C*, Coarse footprints of X and diamond images and the difference image obtained by pixel-wise subtraction of diamond from X. *D, E*, The behavioral and neuronal discrimination indices as measured experimentally are plotted against the coarse footprint difference index for 16 image pairs. The numbers appended to the points have the same significance as in Figure 5. Note that the color pair from set 1 (point 1) was excluded from the analysis because the coarse footprint index can be straightforwardly calculated only for monochromatic images. *F*, Correlation between behavioral and neuronal discrimination as a function of degree of blur.

neuronal selectivity (supplemental Section 7, available at www.jneurosci.org as supplemental material). We conclude that the ability of humans to discriminate between image pairs with differing global organization closely parallels the ability of IT neurons to discriminate between them.

A physical metric predicting image discriminability

Why are some pairs of images differing in global arrangement well discriminated while others are not? Is there some identifiable metric of the difference in global arrangement between two images that can explain this outcome? We assessed the ability of two metrics of image difference to explain the data. One is based on the Fourier power spectrum and the other on the distribution of brightness across pixels in image space.

Fourier power spectrum

The discriminability of images differing in global organization might be proportional to the difference in their Fourier power spectra. This would be consistent with the idea proposed for area V4 that pattern selectivity arises from neurons' possessing restricted receptive fields in orientation and spatial frequency space

(David et al., 2006, 2008). To explore this idea, we first computed for each pair of images used in this study a map of power in Fourier space (Fig. 6*A, B*). Next, we generated a difference map by subtraction (Fig. 6*C*). Finally, by rectifying and integrating across the difference map, we derived a scalar index of the difference between the images that could in principle range from 0 to 2. Further details are given in Materials and Methods. Upon comparing the Fourier difference index to the neuronal and behavioral discrimination indices across image pairs (Fig. 6*D, E*), we discovered that the correlation, although positive, was weak and did not attain significance (neuronal index: $r = 0.40$, $p = 0.11$; behavioral index: $r = 0.39$, $p = 0.12$).

Coarse footprint

The discriminability of images differing in global organization might be inversely proportional to the degree to which their footprints overlap when superimposed. Differences measured in image space, unlike those measured in the Fourier power domain, contain information about spatial phase. A pixel-based measure could thus explain the observation that IT neurons are able to discriminate between images that differ only with regard to spatial phase—for example mirror images (Rollenhagen and Olson, 2000). To explore this idea, we first low-pass filtered each image of a pair, using a Gaussian blur function with a SD equal to 0.08 times the longer dimension (Fig. 7*A, B*). This reduced the spectral power by >80% beyond spatial frequencies of 3 cycles per object. Next we created a difference image by pixel-wise subtraction (Fig. 7*C*), and rectified and integrated it to obtain a difference index that could range from 0 to 2

(see Materials and Methods). The full results are provided in supplemental Materials (Section 10, available at www.jneurosci.org as supplemental material). Plotting the neuronal and behavioral discrimination indices against this index across image pairs (Fig. 7*D, E*) revealed highly significant positive correlations (neuronal index: $r = 0.89$, $p < 0.000005$; behavioral index: $r = 0.89$, $p < 0.000005$). The strength of the correlations was reduced by blurring the images with Gaussian blur functions having SDs greater or less than 0.08 (Fig. 7*F*). We conclude that a simple measure based on the difference between the coarse (3 cycle per object) footprints of two images predicts the ability of IT neurons and humans engaged in visual search to detect a difference in global organization.

A model linking neuronal discrimination to behavioral discrimination

There is a simple mechanism by which better discrimination between a pair of images at the level of IT could give rise to a shorter reaction time for detecting one shape among distractors having

the other shape. It is based on stimulus-specific spatial-contrast-enhancing lateral interactions between neurons responding to images at different locations. Neurons throughout the visual system respond more strongly to a preferred stimulus presented in the classic receptive field if it is different from items in the surround than if it is the same (Lee et al., 2002; Constantinidis and Steinmetz, 2005). Surround modulation in low-order visual areas is conditional on the nature of low-order features such as orientation (Li, 1999; Nothdurft et al., 1999; Bair et al., 2003). It is possible, by analogy, that surround modulation in high-order visual areas depends on global attributes for which the neurons in those areas are selective. The effect would be to enhance the response to an oddball distinguished from distractors by its global attributes, with the degree of enhancement dependent on the degree to which neurons differentiate between the two.

