

Paired neuron recordings in the prefrontal and inferotemporal cortices reveal that spatial selection precedes object identification during visual search

Ilya E. Monosov^{a,1}, David L. Sheinberg^b, and Kirk G. Thompson^a

^aLaboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD 20892; and ^bDepartment of Neuroscience, Brown University, Providence, RI 02906

Edited* by Thomas D. Albright, The Salk Institute for Biological Studies, La Jolla, CA, and approved June 7, 2010 (received for review March 5, 2010)

We addressed the question of how we locate and identify objects in complex natural environments by simultaneously recording single neurons from two brain regions that play different roles in this familiar activity—the frontal eye field (FEF), an area in the prefrontal cortex that is involved in visual spatial selection, and the inferotemporal cortex (IT), which is involved in object recognition—in monkeys performing a covert visual search task. Although the monkeys reported object identity, not location, neural activity specifying target location was evident in FEF before neural activity specifying target identity in IT. These two distinct processes were temporally correlated implying a functional linkage between the end stages of “where” and “what” visual processing and indicating that spatial selection is necessary for the formation of complex object representations associated with visual perception.

attention | cognition | object recognition | perception

We continuously search for and identify objects in highly complex and variable natural environments to guide behavior. An open question is whether selection of an object precedes identification, or whether object identification precedes visual selection. This debate is made evident by two theories of selective attention, early and late selection (1). Early selection theories postulate that visual selection precedes object identification (2, 3). In this scheme a location-based selection process controls which object has access to limited resources for further processing. Preattentive low-level feature analyses and internal top-down signals guide the selection process and object identification occurs after the object is selected. Alternatively, late selection theories postulate that object-related processing drives the spatial selection process, which directs attention to the task-relevant object for further processing (4, 5). Late selection predicts that a neural representation of an object's identity would precede the selection of the location of that object. Subsequently it has been noted that early and late theories of attentional selection are not mutually exclusive; there could be multiple loci of selection at different levels of processing (6–8).

We investigated the temporal relationship of selective processing that occurs between the end stages of visual processing related to “where” a visual object is and “what” it is. We simultaneously recorded the activity of single neurons in the frontal eye fields (FEFs) and the inferotemporal cortex (IT) of monkeys performing a covert visual search task in which they manually reported the identity of a learned target, regardless of its location (Fig. 1). FEF is a part of the dorsal frontoparietal attention network (9) and plays a key role in the visual spatial selection process. Spatially selective signals in FEF guide overt gaze shifts (10) and covert spatial attention (11). IT is the culmination of the ventral “what” visual processing stream and is necessary for object recognition (12, 13). IT is likely the first brain region in which images of natural objects, such as those we used in this study, are represented (14). IT neurons have spatially expansive receptive fields (RFs), generally including the central visual field and large extents of the contralateral hemifield (15, 16). In complex environments,

they become active as their preferred object becomes the target for a saccade (17). We addressed the question of which cognitive process happens first during visual search for a learned object—spatial selection, as predicted by the early selection hypothesis or object identification, as predicted by the late selection hypothesis.

Results

Data were collected from two monkeys (M1 and M2) performing the covert visual search task shown in Fig. 1. On each trial, a target was drawn from a pool of 20 targets previously associated with a left or right lever turn (Fig. 1A). The monkeys' task was to maintain fixation on a central stimulus and report the identity of the learned target object among distractors with a manual lever turn (Fig. 1B). A cue array of colored rings was presented 235 ms before the presentation of the search array. One object of the search array appeared in each of the cue array rings. In half of the trials, all of the rings were green (neutral cue trials). In the remaining trials, the location of the target object was indicated with a red ring (cued trials).

Based on the IT neuron's responses, we identified a *preferred* target object and a *nonpreferred* target object associated with opposite lever turn directions, and four distractor objects that elicited similar responses to the nonpreferred target using a passive viewing task (*SI Text*). On each trial, three of the four identified distractors were chosen at random. On 75% of the trials, the preferred target and the nonpreferred target had an equal chance of being chosen as the target of the search array. On 25% of the trials, the target was chosen at random from the 18 remaining objects in the target pool. These control trials were used to discourage the monkeys from adapting search strategies specific to the preferred and nonpreferred target objects chosen for that session (18).

