Noticing Familiar Objects in Real World Scenes: The Role of Temporal Cortical Neurons in Natural Vision

David L. Sheinberg and Nikos K. Logothetis

Max Planck Institute for Biological Cybernetics, Tübingen, Germany 72076

During natural vision, the brain efficiently processes views of the external world as the eyes actively scan the environment. To better understand the neural mechanisms underlying this process, we recorded the activity of individual temporal cortical neurons while monkeys looked for and identified familiar targets embedded in natural scenes. We found a group of visual neurons that exhibited stimulus-selective neuronal bursts just before the monkey's response. Most of these cells showed similar selectivity whether effective targets were viewed in isolation or encountered in the course of exploring complex scenes. In addition, by embedding target stimuli in natural scenes, we could examine the activity of these stimulus-selective cells during visual search and at the time targets were fixated and identified. We found that, during exploration, neuronal activation sometimes began shortly before effective targets were fixated, but only if the target was the goal of the next fixation. Furthermore, we found that the magnitude of this early activation varied inversely with reaction time, indicating that perceptual information was integrated across fixations to facilitate recognition. The behavior of these visually selective cells suggests that they contribute to the process of noticing familiar objects in the real world.

Key words: natural scenes; object recognition; visual search; inferotemporal cortex; perceptual integration; saccadic eye movements; visual attention; monkey

Convergent evidence from behavioral, neuropsychological, and neurophysiological experiments indicates that, in the primate visual system, neurons located in the inferior areas of the temporal lobes play a critical role in the representation and analysis of visual objects (Gross, 1994; Logothetis and Sheinberg, 1996; Tanaka, 1997). Cells recorded from both anesthetized and behaving monkeys can be selectively responsive to the presentation of particular complex visual forms, implying that they may be critical for recognizing these forms (Perrett et al., 1982; Desimone et al., 1984; Tanaka et al., 1991). Although numerous experiments have suggested that the properties of these cells may help explain how humans and nonhuman primates are able to recognize an object despite changes in viewpoint or configural properties (Lueschow et al., 1994; Tovee et al., 1994; Ito et al., 1995), this study focuses on another problem that the visual system must solve: that of locating and identifying forms in complex environments.

In the real world, objects rarely appear instantaneously or in isolation. Instead, they are usually encountered in the course of exploring visually complex environments. What happens in the brain as familiar objects are searched for, noticed, and then identified? We reasoned that if neurons in the inferior temporal lobes are directly involved in the identification of forms encountered in the real world, their responses should be indifferent to the complexity of the surrounding environment, but these responses should only occur once these forms are noticed. Thus, this study had two major objectives. First, we wanted to determine whether the response selectivity of temporal cortical neurons for objects

flashed in isolation would be maintained in more natural contexts. Second, we wanted to characterize the spatiotemporal response profile of visually selective temporal cortical neurons to better understand the role they may play in natural visual processing. To approach these issues, we designed a task incorporating three critical aspects of real world vision: a complex environment, unconstrained fixation, and goal-directed behavior. We then compared neural activity observed during this task with activity recorded in a more conventional recognition paradigm. Although previous neurophysiological studies of temporal cortical neurons have examined the effects of some of these elements (Sato, 1989; Miller et al., 1993; Rolls and Tovee, 1995; DiCarlo and Maunsell, 2000), none has investigated them in a single, unified task. Our results show not only that the properties of some visual neurons are robust to these unconstrained conditions but also that their activity in such a task can help explain previous behavioral findings about the nature of perceptual integration during active vision.

A brief report of these results appeared previously (Sheinberg and Logothetis, 1999).

MATERIALS AND METHODS

Subjects and surgery. Two adult male rhesus macaques were trained to move from their home cage into primate chairs. After this initial training, the monkeys underwent sterile surgery for implantation of a custom-designed titanium implant for head restraint (Max Planck Institute, Tübingen, Germany) and a scleral search coil for eye position monitoring (Robinson, 1963). After behavioral training was complete, a titanium ball-and-socket recording chamber (Logothetis et al., 1995; Sheinberg and Logothetis, 1997) was surgically implanted in each monkey; this provided chronic guide-tube access to a conical cortical region with a cross-sectional diameter of ~12 mm at the level of the lower bank of the superior temporal sulcus. All surgeries were conducted in accordance with the policies and procedures set forth in the *U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals* and the *National Institutes of Health Guide for the Care and Use of Laboratory*
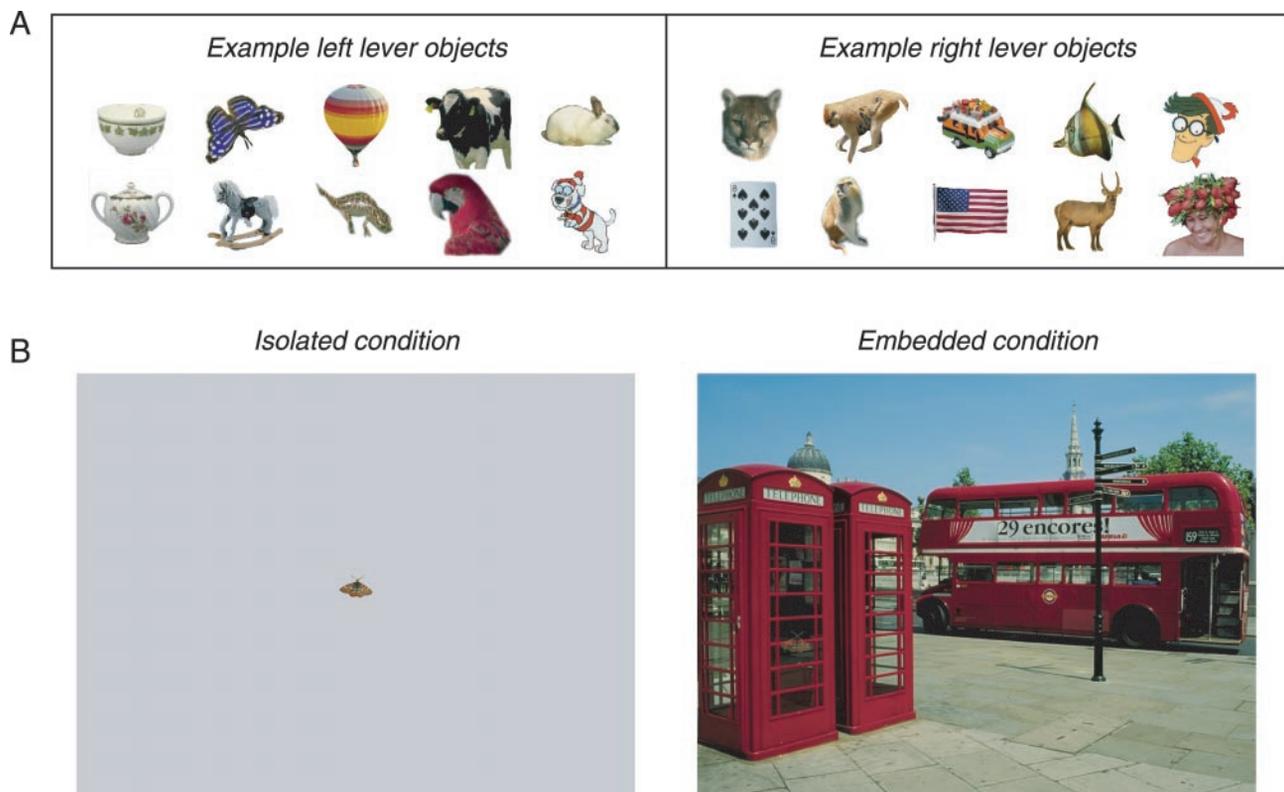
Figure 1. Stimuli and conditions used in the experiment. *A*, Subset of target images. Monkeys were trained to pull one of two levers whenever they noticed any 1 of 70 target objects on the computer monitor set before them. Lever mappings (*left* or *right*) were arbitrarily assigned, but specific exemplars from the same basic-level category (e.g., *playing cards* or *parrots*) were mapped to the same lever. *B*, Two basic conditions were used. In the isolated condition (*left*), the target stimulus was presented alone in the center of the screen, against a gray background. In the embedded condition (*right*), the target was blended into 1 of 100 real world scenes.

Animals, as adopted by the Society for Neuroscience in its Policy on the Use of Animals in Neuroscience Research.