As proof of principle, we studied a network consisting of six units selective for the same image, with receptive fields at the six locations of the hexagonal search array (Fig. 8). The level of activation of each unit is the sum of its bottom-up excitatory input and lateral inhibitory input originating from the other five neurons (Materials and Methods, Neural net model of visual search). If the oddball is the preferred stimulus and the distractors are nonpreferred stimuli, then the level of activation of the unit with the oddball in its receptive field will depend on (1) the strength of bottom-up excitatory input driven by the presence of the preferred stimulus in its receptive field and (2) the strength of lateral inhibition driven indirectly by the presence of the nonpreferred stimulus in the receptive fields of the other units. The greater the discriminative capacity of the units (the greater the difference between bottom-up excitation elicited by the preferred stimulus and bottom-up excitation elicited by the nonpreferred stimulus) the greater the level of activation of the unit with the oddball in its receptive field. The strength of activation can be transformed into a reaction time by accumulating it through an integrator until a decision threshold is reached. The reaction time is shorter for stimuli well discriminated by the units (Fig. 8*A*) than for poorly discriminated stimuli (Fig. 8*B*). The model thus provides a mechanistic transformation of neuronal discrimination ability into reaction time. Fitting the two free parameters of the model to data obtained with 17 image pairs yielded an extremely good fit (Fig. 5, dotted white curve superimposed on best-fit line). The assumptions on which this demonstration rests—that there is stimulus-specific lateral inhibition and that a population of neurons with variable discriminative capacity can be modeled by a few neurons with discriminative activity equal to the average across the population—remain to be tested. Nevertheless the demonstration makes clear that a simple mechanism based on contrast-enhancing lateral inhibition could account causally for the relation between neuronal discriminability in IT and human visual search efficiency as observed in these experiments.