We simultaneously recorded the activity of single neurons in FEF and IT during 24 recording sessions in two monkeys (M1: 20 neuron pairs, M2: 4 neuron pairs). More information about the locations at which the neuronal recordings were performed is provided in *Methods* and Fig. S1. The behavioral performance of the two monkeys was similar. Overall, the monkeys performed correctly on 94% of cued trials and 92% of neutral cue trials. Median reaction time across sessions was faster on cued trials than on neutral cue trials (mean \pm SD; neutral cue trials: 352 ± 56 ms; cued trials: 333 ± 56 ms), and this difference was significant (paired *t* test: $P < 0.001$). Performance accuracy and reaction times from the control trials with random target images did not

Author contributions: I.E.M. designed and performed research; and I.E.M., D.L.S., and K.G.T. analyzed data and wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: ilya.monosov@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002870107/-DCSupplemental.

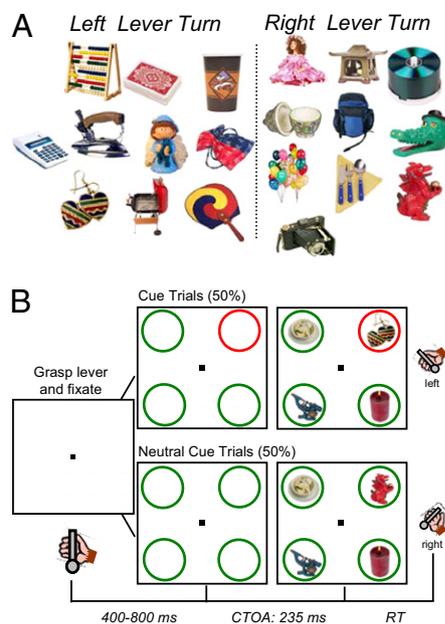


Fig. 1. Target objects, task timeline, and a depiction of valid and neutral cue trials. During each trial any one of 20 target stimuli (A) could appear along with three distractors, each located randomly at the four search array locations (B). While maintaining fixation on the central spot, the monkeys identified the target with a left or a right lever turn, on the basis of a learned arbitrary association. CTOA, cue target onset asynchrony.

differ significantly from that obtained from the preferred and nonpreferred target trials during each of the 24 recording sessions (*SI Text*). This result indicates that the monkeys' behavioral report was guided by an object recognition process rather than by the detection of low-level features specific to the preferred and nonpreferred target objects.

Neuron activity in FEF and IT was recorded simultaneously to compare the time course of spatial selection and object identification during the same trials. The term "spatial selection" refers to the neuronal activity in FEF that discriminates whether a target or a distractor is in the neuron's response field (10). "Object identification" refers to the neuronal activity in IT that differentiates between the preferred versus the nonpreferred target (18).

For the 24 recording sessions, the FEF neuron reliably signaled target location and the IT neuron reliably signaled target identity. The relationship between object identification in IT and the selection of the location of that object in FEF can be observed on neutral cue trials. Fig. 2 plots the average activity of an example neuron in neutral cue trials when the preferred target (Fig. 2*A* and *B*) and the nonpreferred target (Fig. 2*C* and *D*) is presented at each of the four possible target locations. After the presentation of the search array, the activity of the IT neuron differentiated which object was in the RF by exhibiting a robust response for the preferred object in the contralateral hemifield (Fig. 2*B*) and the absence of a response for the nonpreferred object at the same locations (Fig. 2*D*). The activity of the FEF neuron signaled the location of the target object by exhibiting greater activity when the target was in the RF than when a distractor was in the RF, regardless of which object was the target (Fig. 2*A* and *C*). Another example FEF–IT pair is shown in Fig. S2. The response of the IT neuron shown in Fig. S2 was insensitive to spatial location even though the timing, size, and eccentricity of the search and cue arrays were the same as for the neuron shown in Fig. 2.