   Behavioral paradigm. After recovery from the initial surgery, the monkeys were familiarized with images of 70 natural and man-made objects and trained to respond to each of the objects by pulling one of two levers. Objects were sorted by natural category but randomly assigned to either the left lever or right lever class so that, for example, the monkeys would pull the left lever whenever they saw a butterfly or the right lever whenever they saw a mountain lion (Fig. 1*A*). Objects were considered familiar once performance for isolated presentations was >90% throughout an entire training session. Once the animals learned to associate an object with the correct lever, we presented the object either alone (isolated condition) or randomly placed in one of 100 natural scenes (embedded condition) (Fig. 1*B*). In the embedded condition, the animal's task was to search out any familiar object and pull the correct lever on finding it. One target object was present on every trial, but, in contrast to other search paradigms, the monkey did not know its identity before the trial began. To encourage natural exploratory behavior, no fixation constraints were imposed after a trial began, but eye position was recorded throughout the experiment.

   Stimuli were presented on a dedicated graphics workstation (TDZ 2000; Intergraph Systems, Huntsville, AL) at a resolution of 1280 × 1024 at 85 Hz refresh, running an OpenGL-based stimulation program under Windows NT. Behavioral control for the experiments was maintained by a network of interconnected PCs running the QNX realtime OS (QSSL, Ontario, Canada). This system included a high-resolution clock–timer (GT401; Guide Technologies), a sound generator (Yamaha SW60XG), an interrupt-driven digital input (PIO-INT; Keithley Instruments, Cleveland, OH), and a 12-bit analog input for eye position signals (PCL-818; Advantech, Sunnyvale, CA). Communication with the graphics computer was by dedicated Fast-Ethernet. All events relevant to the experiment, including lever presses, analog eye position, and stimulus information, were both streamed to disk and available for on-line monitoring.

   Individual target objects were selected from a set of stock photo CDs (Corel, Ottawa, Ontario, Canada), and an α channel was added, allowing

nonrectangular blending operations between targets and either the blank or scene backgrounds. Targets and background were composed on-line before each trial, with targets being blended into the backgrounds at a ratio of 60:40 (target/background) to minimize sharp edges introduced by the blending procedure and increase the difficulty of locating target objects. Targets subtended ~1.5° visual angle and the scenes were 25° across. No fixation spot was present during the behavioral trials, but a trial only began after the monkey entered a virtual fixation window at the center of the screen (diameter, 3° for isolated trials and 12° for embedded trials). No other control of eye position was imposed. Trials ended (and the visual stimulus was turned off) after the monkey pulled a lever or 15 sec elapsed, whichever came first. Feedback was provided on all trials because juice was delivered only when the monkey correctly identified the target present on a given trial. Isolated and embedded trials were presented in interleaved blocks consisting of at least 60 trials each. Within a block of trials, the number of left and right lever targets was always evenly divided. The monkeys performed between 1000 and 2500 trials per session.

   Eye position was digitized at 1 kHz, and running averages were written to disk for every fifth sample (200 Hz). At the beginning of each session, offsets were adjusted by having the monkey fixate a small square (0.3° per side) positioned at the center of the screen. A calibration procedure was then performed by having the monkey repeatedly saccade to small squares at one of 24 positions on the screen; during this time the gain of the eye position system was iteratively adjusted to minimize estimated position error.

   Recording methodology. Recordings were made in the region between 15–20 anterior (A) and 16–19 lateral (L) in the right hemisphere of each monkey (116 penetrations in monkey Q and 56 in monkey S), and structural magnetic resonance imaging scans were used to estimate the location of electrode track positions. Electrodes consisted of a PtIr (90/10; A-M Systems) core that was coated in glass (Corning, Corning, NY). Neural signals were conditioned using a standard amplifier system with remote probe (Model A-1; BAK Electronics, Germantown, MD) and an active filter (Krohn-Hite, Avon, MA; high-pass cutoff, 100 Hz/12

dB; low-pass cutoff, 8 kHz/24 dB). Single, and often multiple single, cells were isolated using a software-based time-amplitude window discriminator. In this study, only the largest single cell isolated at a particular site was included in the analysis. All analog neural data were streamed to disk at 22 kHz, and spike times reported here were based on off-line analysis of this signal. During physiological recording sessions, we actively searched for cells responsive to any items in our target set while the monkey performed the isolated recognition task. Many cells were bypassed in this effort because the number of trials in any one session was limited by the monkey's behavioral performance, and previous experimentation had shown that finding cells with robust stimulus-selective responses to particular stimuli requires extensive exploration.

Once a cell was selected, an initial screen of the target set was presented in the isolated task, and on-line peristimulus rasters were generated. After the initial presentation of the whole set, a subset of between 8 and 16 target stimuli was chosen; the two most effective targets were included. We limited the number of target stimuli used to ensure that an adequate number of trials would be acquired for each target. For the embedded trials, the same targets were used and integrated into a randomly chosen subset of the 100 background scenes.

Data analysis. In addition to conventional event-triggered spike density estimates, we used the Poisson spike train analysis (Legéndy and Salcman, 1985) as a method to find unusual epochs of neural activity occurring at any time during the behavioral trial. Note that the application of Poisson analysis does not imply that the interspike intervals (ISIs) of individual cells are truly distributed as a Poisson process. Indeed, previous analyses and our own have shown that extremes of this distribution are so common that the approximation does not fairly describe the behavior of many cortical neurons. The analysis was thus used simply to search objectively for these unusual events. We analyzed our spike trains using a modified algorithm of that originally described by Legéndy and Salcman (1985) and adapted by Hanes et al. (1995). The formula used for calculating S, the surprise index, was $S = -\log(P)$, where P represents the probability of observing n spikes in a time interval T, given a mean rate r, if those events were distributed according to a Poisson distribution. Determining r, the mean discharge rate, is nontrivial because many of our cells had extremely low spontaneous rates but could be reliably activated with appropriate stimulation. For consistency, we calculated one value of r for each cell. Although this approach ignored overall changes in baseline firing, it allowed us to compare surprise values across all trials without regard to the particular stimuli present on a single trial. Using the estimated discharge rate, we searched through spike trains for a minimum of two consecutive ISIs, each of which was less than half the mean ISI. We then continued to add spikes to the burst as long as the surprise index continued to increase. At this stage, the earliest spikes were eliminated one by one as long as this increased the surprise value. The identified bursts were then characterized by their start and stop times, their length, and their surprise index, S. Intuitively, high values of S indicate periods of unusually high activity. A comparison of the responses in Figure 2, *A* and *B*, with Figure 4, *A* and *B*, (◎), illustrates the correspondence between the surprise measure and more traditional estimates of spike rate.

RESULTS

After training, performance in both the isolated and embedded conditions was nearly perfect (>95% correct), and the median reaction time in the isolated condition was 390 msec (387 and 393 msec for monkeys Q and S, respectively). Reaction times in the embedded condition were naturally more variable and ranged between 270 msec and 15 sec (the maximum allowed before the trial was automatically terminated). Despite the differences in reaction times between the isolated and embedded conditions, there were no significant differences in classification performance (96% in both conditions).

Neurophysiological recordings were made by slowly lowering microelectrodes vertically into the lower bank of the superior temporal sulcus and the lateral convexity of the middle temporal gyrus. We isolated 268 single units and assessed their selectivity for any of the learned target images using the isolated recognition task. Figure 2, *A* and *B*, illustrates response profiles of two different cells (one from each monkey) to a subset of the learned

stimuli. In each subplot, the spiking activity of the cell is aligned to the onset of the particular stimulus shown above the plot. For these cells, spontaneous activity was extremely low, but a reliable burst of activity was clearly evident after the presentation of certain stimuli from the test set. These bursts began between 100 and 130 msec after the stimulus appeared and preceded the manual response by ~250 msec. A preferred stimulus for the cells could be identified, but less intense and less consistent activity was still apparent for other, often visually similar, stimuli. One of the central aims of the present experiment was to determine whether the response profiles found with the isolated stimuli would fairly characterize the neural activity recorded when the animal searched for and found the same objects embedded in natural scenes.

Figure 3 illustrates embedded trials for the two cells of Figure 2. In Figure 3*A*, the target object (*inset*) is positioned on the roof of the church (*red circle*), and the *continuous white line* traces the monkey's direction of gaze during the trial. The behavioral and neural activity for this single trial are shown below. The plot of gaze distance to the target as a function of time shows that the eyes were within 6° of the target throughout the entire trial, but it was not until after the sixth saccade, when the monkey looked directly at the target, that he seemed to notice it and pull the lever. The aligned neural activity of the cell supports this conclusion, in that the cell was completely silent during the search until the target was fixated, at which time the characteristic burst occurred, followed by the manual response. Similarly, the cell of Figure 3*B* responded just before the monkey's overt response to the target, after >4 sec of inactivity during the prolonged search.