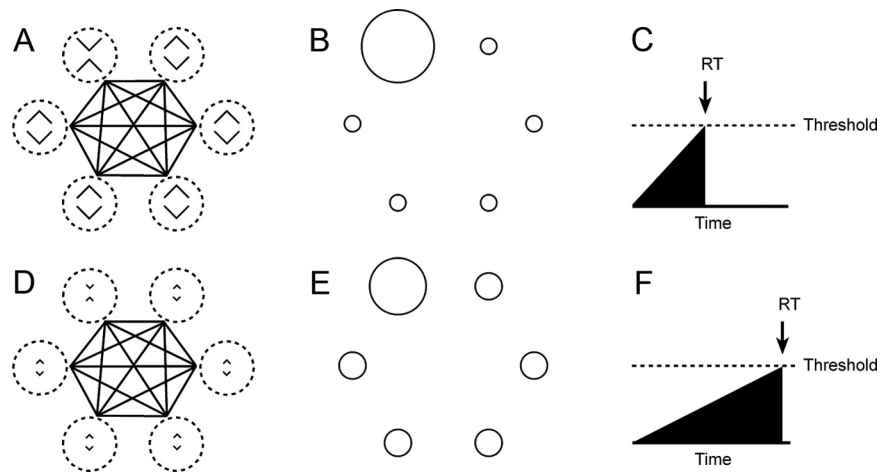


Figure 8. A plausible mechanistic link between the neuronal discrimination signal and visual search reaction time is afforded by a model incorporating contrast-enhancing lateral inhibition. **A**, The model consists of six units with receptive fields (hatched circles) centered on the locations of the six items in the search array (in this case an X oddball among diamond distractors). The units mutually inhibit each other (as indicated by the black lines connecting the receptive fields). **B**, Responses of the network to the display shown in **A** on the assumption that the units are X-selective. The unit with the oddball in its field is strongly active because (1) it receives strong bottom-up excitation from the X in its receptive field and (2) the other units, which inhibit it, receive only weak bottom-up excitation from the diamonds in their receptive fields. **C**, Strong activation from the unit with the oddball in its receptive field, fed to an integrator, causes the output of the integrator to rise rapidly toward a decision threshold with the result that the reaction time (RT) is short. **D–F**, When the same units are stimulated with a display in which the X and diamonds are less discriminable, the unit with the oddball in its field is less strongly active because the other units, which inhibit it, receive stronger bottom-up excitation from the diamonds in their receptive fields. Consequently, the output of the integrator rises more slowly toward a decision threshold with the result that the RT is longer.

Discussion

We have assessed the ability of neurons in monkey IT and of humans engaged in visual search to discriminate between images that differ exclusively at the level of global organization which we define in the following manner. Imagine systematically scanning two images with a window of fixed diameter and characterizing each image as the collection of details seen through the window without regard to the specific location of any detail. A pair of images differs solely at the global level if, to detect a difference between them, it is necessary to scan them with a relatively large window. For the image pairs in Figure 1, discrimination would be just possible with a window having a diameter one quarter of the image's height and would become progressively more robust as the diameter increased beyond that limit. The first essential finding of this study is that IT neurons are more sensitive to some global differences than to others. The second essential finding is that the ability of IT neurons to discriminate between a pair of such images is correlated with the ability of humans to discriminate between them during visual search.

Why are some global differences better discriminated than others?

The mere demonstration of a correlation between behavioral discrimination and neural discrimination is generally considered a sufficient endpoint in such studies (Edelman et al., 1998; Op de Beeck et al., 2001; Allred et al., 2005; Kayaert et al., 2005b; Haushofer et al., 2008). However, it is worthwhile to consider the underpinnings of the correlation by asking why, in physical terms, some image pairs are easier to discriminate than others. We explored this issue by applying to images in our experimental set two measures that are “global” in the sense that they depend on the distribution of content across the image as a whole but are also “low level” in the sense that they depend on convolving the

image with simple filters. One of these, based on the Fourier power spectrum, provided a poor fit to the data. The other, based on the coarse footprint, provided a good fit. Although the coarse footprint measure worked well for images in this particular set, it is only a first step toward a general account of discrimination based on global structure. To explain results obtained with high-pass-filtered images will require refining the measure (supplemental Section 8, available at www.jneurosci.org as supplemental material). Furthermore, any complete approach will have to take perceptual organization into account (Palmer, 1999; Kimchi et al., 2003). Neurons in IT are sensitive to figure-ground organization (Baylis and Driver, 2001). So are humans engaged in visual search (Enns and Rensink, 1990; He and Nakayama, 1992; Davis and Driver, 1994; Rensink and Enns, 1998). Figure-ground organization may have played a role even in the present experiment. For example, the line segments in the X may have been construed as having figural status and the line segments in the diamond may have been construed as the boundaries of an enclosed figure.

Does the human homolog of IT guide search based on global attributes?

The simplest possible interpretation of the neuronal-behavioral correlation is (a) that humans possess an area homologous to IT, (b) that neurons in this area represent global image structure in a code equivalent to the code in IT, and (c) that humans base global search on activity in this area. These points seem plausible but each must be qualified. (a) The human lateral occipital and fusiform regions are generally regarded as homologous to macaque IT on the grounds of their location and object-selective BOLD responses (Grill-Spector et al., 2001; Tootell et al., 2003; Orban et al., 2004; Pinsk et al., 2009) but no homology is certain. (b) The idea that representations in lateral occipital and fusiform cortex are similar to those in IT remains to be established. One approach to characterize the object representation in humans is to assess the similarity in the BOLD activation patterns elicited by individual objects (Edelman et al., 1998; Williams et al., 2007; Haushofer et al., 2008). These similarity patterns in humans reveal a categorical structure that is correlated with the similarity patterns observed in monkey IT (Kriegeskorte et al., 2008). However, it is not clear whether this categorical structure reflects the geometry of objects or simply their category. (c) Even findings based on this approach would not clinch the argument that humans engaged in global search rely on representations in lateral occipital and fusiform cortex. This might be accomplished by demonstrating selective impairment of visual search based on global image attributes after occipitotemporal injury. Visual search is certainly impaired in patients with visual agnosia arising from occipitotemporal damage (Humphreys et al., 1992; Saumier et al., 2002; Kentridge et al., 2004; Ballaz et al., 2005; Foulsham et al., 2009). However, no tests have involved targets and distractors similar to the ones used in this study. We note finally that even if humans are guided by representations in IT-homologous cortex during visual search based on global attributes, it may still be the case that search based on local features depends on areas of lower order (supplemental Section 11, available at www.jneurosci.org as supplemental material).