For each neuron pair, we compared the time of object identification by the IT neuron to the time of spatial selection by the simultaneously recorded FEF neuron (Fig. 2*E* and *F*). Because

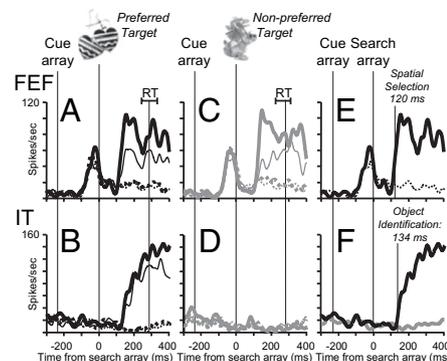


Fig. 2. Activity of an FEF–IT neuron pair recorded simultaneously in monkey 2 during neutral cue trials. The locations of the search array items were as shown in Fig. 1*B*. (Upper row, activity from the FEF neuron; Lower row, activity from the IT neuron.) (A and B) Activity from the FEF neuron (A) and the IT neuron (B) on trials that the IT neuron's preferred target was presented (black lines). (C and D) Activity from the FEF neuron (C) and the IT neuron (D) on trials that the IT neuron's nonpreferred target was presented (gray lines). Solid lines represent activity on trials in which the target was in the visual hemifield contralateral to the recording sites. Dashed lines show activity on trials in which the target was in the ipsilateral visual hemifield. Thick lines show activity on trials in which the target was in the upper visual hemifield. Thin lines show trials in which the target was in the lower visual hemifield. The range of lever turn reaction times (RT) are indicated above the plots; the median reaction time is shown by the vertical line. (E) Spatial selection time calculation. The time of spatial selection in FEF was calculated by comparing the activity for the preferred target (shown in plot A) presented at the location that evoked the maximum response (thick black solid line) to that associated with the minimum response (thin black dashed line). The time of spatial selection for this neuron is 120 ms. (F) Object identification time calculation. The time of object identification in IT was calculated by comparing the strongest activity for the preferred target (thick black line in plot B) to the activity in trials in which the nonpreferred target (thick gray line in plot D) was at the same search array location. The time of object identification for this neuron is 134 ms.

the responses of many IT neurons vary with location (15, 16, 19), we structured this analysis so that we would obtain the earliest possible time of object identification by the IT neuron. The time of object identification in IT was calculated during trials in which the preferred target was presented at the array location that elicited the greatest and earliest response from the IT neuron, and was defined as the first time when the preferred target produced a significantly greater response than the nonpreferred target (Fig. 2*F*). For the analysis of spatial selection in FEF, we were constrained by the spatial extent of the FEF neuron's receptive field, which was limited to one or sometimes two array locations. The time of spatial selection in FEF was defined as the first time the FEF neuron's activity during trials in which the preferred target was inside the receptive field was significantly greater than the activity during trials in which the preferred target was outside of the receptive field (Fig. 2*E*). The target locations for which the FEF and IT activity was analyzed did not match in 8 of the recording sessions from monkey 1 and in one session from monkey 2. The results from these 9 sessions did not differ from the results from the remaining 15 recording sessions in which the best locations matched. The FEF and IT neuron pair shown in Fig. 2 exhibited the strongest responses to the preferred target at the same location in the search array. For this neuron pair, spatially selective activity in the FEF neuron became significant at 120 ms (Fig. 2*E*) and object selective activity became significant at 134 ms (Fig. 2*F*) following the presentation of the search array.

The time course of object identification and spatial selection on cued and neutral cue trials can be seen in the pooled average activity from all 24 recording sessions shown in Fig. 3*A–D*. The

average differences in activity on cued and neutral cue trials that define the spatially selective response in FEF and the object selective response in IT are compared in Fig. 2 *E* and *F*. On cued trials, a red ring cued the location of the target. Because activity in FEF signaled the presence of the red cue ring in the receptive field (11), the time of spatial selection on cued trials measures the selection of the cue stimulus and occurs well before the appearance of the target array (Fig. 3*A*).