Figure 3 also illustrates the problem of identifying an appropriate measure for characterizing the behavior of a cell in these extended, subject-controlled trials. In our initial analysis of the data, we wanted to avoid the traditional approach of analyzing a specific, but arbitrary, epoch of each trial, because any such selection would assume that activity during this time was more important than that at other times throughout the trial. Instead, we used the statistical properties of the spike train of a cell to locate periods of significantly elevated discharge. This method, called Poisson surprise or burst analysis (Legéndy and Salcman, 1985; DeBusk et al., 1997), results in an enumeration of time intervals during which cell activity was unusually intense, based on the average firing rate of the cell. Each interval is assigned a surprise index, S, which is a measure of how unlikely such a period of elevated activity would be for the cell in question. The potential importance of bursting activity in general neural computation has been previously emphasized (Lisman, 1997), as has its applicability to behavioral studies (Hanes et al., 1995; Livingstone et al., 1996). Here we demonstrate its utility in analyzing the activity of temporal cortical neurons during active recognition.

In both the isolated and embedded conditions, every trial contained a single target, and we could thus correlate the bursting activity of a cell with the particular target identified on that trial. Because more than one burst could occur during a trial, we chose to describe a trial by the magnitude of the most surprising burst that occurred anywhere in the course of the trial. Trials with no burst were assigned a value of zero. Figure 4 shows the results of such an analysis for four cells (including the two shown in Fig. 2). The maximum burst surprise values on each trial are shown for each of four target stimuli used in both the isolated and embedded conditions. *Open squares* indicate responses on single trials to the stimuli presented in isolation, and *gray circles* show the responses for the same object embedded in a natural scene. Despite
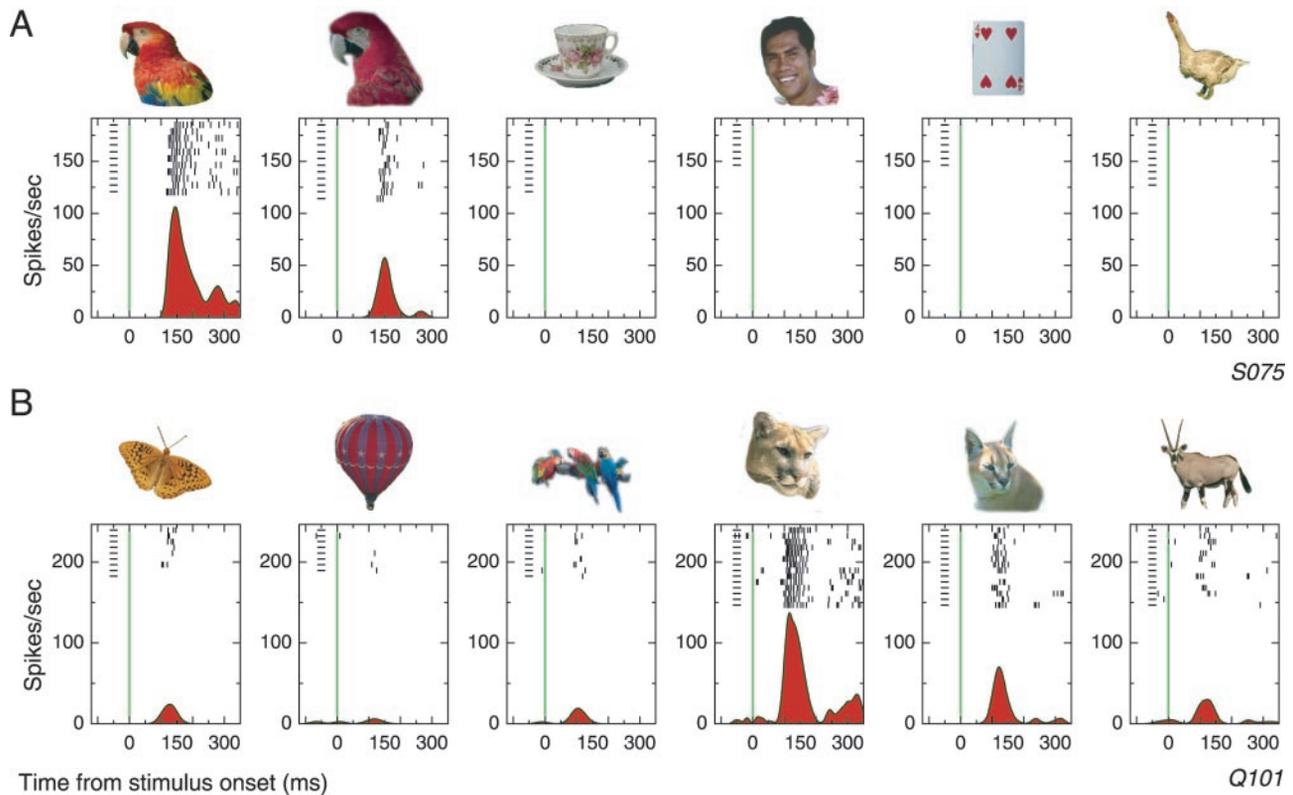
Figure 2. Target-selective visual responses. *A*, Response of a single cell from monkey S to six of the target objects from the test set presented in isolation. The first three targets were mapped to the left hand, and the second three were mapped to the right hand. Targets were presented in pseudorandom order against a gray background, and the monkey was rewarded for pulling the preassigned lever for each object. Trials were sorted by target object and aligned to the onset of the stimulus (*green vertical line*). *Vertical ticks* denote the time of occurrence of action potentials in each trial, and *small horizontal lines* simply indicate the presence of a trial (which may have included no spikes). Beneath each set of rasters is an estimate of instantaneous firing rate calculated using adaptive kernel estimation (Richmond et al., 1990). *B*, Selective response of a cell from monkey Q to a different subset of stimuli.

the presence of complex natural surrounds, the bursting activity of the cells was limited to trials containing effective target stimuli. The background scenes used to embed both effective and ineffective targets had little impact on the response selectivity of the cells because scenes that contained an ineffective stimulus were no more likely to include a burst than an isolated trial with the same stimulus. Furthermore, on each trial, the burst of the cell and the monkey's lever pull tended to occur with about the same latency after the onset of the visual stimulus (Fig. 4*A–D*, *inset plots*), supporting the view that these two events are related. Observe, however, that this burst activity was not directly related to the monkeys' motor behavior, because multiple stimuli resulted in the same manual response, but only select stimuli elicited a reaction from the cells. The strong relationship between the behavioral confirmation of target identification and the bursting behavior of cells, together with the fact that background features fixated during exploration did not elicit significant bursts, indicates that the activity of these cells is correlated with the animal's noticing and responding to particular stimuli.

We examined the behavior of the population of 268 cells (170 and 98 from monkeys Q and S, respectively) using the burst analysis described above. Of these 268 cells, 62 showed significant differences ($p < 0.01$, paired t test) between spontaneous firing and visually elicited burst behavior and are the subject of the following analysis, which examines the effect of target type (effective/ineffective) and background scene (isolated/embedded) on burst activity. Of the 62 cells, 49 (79%) showed a significant difference between their responses to the most effective and least

effective stimulus presented in isolation. For 65% of these cells (32 of 49; 23 from monkey Q, 9 from monkey S), this target selectivity was also observed in the embedded trials. A few of these cells (4 of 49 or 8%) exhibited stronger bursts when the preferred stimulus was embedded in a natural scene than when it was presented in isolation, whereas ~20% (10 of 49) showed a significant reduction in burst response to the most effective target in the embedded condition. This suppression is consistent with previous physiological investigations in both temporal cortical (Sato, 1989; Miller et al., 1993; Rolls and Tovee, 1995) and earlier visual areas (Gallant et al., 1998) that have shown that the presence of more than one object in the visual field can have a significant suppressive influence on the response of a cell to the object alone.