What is the relation of these findings to classic theories of visual search?

Feature integration theory (FIT) as propounded by Treisman and colleagues (Treisman and Gelade, 1980; Treisman and Gormi-

can, 1988; Treisman and Sato, 1990; Treisman, 2006) has been the dominant theoretical framework for understanding visual search over recent decades. The data that we have presented and the model with which we have been able to fit the data are, however, contrary in spirit to FIT. Our results are best understood as adding to an accumulating body of evidence that calls into question four intertwined assumptions embodied in FIT.

Features versus conjunctions

FIT posits that search is efficient if the target is distinguished from the distractors by a unique low-level feature but not if it is distinguished by a conjunction of features. In fact search for some conjunctions is efficient and search for some features is not. Xs and diamonds (different conjunctions of identical parts and locations) (von der Malsburg, 1999) pop out from each other (supplemental Section 6, available at www.jneurosci.org as supplemental material). So do different conjunctions of the same spatial frequencies and orientations (Sagi, 1988). Conversely, a target distinguished by a unique feature will not pop out if the featural difference from the distractors is too small (Moraglia, 1989; Nagy and Sanchez, 1990; Bauer et al., 1996; Nagy and Cone, 1996).

Striate versus extrastriate cortex

FIT posits that search is efficient if the target is distinguished from the distractors by an attribute represented explicitly by neurons in primary visual cortex (Li, 1999; Nothdurft et al., 1999) and not otherwise. In fact, some complex image attributes to which striate neurons are insensitive support popout (Wolfe and Horowitz, 2004). These include the arrangement of parts in the image plane (Pomerantz et al., 1977; Heathcote and Mewhort, 1993; Davis and Driver, 1994; Rensink and Enns, 1998; Conci et al., 2006). The occurrence of popout in these cases implies that visual areas outside striate cortex can guide efficient search (Hochstein and Ahissar, 2002).

Parallel versus serial search

FIT posits that two measured behavioral phenomena (search time independent of set size vs increasing with set size) indicate two modes of search (parallel vs serial). In fact, a purely parallel model can shift gradually from seemingly parallel to seemingly serial behavior as the difference between the target and the distractors decreases (Deco and Zihl, 2006). During conjunction search, which FIT supposes to be serial, neuronal activity in monkey extrastriate cortex shifts steadily as attention converges on the target and not discontinuously as would be expected from serially allocating attention to different items (Bichot et al., 2005).

Preattentive versus attentive vision

FIT posits that there is a qualitative difference between the representation of an image formed during preattentive vision (when attention is distributed across the array) and the representation formed during attentive vision (when it is the sole object of attention). In fact, when multiple items appear in the visual field, all other things being equal, a neuron fires at a rate corresponding to the average of its responses to the individual items (Zoccolan et al., 2005). Attention to one of the items induces a quantitative increase in its degree of influence (Reynolds et al., 2000) but not a qualitative change such as would occur if neurons were sensitive only to the collection of basic features in an object forming part of an array but were sensitive to conjunctions of features in an isolated object.