A comparison of the means \pm SDs of spatial selection times in FEF, object identification times in IT, and reaction times on cued and neutral cue trials is plotted in Fig. 3*G*. The average time of spatial selection (mean \pm SD) measured from FEF activity on cued trials was -139 ± 21 ms relative to the presentation of the search array. The average time of spatial selection in FEF on neutral cue trials was 116 ± 24 ms after the presentation of the search array. Reaction times were, on average, 18.5 ms faster on cued trials than on neutral cue trials. This indicates that spatial attention was directed to the spatial cue. The time of object

identification in IT on cued trials was 18.3 ms faster than on neutral trials (cued trials: 133 ± 26 ms; neutral cue trials: 151 ± 31 ms). The nearly equivalent difference demonstrates that the presence of an attentional cue affects the latency of object identification in IT and the reaction time difference can be attributed to the time it takes to form an object representation in IT.

For the remainder of this report we concentrate on neutral cue trials, trials that produced only target-related spatial selection in FEF neuron activity (Fig. 3*B*). From the activity recorded on neutral cue trials we were able to compare the time course of object identification for the preferred target in IT to the time course of spatial selection elicited by that target in FEF. Overall, spatial selection in FEF preceded object identification in IT by 34 ± 27 ms (Fig. 3*G*). After the presentation of the search array, some IT neurons had an initial nonselective visual response (Fig. S2). It may be argued that across the population of IT neurons, object identification occurs at the time of IT visual onset. Considering this possibility, we compared the IT visual onset times evoked by the presentation of the search array (relative to the activity before the search array presentation) to the times of FEF spatial selection. IT visual onset times occurred after FEF selected the location of the target to be identified ($P = 0.02$; *SI Text*).

Reaction times varied from session to session, most likely because the stimuli used in each session were tailored to the IT neuron's response. This variation in the combinations of targets and distractor stimuli changed the difficulty of the task from day to day. This variability allowed us to examine the temporal relationships between reaction time, object identification in IT, and spatial selection in FEF across recording sessions. Both the spatial selection times in FEF and the object identification times in IT were correlated with the median reaction times across sessions (Fig. 4*A*; FEF vs. reaction time: $r = 0.42$, $P = 0.04$, IT vs. reaction time: $r = 0.52$, $P = 0.009$). This suggests that the neural activity in FEF and in IT contributed to the object identification process leading to the behavioral report in our task. The stronger correlation to reaction time for IT than for FEF may indicate that neuronal activity in IT represents a processing stage that has a greater influence over the object identification process than does FEF.

Next, we performed a correlation analysis between the spatial selection times obtained from FEF and the object identification times obtained from IT using activity recorded on the same trials (Fig. 4*B*). The best locations for the FEF neuron and the IT neuron matched for 15 of the 24 FEF–IT neuron pairs, as they do for the FEF–IT neuron pair shown in Fig. 2. For these 15 neuron pairs (M1: 12 pairs; M2: 3 pairs), the strongest object selective response in IT was obtained at a search array location inside the FEF neuron's receptive field. These 15 neurons are indicated by the solid data points in Fig. 4*B*. For an additional five neuron pairs (M1: 4 pairs; M2: 1 pair) the FEF neurons' receptive field overlapped with a location at which the preferred object elicited a response from the IT neuron that was nearly as strong as when that object was at the IT neuron's best location. The object selection times used for these IT neurons were, on average, 12.5 ms later than those obtained from the best location and are indicated by the open data points in Fig. 4*B*. We could not obtain object selection times from both the IT neuron and the FEF neuron from the same trials for the remaining 4 sessions, so they were omitted from the correlation analysis. For the 20 neuron pairs with overlapping receptive fields, the spatial selection times in FEF were correlated with object identification times in IT ($r = 0.59$, $P = 0.007$). The correlation was significant when we included only the 15 neuron pairs that exhibited the best responses at the same array locations ($r = 0.65$, $P = 0.009$). For the 15 recording sessions in which the best locations matched, the spatial selection in FEF occurred, on average, 35 ± 25 ms before object identification in IT. It is important to note that the objects used in the task differed from day to day and the monkeys could have used different