We then examined whether burst magnitudes were affected by the variable time it took the monkeys to solve the task on individual trials. Figure 5*A* depicts three trials taken from a single cell, each containing the same (most effective) target. The response time of the monkey on each trial is indicated by the elapsed time between the two vertical bars and clearly varies between the trials. To assess whether bursts occurring at different times after the start of a trial were stronger or weaker than the average burst, we pooled all trials for the 32 cells that showed stimulus-selective burst modulations in the embedded condition and selected those trials containing the most effective stimulus ($n = 1089$). We normalized all bursts for a single cell to the average burst magnitude for that cell. We then partitioned the trials according to the time of occurrence of the maximum burst.
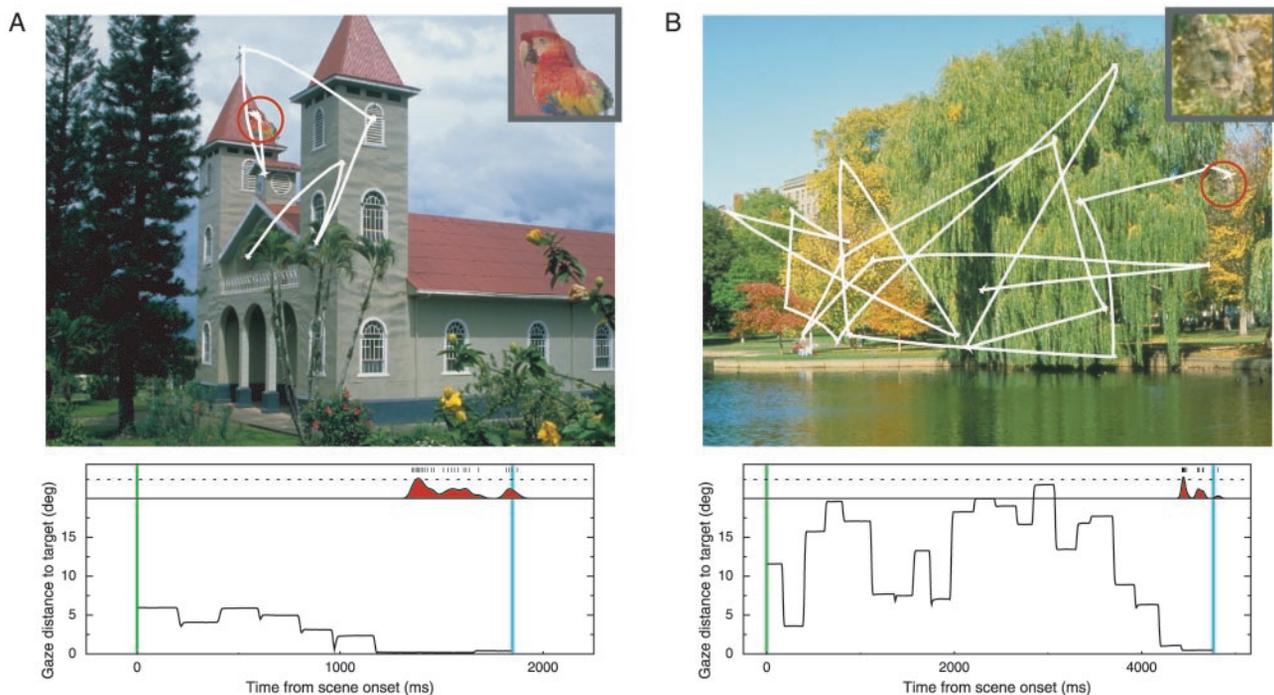
Figure 3.　The embedded search task: associated behavior and neuronal response. *A*, Representation of the actual visual stimulus presented on a single trial with eye movement records superimposed. On each trial in the embedded condition, one of the target objects was placed at a random location within one of 100 background scenes. The *eye trace* starts near the center of the screen and ends atop the embedded target, which on this trial was the most effective stimulus of the cell (as indicated by Fig. 1*A*). The *red circle* and *inset* show the location and identity of the target (neither was present in the experiment). Before the trial, the monkey did not know which target would be present or where it would appear. The behavioral and neural responses for this trial are depicted in the *graph* below. The *bottom section* of the *plot* shows the distance from the center of gaze to the target as a function of time. The stimulus appeared at time 0 (*green vertical line*); the *vertical blue line* denotes the time of the lever pull and stimulus disappearance. In this plot, saccades are indicated by sharp transitions in the distance function, and fixations are indicated by the flat epochs. Spikes occurring on this trial are shown above the plot, as is a spike density function estimate. The *dotted line* in this and subsequent plots corresponds to an estimate of 100 spikes per second. Note the silence of the cell between scene onset and the time at which the animal appears to notice the target and responds. *B*, A single embedded trial for the cell of Figure 2*B*.

As shown in Figure 5*B*, burst magnitudes were unaffected by when, in time, the burst occurred. In addition, Figure 5*B* shows the average time of occurrence of each of the first six saccades after the presentation of the scene. The saccade data (*n* = 20,510 saccades) show a linear relationship between the number of saccades made and time, indicating that eye movements were programmed at a relatively fixed rate. The slope of the linear fit was 241 msec/saccade, or just over four eye movements per second. This means, for instance, that if the maximum burst began 1500 msec after the beginning of a trial, there would be, on average, six saccades preceding that burst (Fig. 5*A*, *bottom plot*). Taken together, we see that burst magnitude does not change as a function of time or the number of preceding saccades.

Next we asked how closely the observed changes in neural activity related to overt aspects of the animals' behavior. For this analysis, we constructed event-triggered averages aligned on the monkeys' visual acquisition of the target and on their manual lever pulls. Figure 6 illustrates the population activity averaged across the 32 cells that showed significant selective bursting behavior in both the isolated and scene conditions. For each cell, trials containing the most effective and least effective stimuli were extracted. Spike density estimates were calculated using the adaptive kernel procedure (Richmond et al., 1990) and were normalized to the maximum rate observed in the effective/isolated condition on a cell-by-cell basis. In Figure 6*A*, time 0 marks the arrival of gaze direction to within 1.5° of the center of the target (*target acquisition*), whereas in Figure 6*B*, the same trials are aligned on the manual response. In the isolated trials, the monkeys acquired the target at the moment the stimuli appeared (or the trial was aborted), and the *solid gray line* in Figure 6*A* closely resembles the spike density functions for the single cells shown in Figure 2. For both the isolated (*gray*) and embedded (*black*) conditions, there is a clear difference between trials with effective targets (*solid lines*) and those with ineffective targets (*dotted lines*), demonstrating the selectivity of this cell population and the fact that in the absence of an appropriate stimulus, little if any modulation in activity is observed.

We estimated the time at which the population response to the effective targets diverged significantly from the response to ineffective stimuli for both the isolated and embedded conditions by conducting repeated pairwise *t* tests for consecutive 10 msec epochs spaced 5 msec apart (Fig. 6*A*, *gray* and *black asterisks* along the *abscissa*). In the isolated condition, repeated significant differences (*p* < 0.01) began 100 msec after stimulus onset, providing an estimate of the response latency of our cells to conventionally presented effective stimuli. However, in the embedded condition, differential response to effective targets began 95 msec before the eyes acquired the target. The shallower slope for the effective/embedded conditions indicates that this preactivation did not occur at the same time on every trial. A second peak in this activity profile, starting ~100 msec after target acquisition, aligns almost perfectly with the activity profile in the isolated condition, suggesting that a second round of processing began only after the eyes landed on the target.
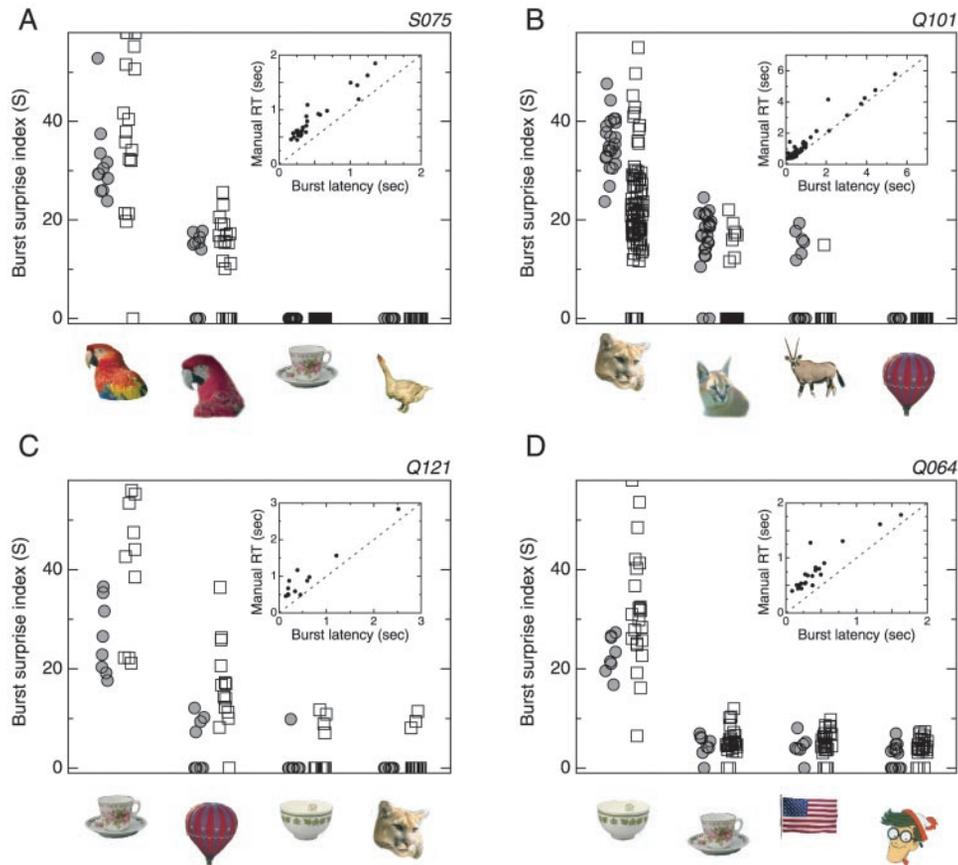
Figure 4. Comparison of single trial responses for isolated and embedded conditions. *A*, Each behavioral trial was assigned a burst surprise value, *S* (ordinate), corresponding to the largest amplitude burst detected on that trial. Large burst values correspond to short periods of high activity. Trials were sorted by target stimulus (depicted by *target images* below the *graph*) and then by condition. The *gray circles* represent isolated trials, and the *open squares* represent embedded trials. Small horizontal displacements of the data points are for visualization purposes only. For a given cell, the same set of scenes was used to test the response to all targets. Response differences to presentations of the most effective and the less effective targets are maintained during the embedded trials, showing that the complex surrounds did not eliminate the selectivity of the cell. *Insets*, Timing of individual bursts occurring in embedded trials. Maximal bursts began ~300 msec before the manual response, independent of when the response actually occurred. The linear relationship strongly suggests that the activity bursts were coupled to the act of noticing and responding to the preferred targets and not just to their physical presence. *B–D*, Three other example cells demonstrating stimulus selectivity in the isolated and scene conditions, as well as the temporal relationship between bursts and responses.