The results of our study and other observations as noted above agree in supporting an alternative to FIT put forward by Duncan and Humphreys (1989, 1992). In their scheme, all search is parallel and operates on sophisticated representations of objects. Search can occupy any point along a continuum of efficiency, with efficiency increasing as the target and distractors become more dissimilar. A gap in this theory has been the lack of an operational measure of dissimilarity. Our results suggest a plausible measure based on differences in neuronal activity in visual areas including high-order cortex homologous to IT.

References

- Allred A, Liu Y, Jagadeesh B (2005) Selectivity of inferior temporal neurons for realistic pictures predicted by algorithms for image database navigation. *J Neurophysiol* 94:4068–4081.
- Bair W, Cavanaugh JR, Movshon JA (2003) Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J Neurosci* 23:7690–7701.
- Ballaz C, Boutsen L, Peyrin C, Humphreys GW, Marendaz C (2005) Visual search for object orientation can be modulated by canonical orientation. *J Exp Psychol Hum Percept Perform* 31:20–39.
- Bauer B, Jolicoeur P, Cowan WB (1996) Visual search for colour targets that are or are not linearly separable from distractors. *Vision Res* 36:1439–1465.
- Bauer B, Jolicoeur P, Cowan WB (1998) The linear separability effect in color visual search: Ruling out the additive color hypothesis. *Percept Psychophys* 60:1083–1093.
- Baylis GC, Driver J (2001) Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nat Neurosci* 4:937–942.
- Bichot NP, Rossi AF, Desimone R (2005) Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308:529–534.
- Conci M, Gramman K, Müller HJ, Elliott MA (2006) Electrophysiological correlates of similarity-based interference during detection of visual forms. *J Cogn Neurosci* 18:880–888.
- Constantinidis C, Steinmetz MA (2005) Posterior parietal cortex encodes the location of salient stimuli. *J Neurosci* 25:233–238.
- David SV, Hayden BY, Gallant JL (2006) Spectral receptive field properties explain shape selectivity in Area V4. *J Neurophysiol* 96:3492–3505.
- David SV, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron* 59:509–521.
- Davis G, Driver J (1994) Parallel detection of Kanizsa subjective figures in the human visual system. *Nature* 371:791–793.
- De Baene W, Premereur E, Vogels R (2007) Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial presentation. *J Neurophysiol* 97:2900–2916.
- Deco G, Zihl J (2006) The neurodynamics of visual search. *Vis Cogn* 14:1006–1024.
- Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. *Psych Rev* 96:433–458.
- Duncan J, Humphreys GW (1992) Beyond the search surface: visual search and attentional engagement. *J Exp Psychol Hum Percept Perform* 18:578–588.
- Edelman S, Grill-Spector K, Kushnir T, Malach R (1998) Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26:309–321.
- Enns J (1986) Seeing textons in context. *Percept Psychophys* 39:143–147.
- Enns JT, Rensink RA (1990) Influence of scene-based properties on visual search. *Science* 247:721–723.
- Foulsham T, Barton JJ, Kingstone A, Dewhurst R, Underwood G (2009) Fixation and saliency during search of natural scenes: The case of visual agnosia. *Neuropsychologia* 47:1994–2003.
- Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41:1409–1422.
- Haushofer J, Livingstone MS, Kanwisher N (2008) Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLOS Biol* 6:1459–1467.
- He ZJ, Nakayama K (1992) Surfaces versus features in visual search. *Nature* 359:231–233.
- Heathcote A, Mewhort DJ (1993) Representation and selection of relative position. *J Exp Psychol Hum Percept Perform* 19:488–516.
- Hochstein S, Ahissar M (2002) View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron* 36:791–804.
- Humphreys GW, Riddoch MJ, Quinlan PT, Price CJ, Donnelly N (1992) Parallel pattern processing and visual agnosia. *Can J Psychol* 46:377–416.
- Julesz B (1981) Textons, the elements of texture perception, and their interaction. *Nature* 290:91–97.
- Kayaert G, Biederman I, Op de Beeck HP, Vogels R (2005a) Tuning for shape dimensions in macaque inferior temporal cortex. *Eur J Neurosci* 22:212–224.
- Kayaert G, Biederman I, Vogels R (2005b) Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cereb Cortex* 15:1308–1321.
- Kentridge RW, Heywood CA, Milner AD (2004) Covert processing of visual form in the absence of area LO. *Neuropsychologia* 42:1488–1495.
- Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97:4296–4309.
- Kimchi R, Behrmann M, Olson CR (Eds.) (2003) *Perceptual organization in vision: behavioral and neurological perspectives*. Hillsdale, NJ: Erlbaum.
- Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856–867.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141.
- Lee TS, Yang CF, Romero RD, Mumford D (2002) Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nat Neurosci* 5:589–597.
- Lehky SR, Sereno AB (2007) Comparison of shape encoding in primate dorsal and ventral visual pathways. *J Neurophysiol* 97:307–319.
- Li Z (1999) Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc Natl Acad Sci U S A* 96:10530–10535.
- Messinger A, Squire LR, Zola SM, Albright TD (2005) Neural correlates of knowledge: Stable representation of stimulus associations across variations in behavioral performance. *Neuron* 48:359–371.
- Moraglia G (1989) Visual search: Spatial frequency and orientation. *Percept Motor Skills* 69:675–689.
- Nagy A, Cone SM (1996) Asymmetries in simple feature searches for color. *Vis Res* 36:2837–2847.
- Nagy AL, Sanchez RR (1990) Critical color differences determined with a visual search task. *J Opt Soc Am A* 7:1209–1217.
- Nothdurft HC, Gallant JL, Van Essen DC (1999) Response modulation by texture surround in primate area V1: Correlates of “popout” under anesthesia. *Vis Neurosci* 16:15–34.
- Op De Beeck H, Vogels R (2000) Spatial sensitivity of macaque inferotemporal neurons. *J Comp Neurol* 426:505–518.
- Op de Beeck H, Wagemans J, Vogels R (2001) Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat Neurosci* 4:1244–1252.
- Orban GA, Van Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn Sci* 8:315–324.
- Palmer SE (1999) *Vision science: photons to phenomenology*. Cambridge, MA: MIT.
- Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, Kastner S (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *J Neurophysiol* 101:2581–2600.
- Pomerantz JR, Sager LC, Stoeber RJ (1977) Perception of wholes and their component parts: some configural superiority effects. *J Exp Psychol Hum Percept Perform* 3:422–435.
- Rensink RA, Enns JT (1998) Early completion of occluded objects. *Vis Res* 38:2489–2505.
- Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26:703–714.
- Sagi D (1988) The combination of spatial frequency and orientation is effortlessly perceived. *Percept Psychophys* 43:601–603.
- Saumier D, Arguin M, Lefebvre C, Lassonde M (2002) Visual object agnosia