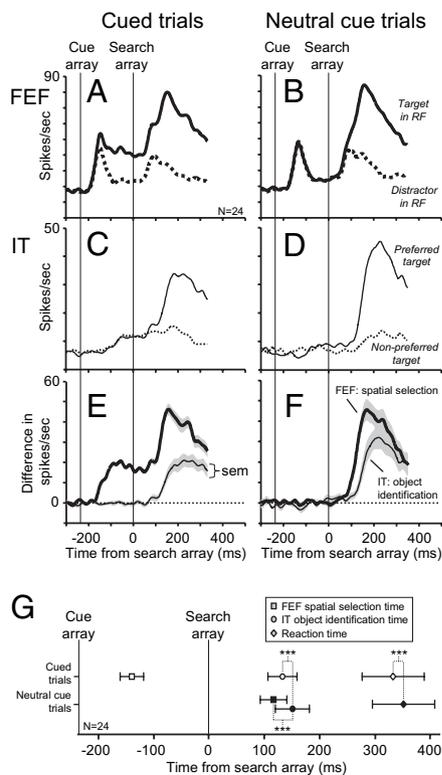


Fig. 3. Population average activity showing the time course of spatial selection and object identification across the 24 FEF–IT neuron pairs. (*A* and *B*) Population average FEF activity during cued (*A*) and neutral cue (*B*) trials. Thick solid line shows average FEF activity during trials in which the IT neurons' preferred target was in the receptive field of the FEF neuron. Thick dashed line shows average FEF activity during trials in which the preferred target was presented outside of the receptive field and a distractor was in the receptive field. (*C* and *D*) Average IT activity from the same recording sessions during cued (*C*) and neutral cue (*D*) trials. Thin solid line shows average IT activity for the preferred target at the location that elicited the best response from the IT neuron. Thin dashed line shows average IT activity for the nonpreferred target at the same location. (*E* and *F*) Activity difference plots capturing the time course of spatial selection and object identification during cued (*E*) and neutral cue trials (*F*). Thick lines represent mean difference of FEF activity shown in *A* and *B*. Thin lines represent mean difference of IT activity shown in *C* and *D*. SE around the mean for each plot is shown in gray shading. (*G*) Average \pm SD of spatial selection time in FEF (squares), object identification time in IT (circles), and median reaction time (diamonds) from the 24 recording sessions. Open symbols are from cued trials. Solid symbols are from neutral cue trials. *** $P < 0.001$, significant differences from paired *t* tests.