When the data are aligned on the manual response (Fig. 6*B*), the average responses to effective targets in the isolated and embedded conditions do not differ significantly at any time for the 800 msec period shown in Figure 6*B*. Both show large activity increases leading up to the response, and it is interesting that the elevation in activity continues well beyond the motor response (and the concurrent disappearance of the stimulus). Activity persisting beyond the response can obviously play no part in a behavior that has already occurred but may contribute indirectly to performance on subsequent trials through the strengthening of connections between coactive cells (Yakovlev et al., 1998).

Although studies of translation invariance in temporal cortical neurons have generally found that overall selectivity is relatively independent of stimulus position (Schwartz et al., 1983; Lueschow et al., 1994; Tovee et al., 1994; Ito et al., 1995; Logothetis et al., 1995), the absolute response of these cells is known to decrease with increasing target eccentricity (Ito et al., 1995). In our task, target eccentricity varied from fixation to fixation, as illustrated in the plots of Figures 3 and 5*A*. We therefore more closely examined the effect of eccentricity on the response to effective targets. In Figure 7*A*, we sorted the effective/embedded data from Figure 6*A* by the distance the eyes were from the target

on the fixation just before acquisition. This resorting yielded six overlapping groups of trials for presaccadic target distances ranging between 1.5 and 14° visual angle (1.5–4°/286 trials, 2–6°/500 trials, 4–8°/584 trials, 6–10°/517 trials, 8–12°/339 trials, 10–14°/168 trials). These groupings uncovered a systematic difference in the early activation noted in Figure 6*A*. Specifically, trials in which target acquisition occurred from nearby positions showed more activity before and just after the eyes landed on the effective target than did trials in which the target was acquired from more distal positions. These differences, which began before the eyes moved, peaked ~45 msec after the target was fixated. The magnitude of the activity between 20 and 70 msec after acquisition decreased systematically with target distance (Fig. 7*B*) and indicates that the amount of extrafoveal information available before a target is fixated varies with distance. We also analyzed the activity between 125 and 175 msec after acquisition (second peak in Fig. 7*A*; comparison not shown) and found no effect of presaccadic target distance. Activity during this period presumably reflected analysis of the newly acquired image and was not affected by how far away the target was before acquisition.

Taking the analysis one step further, we examined whether there were any behavioral correlates of this systematic change in
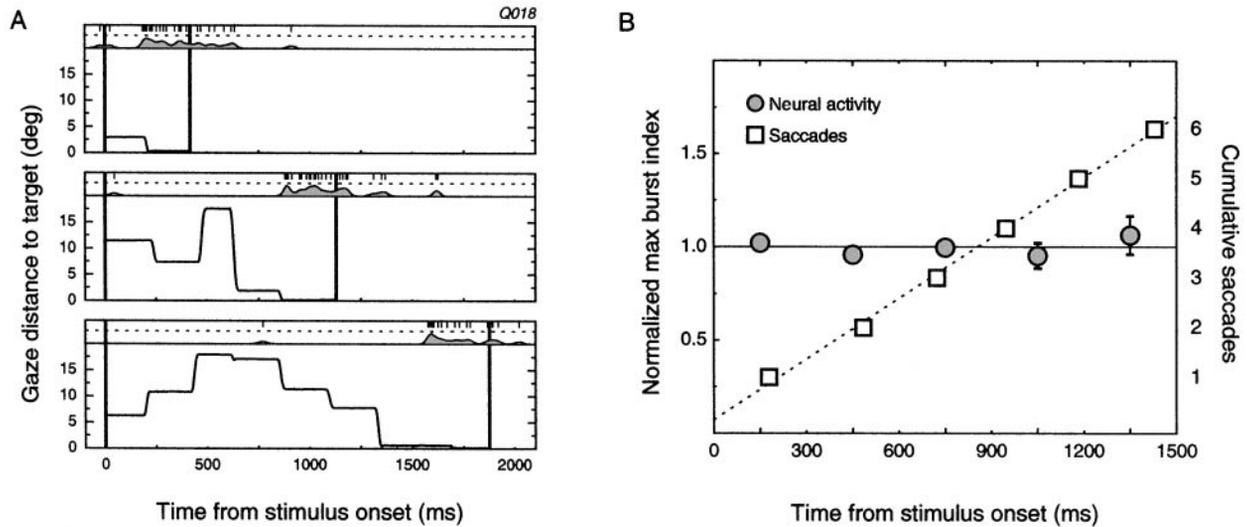
Figure 5. Latency of bursts has no effect on magnitude. *A*, Three trials from the embedded condition for a single cell with the same target stimulus. Each trial was performed correctly, but the search times varied between trials. Plot conventions are as in Figure 3 and show the distance of the eyes from the target as a function of time and the corresponding pattern of neural activity. *B*, To compare the burst magnitudes at different times after stimulus onset, the maximum burst responses to the most effective stimuli were sorted by latency and binned into five groups (0–300, 300–600, 600–900, 900–1200, and 1200–1500 msec). Bursts were normalized by dividing by the mean burst amplitude for each cell, and the average for each time bin is shown by the *filled circles* (error bars denote SEM). Note that there is no effect of latency on magnitude, showing that neither the initial visual transient nor changes in attentional state throughout the trial have any systematic influence on the activity of the cells. The number of saccades executed increases linearly with time (□), illustrating that the number of preceding eye movements does not impact the response of cells during natural search.
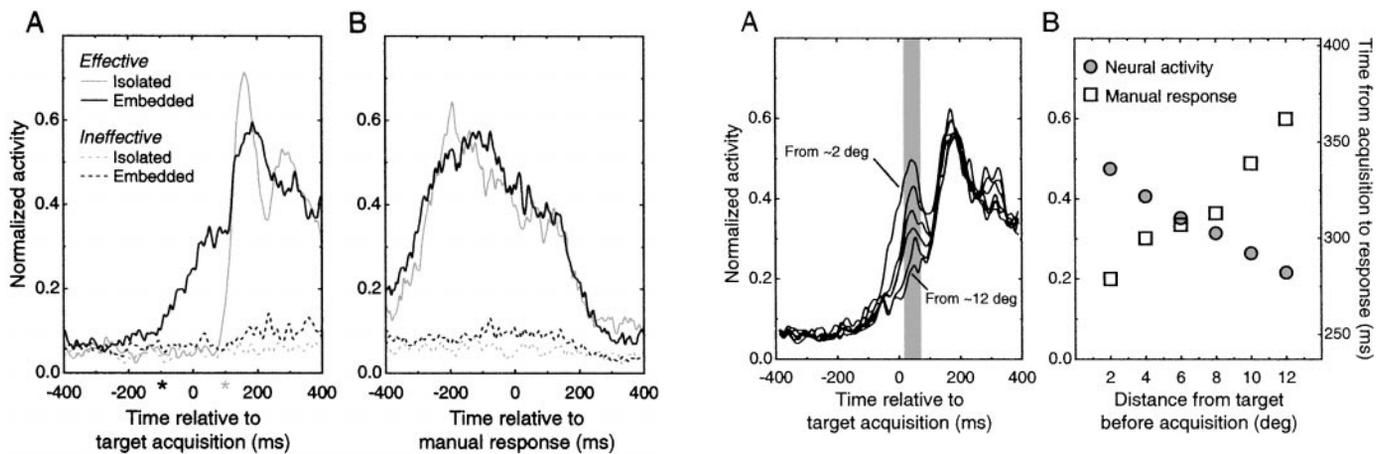