- as a problem in integrating parts and part relations. *Brain Cogn* 48:531–537.
- Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189.
- Tomonaga M (1999) Visual texture segregation by the chimpanzee (*Pan troglodytes*). *Behav Brain Res* 99:209–218.
- Tootell RB, Tsao D, Vanduffel W (2003) Neuroimaging weighs in: humans meet macaques in “primate” visual cortex. *J Neurosci* 23:3981–3989.
- Treisman A (2006) How the deployment of attention determines what we see. *Vis Cogn* 14:411–443.
- Treisman A, Gormican S (1988) Feature analysis in early vision: evidence from search asymmetries. *Psychol Rev* 95:15–48.
- Treisman A, Sato S (1990) Conjunction search revisited. *J Exp Psychol Hum Percept Perform* 16:459–478.
- Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cogn Psychol* 12:97–136.
- von der Malsburg C (1999) The what and why of binding: the modeler’s perspective. *Neuron* 24:95–104, 111–125.
- Williams MA, Dang S, Kanwisher NG (2007) Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci* 10:685–686.
- Wolfe JM, Horowitz TS (2004) What attributes guide the deployment of attention and how do they do it? *Nat Rev Neurosci* 5:1–7.
- Yamane Y, Tsunoda K, Matsumoto M, Phillips AN, Tanifuji M (2006) Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *J Neurophysiol* 96:3147–3156.
- Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci* 25:8150–8164.