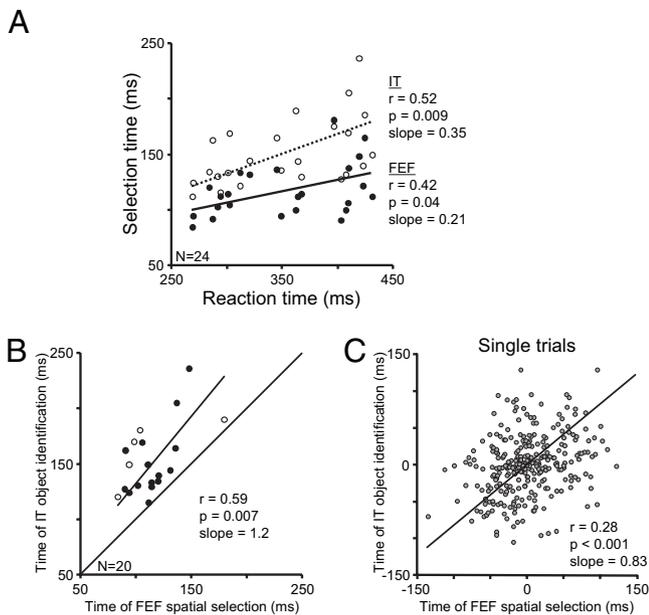


Fig. 4. Correlation analyses. (A) Significant correlations of spatial selection in FEF (solid circles; $r = 0.42$, $P = 0.04$) and object identification in IT (open circles; $r = 0.52$, $P = 0.009$) to reaction time. Results from the 24 FEF–IT neuron pair recording sessions are included. The best fit lines through the data points were calculated by minimizing the orthogonal least-squares distance of each point (FEF: solid line, slope = 0.21; IT: dotted line, slope = 0.35). (B) Significant correlation of spatial selection time in FEF and object identification time in IT across recording sessions from simultaneously recorded trials ($r = 0.59$, $P = 0.007$). Each data point plots the time of object identification in IT as a function of the time of spatial selection of that object in FEF calculated from the same trials recorded in a single recording session ($n = 20$). Solid circles are from recording sessions in which the object identification time of the IT neuron was obtained from the location with the best response. Open circles are from recording sessions in which the object identification time of the IT neuron was obtained from the location with the second best response. The best fit line through the data points is shown (slope = 1.21). (C) Significant partial correlation of spatial selection time in FEF and object identification time IT calculated from the same single trials after subtracting the session means from each data point ($r = 0.28$, $P < 0.001$). Each data point plots the time of object identification from the IT neuron as a function of the time of spatial selection from the simultaneously recorded FEF neuron on the same single trial. The best fit line through the data points is shown (slope = 0.83).

strategies to find the target. Nevertheless, all of the data points in Fig. 4B are above the identity line, indicating that spatial selection in FEF occurred before object identification in IT on the same trials in every experimental session.

It could be argued that the correlation between spatial selection time in FEF and object identification time IT (Fig. 4B) is coincidental, a product of the correlations with reaction time and variations in task difficulty across sessions (Fig. 4A). To address this question, we examined whether there was a correlation between the spatial selection time in FEF and object identification time in IT on individual trials within recording sessions. Simultaneous activity recorded in FEF and IT was analyzed to obtain a spatial selection time and an object identification time for the same trial. The single trial selection time analysis is discussed and illustrated in Fig. S3. We performed a partial correlation analysis on the single trial timing results that removed differences in timing across sessions (Fig. 4C). We included only those sessions in which we obtained a spatial selection time from the FEF neuron and an object identification time from the IT neuron for at least 10 simultaneously recorded trials. This yielded a total of 290 trials from 14 sessions (M1: 10 sessions; M2: 4 sessions). The

partial correlation analysis removed the across-session correlation (shown in Fig. 4B) by subtracting the session means from the spatial selection time and object identification time obtained from each single trial. There was a strong correlation between the times of spatial selection in FEF and object identification in IT on single trials (M1: $r = 0.25$, $P < 0.0001$; M2: $r = 0.42$, $P < 0.0003$; both: $r = 0.28$, $P < 0.00001$). A similar result was obtained when selection times from all sessions, including those with <10 simultaneously recorded single trials, were included in the partial correlation analysis ($n = 321$, $r = 0.22$, $P < 0.0001$). In addition, we performed a single trial Pearson's correlation analysis separately for each session. The correlation coefficients were >0 for all 14 sessions (mean $r = 0.33 \pm 0.06$), of which 4 were significant (M1: 2 sessions; M2: 2 sessions; $P < 0.05$).

Discussion

We simultaneously recorded neural signals in FEF and IT representing two distinct but related cognitive processes leading to a behavioral report of visual object identification during covert visual search without eye movements. The selective signals observed in each area were consistent with those that have been described previously. FEF activity localizes the target object to be identified and is correlated with spatial attention (11, 20), and IT activity represents object identity and is correlated with object recognition (13, 18). The emergence of spatially selective signals in FEF and object representations in IT in our task was correlated with variations in the time it took the monkeys to report target identity across recording sessions with varying levels of difficulty (Fig. 4A). This is also consistent with the results of previous studies (18, 21). The correlations with reaction time suggest that the neuronal signals observed in both of these areas play a role in the target identification process. It should be noted that the correlations to reaction time do not necessarily demonstrate a causal relationship between FEF and IT to choice behavior. In our view, the activity recorded simultaneously in FEF and IT allows us access to the outcome of the brain processes responsible for spatial selection and object recognition, respectively.