Figure 6. Population activity profiles. *A*, Average activity aligned on the time the monkey fixated within 1.5° of the center of the target. Averages comprise normalized activity estimates for each cell that exhibited significant selectivity for at least one target in both the isolated and embedded conditions (n = 32 cells). *Dotted lines* correspond to the response to ineffective targets in the isolated (*gray*) and embedded (*black*) condition, and both show little modulation compared with activity on trials containing the most effective target (*solid lines*). For the isolated trials, target acquisition time coincided with target appearance, whereas in the embedded trials, the monkey could acquire the target any time after the scene onset. The activity for all four conditions for each cell was normalized to the maximum of the activity estimate in the isolated/preferred (*solid gray line*) condition. The pilot kernel for all spike density estimates was 5 msec. *B*, Average activity aligned on the manual response time for the same conditions and cells shown in *A*. The difference between the effective (*solid lines*) and ineffective targets (*dotted line*) is clear, but no significant difference between the isolated (*gray lines*) and embedded (*black lines*) trials was observed.

Figure 7. Perisaccadic activation and reaction times grouped by target distance. *A*, Data from Figure 6 have been sorted by distance from the center of gaze to the target before foveation (presaccadic target distance). The *top curve* represents trials in which the target was acquired from between 1.5 and 4° (*From ~2 deg*), and the *bottom curve* was acquired from trials in which the presaccadic target distance was from 10 to 14° (*From ~12 deg*). Curves are average spike density estimates normalized to the peak activity in the isolated/effective condition (as in Fig. 6). The number of trials included in each curve was (from *top* to *bottom*) 286, 500, 584, 517, 339, and 168. The *gray area* highlights the peak of the eccentricity-dependent early activation (20–70 msec after acquisition). *B*, Early peak activity and manual reaction times plotted against presaccadic target distance. The average activity between 20 and 70 msec after acquisition (*A*, *gray region*) systematically declined as a function of the distance of the target before it was acquired. Median reaction times, starting from the time the target was fixated (□), show that the amount of time necessary to identify the target after it was fixated correlated inversely with the activity observed in the early peak.

early activity across this population. Using the same trial groupings as described above for the neural data, we asked whether reaction times reflected the variation in neural activity. Indeed, we found that the time between target acquisition and manual

response time systematically increased the farther the target had been on the previous fixation. This effect is illustrated in Figure 7B (□). For the most distant grouping (targets acquired from between 10 and 14°), the median time from the first fixation of the
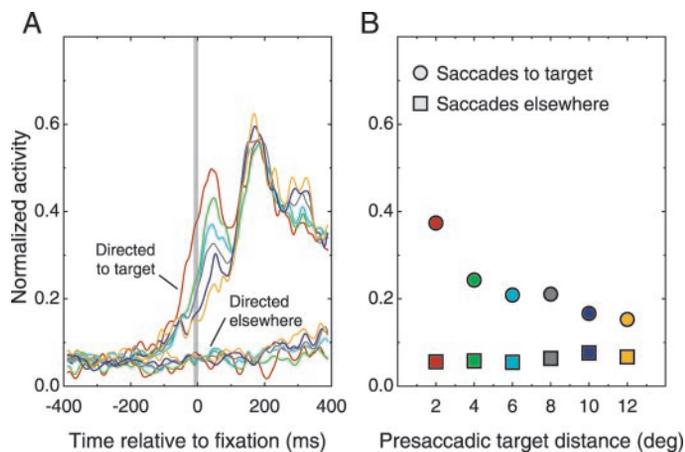
Figure 8. Early activation depends on the goal of the next fixation. *A*, Data from Figure 7*A* are replotted and color-coded by distance (*top set of curves*). The *bottom traces* represent epochs during which the most effective target was present before a saccade but was not the target of the saccade. These trials were sorted by how far the target was from the direction of gaze before the next fixation and color-coded to correspond to the *top traces*. In contrast to the increase in activity found before saccades aimed at the target, no such activity was found when the monkey did not appear to notice the target and looked elsewhere. The number of epochs in the *Directed elsewhere traces* was (from closest to farthest) 135, 281, 437, 529, 532, and 453. More than one epoch could come from the same trial. *B*, Comparison of presaccadic activity for epochs ending in target acquisition and epochs preceding saccades elsewhere. Average activity before the eyes landed (from −10 to 0 msec, before acquisition) shows that when the eyes were headed toward the effective target, neural activity started before the eyes landed on the target, but this activation depended on presaccadic target distance. No such activity was observed when the eyes were not directed to the target, although presaccadic distances were the same as in the acquisition trials.

target to the lever pull was 363 msec, but when the target was acquired from nearby positions (1.5–4°), this time dropped to 279 msec, a decrease of 84 msec. The obvious implication is that information useful for identifying the target could be acquired before the eyes foveated the target and that the amount of useful information decreased with increasing eccentricity, most likely because of reduced acuity. This conclusion complements the conclusion we drew from the neural data above and strongly implicates these cells as active participants in the process of visual recognition.

To bolster this claim, we asked whether the activity of these cells was contingent on whether the animal actually noticed the target. Because many of our trials included extended periods during which the target was present but the animal seemed unaware of its whereabouts, we could use this data to determine whether the physical presence of the target alone was adequate to activate the cells. For all trials containing the effective target, we extracted those epochs aligned on fixations preceding target acquisition and then sorted these by how far the eyes were from the target before the fixation. Figure 8*A* illustrates the results of this analysis. The top set of lines shows the same data as Figure 7*A* recoded by color, and the bottom set of lines, with corresponding color codes, is from fixations not directed at the target. For the latter trials, we hypothesize that although the target was potentially visible, the monkey did not notice it, and he therefore looked elsewhere. By ~100 msec after the fixation, the trials begin to diverge significantly, but this difference can be attributed simply to the fact that by this time the eyes were looking directly at an effective target for the acquisition trials (Fig. 8*A*, *top set of*

lines) but not for the others (*bottom set of lines*). More interesting is the period leading up to the eye movement, because the differences seen here show that target-selective responses begin before the eyes move only when the target appears to have been noticed. Figure 8*B* illustrates that the difference between the targeting and nontargeting fixations was clearly evident for the last 10 msec period of the preceding saccade, during which saccadic suppression (Matin, 1974) would prevent new form-specific information from entering the visual system. To test the significance of this effect, we analyzed the trials from three non-overlapping eccentricity groups (so that no trial was counted twice) with a two-way ANOVA, using condition (toward or away from target) and eccentricity (1.5–4, 4–8, and 8–12°) as factors. This analysis confirmed that both main effects were significant [condition, $F_{(1,1989)} = 110.4$; $p < 0.001$; eccentricity, $F_{(2,1989)} = 8.7$; $p < 0.001$], as was the interaction [$F_{(2,1989)} = 10.0$; $p < 0.001$].

Further evidence for a link between the activity of these cells and the state of noticing visual targets comes from a small subset of the data that we called "double take" trials. In the vast majority of successfully completed search trials, the monkey's eye position followed the pattern illustrated in Figures 3 and 5, wherein the eyes located the target and then fixated it; the monkey made its manual response during this time. We found, however, that on ~4% of the embedded search trials, the eyes fixated the target, passed over it, and then quickly returned, as if the monkey realized he had seen the target only after executing an intervening saccade. The frequency of these return saccades is very close to that reported in a previous behavioral study of search in monkeys (Motter and Belky, 1998). Here, we were interested in how a neuron that was selective for the target overlooked in such a trial would respond and, in particular, when in time a selective burst might occur. Figure 9*A* illustrates a double take trial and shows that the eyes approach the target, fall short, and then very quickly return after an intervening saccade. The pattern of eye movements indicates that sometime between the initial saccade and the intervening saccade, the monkey realized that the target was present. The neural response, shown below the trial, supports this conclusion, because only after the eyes land well away from the target does the characteristic bursting of the cell begin; this is followed by the return saccade and then the lever press. Figure 9, *B* and *C*, illustrates similar trials for two other stimulus-selective cells, and both show that the response of the cells begins just after the eyes move away from the target but before the return saccade. In this experiment, because we had no way of controlling when these trials would occur or which target was present when they did occur, there are too few trials available for a complete analysis. Nonetheless, these few trials support the view that the activity of selective temporal cortical cells best correlates with the state of noticing the presence of familiar forms and that this state cannot be predicted simply by the current position of the eyes with respect to a target stimulus.

DISCUSSION

The purpose of this study was to examine the physiological properties of temporal cortical neurons during exploration of complex scenes. We found visual cells in the anterior regions of the temporal lobes with reliable and selective visual responses for visual objects that the monkeys had learned to recognize. These responses were similar whether the objects were flashed in isolation or found during search, suggesting that the observed activity is related to the process of noticing particular targets, independent of how they are found. In the isolated condition, it is unclear
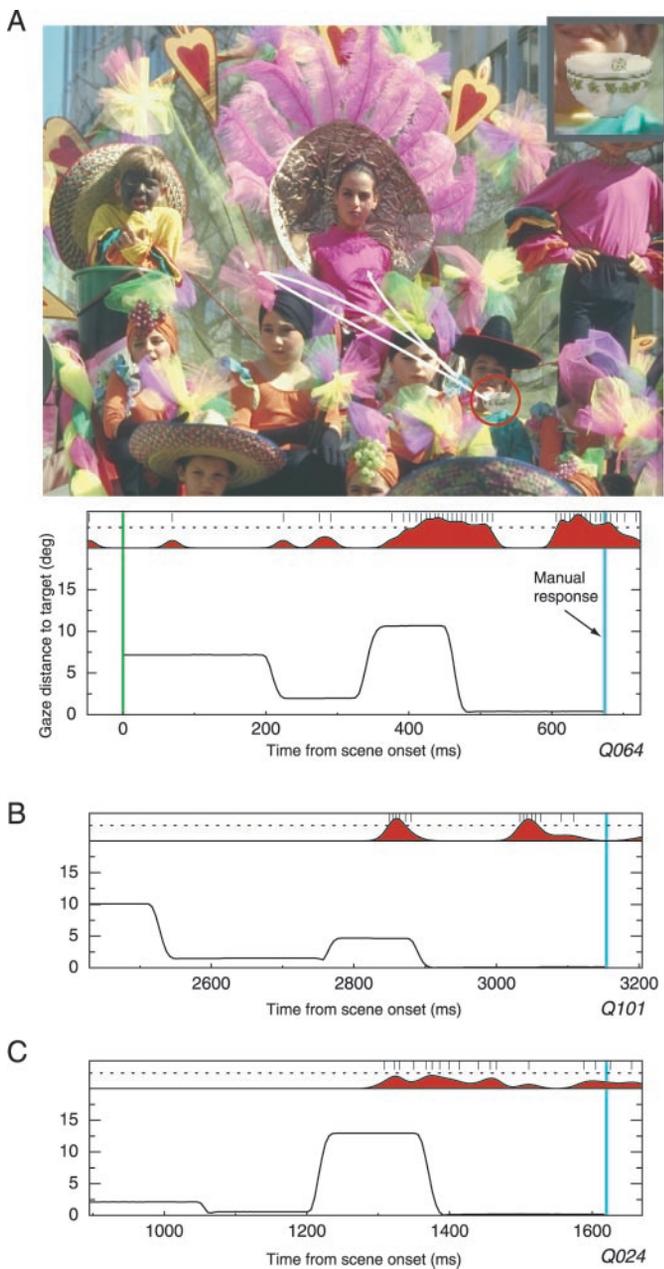
Figure 9. Double-take responses. *A*, Example trial in which the monkey's eye movement pattern indicates that he noticed the stimulus just after looking away from it. The eyes then quickly return to the target location, and the monkey makes his response. The neural burst, shown in the *plot* below the stimulus, occurs during the intervening fixation, after which time the gaze returns to the target. The neural processing of the form of the target continued independent of the eye movement and seems most related to the state of registering the presence of the target. *B, C*, Other examples of double-take trials from two different cells.

how much of the observed response results from the sudden onset of a single target, because this external event presumably captures the attention of the entire visual system. Analysis of neural activity during search helped clarify this issue because, under these conditions, targets were often noticed only hundreds of milliseconds after the stimulus initially appeared. Nevertheless, even without abrupt external transients and in the presence of unconstrained eye movements and complex visual surrounds, stimulus-selective neurons still responded shortly before the mon-

key's overt manual response. Closer inspection of the precise timing of this response revealed that information about the identity of targets was sometimes extracted before the eyes acquired the target, but only if the monkey was about to fixate the target. Behaviorally, this preview led to speeded reaction times, indicating that the information not only was available to the visual system but also was used to guide behavior.

Although most studies of temporal cortical neurons have concentrated on the responses to complex but isolated figures, we specifically set out to determine how these cells would respond during exploration of equally complex backgrounds. Gallant et al. (1998) previously examined the effect of free viewing of natural scenes on neural activity in visual areas V1, V2, and V4 and reported an overall reduction in activity during exploration, which they attributed to both suboptimal stimulation and surround inhibition. Our results in the temporal cortex are compatible with these findings, because very little discharge activity was observed while the monkeys explored the scenes before finding effective target stimuli. We analyzed the entire period encompassing the active search and were struck by the fact that incidental objects encountered in these epochs rarely led to bursts of activity similar in magnitude to the discharges elicited by particular effective targets. If these bursts had occurred, then their presence would be evident, for example, in Figure 4 for the embedded trials with ineffective targets. Instead, the visually selective cells did not contribute in any obvious way to the representation of random features or other objects located in the scenes. One interpretation of these results is that visual neurons in the temporal lobes are more involved in connecting particular feature configurations with learned actions or other mental associations than they are with the analysis of all visual patterns.

Previous studies have shown that the presentation of multiple isolated stimuli can have suppressive effects on cell responses in both early visual areas (Reynolds et al., 1999) and temporal cortex (Sato, 1989; Miller et al., 1993; Rolls and Tovee, 1995; Missal et al., 1999). These experiments demonstrate that the response of a cell to multiple stimuli cannot be predicted by the response to each of the constituent stimuli alone. Instead, interactions between multiple stimuli appear to compete for neural representation (Chelazzi et al., 1998; Reynolds et al., 1999). In this study, we also found that for ~20% of cells that were stimulus selective in isolation, response magnitudes to effective stimuli were significantly reduced in the presence of the complex surrounds. This effect was observed even when the monkeys looked directly at the target stimulus and correctly identified it. One possibility is that that response selectivity of cells, which appears to be plastic and modifiable by experience (Sakai et al., 1994; Logothetis et al., 1995; Booth and Rolls, 1998; Kobatake et al., 1998), must also adapt to respond under conditions of complex surrounds. In our experiments, the monkeys had repeatedly experienced targets that were both in isolation and embedded in scenes by the time-selective cells were recorded; presumably, many but not all stimulus-selective cells could have adapted to both conditions. One prediction of this hypothesis is that the amount of suppression observed in an experiment will depend on the level of experience the monkey has had with the test objects in complex environments.

Previous studies also suggest that in the course of visual search, the observed competitive effects may also be controlled by the active selection of targets by the perceiver for subsequent processing. Reynolds et al. (1999), for example, found that by directing the monkey to attend to a particular stimulus, the competitive

effects found for cells in early visual areas could be mitigated so that the response of cells was biased toward the response to the attended stimulus alone. Additionally, for face-selective cells in more anterior visual areas, Rolls and Tovee (1995) found that competition between stimuli was biased in favor of objects appearing at the fovea, possibly resulting from the overrepresentation of central vision in earlier cortical areas. A bias for stimuli projected onto the fovea is particularly relevant when one considers how the visual system naturally extracts information from complex scenes—with rapid shifts of gaze. As the eyes actively scan the environment, the representations of stimuli at the fovea may dominate over peripheral targets, thus providing one method for effectively transferring localized information from the visual system into either motor or memory systems (Rolls and Tovee, 1995).

The importance of eye movements during natural vision is obvious, given their ubiquity, but only a few studies have directly investigated how unconstrained fixation affects the activity of visual neurons (Livingstone et al., 1996; Gallant et al., 1998). A recent study (DiCarlo and Maunsell, 2000) of temporal cortical neurons reported no direct effects of saccadic eye movements on the neural selectivity for transiently presented targets. In that experiment, the authors compared responses to figures presented immediately after the monkey had executed a saccade ("free viewing") with responses observed when the same stimuli were presented during controlled fixation. They found essentially no difference between the two conditions and concluded that neuronal responses were indistinguishable between controlled and free viewing. In the present study, we were specifically interested in the interaction between eye movements, the natural visual scene, and the response of neurons as objects were noticed. In natural vision, eye movements serve the general function of bringing already present stimuli into the center of gaze and are not simply isolated motor acts. Under the conditions used in the current experiment, we found that the dynamics of the neural response in the isolated and free viewing conditions differed substantially (Fig. 6*A*), and we believe this difference to be both behaviorally relevant and crucial to our understanding of natural visual processing.

A comparison of two simplified models of how visual information may be extracted during exploration clarifies our position. One model would totally dissociate the process of selecting targets for fixation from the process of target identification. If this model were applied to the current task, we would predict that the process of identifying objects would begin after each fixation. In the isolated task, this would occur at the beginning of the trial, but in the embedded task, this would only begin once a target had been foveated. If this mode of processing were correct, we should have found no differences between the isolated and embedded conditions when the data were aligned to the time the target was fixated. Instead, an alternative model in which object identification can begin before the eyes actually fixate a target better accounts for our results. In this model, eye movements and shifts of visual attention are naturally coupled but not precisely synchronized (Kowler et al., 1995). There is clear physiological support for such a model because presaccadic modulation of neural activity dependent on the goal of impending saccades has been reported for cells in many cortical areas (Wurtz and Mohler, 1976; Robinson et al., 1978; Fischer and Boch, 1981; Colby et al., 1996). Recently, Moore and colleagues (Moore et al., 1998; Moore, 1999) found that cells in area V4 responded selectively both to the initial presentation of an optimally positioned bar and

just before a delayed saccade to the same stimulus. Because V4 is a major source of input to visual areas in the temporal lobe, convergent presaccadic activity arriving from this area is likely the basis for the early activation reported in this study. Furthermore, psychophysical studies have reported a significant benefit in naming latencies for visual objects previewed extrafoveally (Pollatsek et al., 1984; Henderson et al., 1987), and it is known that stimulus features can be used to guide saccadic eye movements (Motter and Belky, 1998; Moore, 1999). Our results provide strong evidence that neurons in the temporal lobes can begin processing specific peripheral targets before they are fixated, but only when they are the goal of the next saccade (Fig. 8).

The results from the present experiment augment our conclusions from a previous study in which we showed that during ambiguous stimulation, the activity of temporal cortical cells better correlates with the perceptual state of the animal than with the physical stimulus (Sheinberg and Logothetis, 1997). Here we have demonstrated that these stimulus-selective cells only become active when effective targets are actually noticed by the visual system. Although we cannot say what causal role these cells play in this process, their activity does seem tightly coupled to the process of transforming perceived wholes into learned reactions. Further studies of these cells and their interactions should prove useful in refining our definition of the elusive process that we call recognition.

REFERENCES

Booth MCA, Rolls ET (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb Cortex 8:510–523.

Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol 80:2918–2940.

Colby CL, Duhamel J-R, Goldberg ME (1996) Visual, presaccadic and cognitive activation of single neurons in monkey lateral intraparietal area. J Neurophysiol 76:2841–2852.

DeBusk BC, DeBruyn EJ, Snyder RK, Kabara JF, Bonds AB (1997) Stimulus-dependent modulation of spike burst length in cat striate cortical cells. J Neurophysiol 78:199–213.

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051–2062.

DiCarlo J, Maunsell JHR (2000) Form representation in monkey infero-temporal cortex is virtually unaltered by free viewing. Nat Neurosci 3:814–821.

Fischer B, Boch R (1981) Enhanced activation of neurons in prelunate cortex before visually guided saccades of trained rhesus monkeys. Exp Brain Res 44:129–137.

Gallant JL, Connor CE, Van Essen DC (1998) Neural activity in areas V1, V2, and V4 during free viewing of natural scenes compared to controlled viewing. NeuroReport 9:2153–2158.

Gross CG (1994) How inferior temporal cortex became a visual area. Cereb Cortex 5:455–469.

Hanes DH, Thompson KG, Schall JD (1995) Relationship of presaccadic activity in frontal eye field and supplementary eye field to saccade initiation in monkey: Poisson spike train analysis. Exp Brain Res 103:85–96.

Henderson JM, Pollatsek A, Rayner K (1987) Effects of foveal priming and extrafoveal preview on object identification. J Exp Psychol Hum Percept Perform 13:449–463.

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.

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.

Kowler E, Anderson E, Dosher B, Blaser E (1995) The role of attention in the programming of saccades. Vision Res 35:1897–1916.

Legéndy CR, Salcman M (1985) Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons. J Neurophysiol 53:926–939.

Lisman JE (1997) Bursts as a unit of neural information: making unreliable synapses. Trends Neurosci 20:38–43.

Livingstone MS, Freeman DC, Hubel DH (1996) Visual responses in V1

of freely viewing monkeys. Cold Spring Harb Symp Quant Biol 61:27–37.

Logothetis NK, Sheinberg DL (1996) Visual object recognition. Annu Rev Neurosci 19:577–621.

Logothetis NK, Pauls JM, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. Curr Biol 5:552–563.

Lueschow A, Miller EK, Desimone R (1994) Inferior temporal mechanisms for invariant object recognition. Cereb Cortex 5:523–531.

Matin E (1974) Saccadic suppression: a review. Psychol Bull 81:899–917.

Miller EK, Gochin PM, Gross CG (1993) Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. Brain Res 616:25–29.

Missal M, Vogels R, Li C-Y, Orban GA (1999) Shape interactions in macaque inferior temporal neurons. J Neurophysiol 82:131–142.

Moore T (1999) Shape representations and visual guidance of saccadic eye movements. Science 285:1914–1917.

Moore T, Tolias AS, Schiller PH (1998) Visual representations during saccadic eye movements. Proc Natl Acad Sci USA 95:8981–8984.

Motter BC, Belky EJ (1998) The guidance of eye movements during active visual search. Vision Res 38:1805–1815.

Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342.

Pollatsek A, Rayner K, Collins WE (1984) Integrating pictorial information across eye movements. J Exp Psychol Gen 113:426–442.

Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci 19:1736–1753.

Richmond BJ, Optican LM, Spitzer H (1990) Temporal encoding of two-dimensional patterns by single units in primary visual cortex. I. Stimulus-response relations. J Neurophysiol 64:351–369.

Robinson DA (1963) A method of measuring eye movement using a scleral search coil in a magnetic field. IEEE Trans Biomed Eng 10:137–145.

Robinson DL, Goldberg ME, Stanton GB (1978) Parietal association cortex in the primate: sensory mechanisms and behavioral modulations. J Neurophysiol 41:910–932.

Rolls ET, Tovee MJ (1995) The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. Exp Brain Res 103:409–420.

Sakai K, Naya Y, Miyashita Y (1994) Neuronal tuning and associative mechanisms in form representation. Learn Mem 1:83–105.

Sato T (1989) Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. Exp Brain Res 77:23–30.

Schwartz EL, Desimone R, Albright TD, Gross CG (1983) Shape recognition and inferior temporal neurons. Proc Natl Acad Sci USA 80:5776–5778.

Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci USA 94:3408–3413.

Sheinberg DL, Logothetis NK (1999) Activity of temporal cortical neurons during visual exploration in the monkey. Soc Neurosci Abstr 24:530.

Tanaka K (1997) Mechanisms of visual object recognition: monkey and human studies. Curr Opin Neurobiol 7:523–529.

Tanaka K, Saito H, Fukada H, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 66:170–189.

Tovee MJ, Rolls ET, Azzopardi P (1994) Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert monkey. J Neurophysiol 72:1049–1060.

Wurtz RH, Mohler CW (1976) Enhancement of visual response in monkey striate cortex and frontal eye fields. J Neurophysiol 39:766–772.

Yakovlev V, Fusi S, Berman E, Zohary E (1998) Inter-trial neuronal activity in inferior temporal cortex: a putative vehicle to generate long-term visual associations. Nat Neurosci 1:310–317.