Recently we demonstrated that the magnitude of FEF activity is correlated with the speed and accuracy of object identification and inactivation of FEF produced deficits in object identification (11). Here we analyzed concurrent activity recorded in FEF and IT on the same trials providing insight about the neural processes leading to object identification that cannot be addressed from the analysis of either signal alone. The emergence of object representations in IT and the time of selection of that object's location in FEF on the same trials are correlated across recording sessions (Fig. 4B) and on individual trials within recording sessions (Fig. 4C). The task used in this study did not explicitly require a spatial component; the monkeys reported target object identity. Nevertheless, visual spatial selection signals in FEF were present about 35 ms before object identification signals emerged in IT. This suggests that spatial selection is an integral part of the object recognition process and is consistent with the "early selection" hypothesis of attentional selection, that an object must be within the focus of spatial attention to bind simple visual features (e.g., color, shape) into a complex object representation (22–26).

We found significant correlations between the time of neuronal object identification and the monkeys' reaction times across recording sessions whereas other studies have found little covariation between reaction time and IT neuronal activity (27–29). In these other studies, the task was relatively easy; the monkeys identified a single stimulus at the center of the screen. In contrast, we required the monkeys to perform a difficult search task in which the monkeys had to covertly search for the object among distractors and identify it; and on any given trial any one of 20 targets could appear. Consistent with our study, a recent study also found a relationship between object-related activity in IT and reaction time in monkeys performing a near threshold object

recognition task (18). We hypothesize that differences in task difficulty can explain why some studies observe a correlation between reaction time and IT activity and others do not.

The results of our study seem to be at odds with the results of studies by Chelazzi and colleagues (30, 31). In these studies, neurons in IT were recorded in monkeys performing a memory-guided visual search task in which a cue stimulus was followed, after a delay, by a search array with two or more stimuli. The animal was rewarded for making a saccade to the target stimulus in the search array that matched the cue stimulus. The target stimulus could be “good” or “poor” depending on how effective it was in driving the neuron’s activity. They showed that there was strong early activation for the good stimulus even when it was a distractor. The activity of the IT neuron then changed to signal the poor stimulus as the poor stimulus was selected as the target for the saccade. Their results are consistent with the “late selection” hypothesis in which there is parallel processing of both the target and distractor stimuli before the selection of the target stimulus that matched the previous sample.

Our results and those of Chelazzi et al. (31) can be reconciled by considering differences in the tasks and by the fact that they did not simultaneously record spatially selective activity in an area such as FEF that indexes the outcome of the spatial selection process. Also, the memory-guided search task used by Chelazzi et al. (31) relied on visual working memory. The monkeys were shown the item to search for at the beginning of every trial, and the IT neurons exhibited elevated activity related to this knowledge. Visual working memory for the target object was not a factor in our task.

On the basis of Chelazzi et al. (31), Hamker (32) proposed the *reentry model* of visual search in which spatial selection by FEF neurons is required for object detection and identification (32). Predictions of the reentry model are in line with our finding that the times of spatial selection in FEF and object identification in IT are correlated on single trials (Fig. 4C). Hamker (32) constrained this model with the assumption that neurons in IT compete in parallel and that the spatial signal in FEF resolves this competition, which results in the perception of the target. We found that the neural representation of the preferred target object’s identity in IT emerges after spatial selection has taken place.

It has been suggested that there are different types of selection for different cognitive processes (6–8). It is possible that the matching of an object representation held in working memory requires an object representation in IT and is thereby consistent with a late selection mechanism. We showed that the emergence of these object representations in IT require spatial selection signals and are therefore consistent with an early selection mechanism. This hypothesis predicts that distractors in our task are not recognizable unless they are first selected by spatial selection processes. Whether this prediction holds true will require further experiments. But this view is consistent with the results from a study by Sheinberg and Logothetis (17). Although their study could not determine whether spatial selection or object identification happens first, they showed that object representations emerge in IT when viewing complex visual scenes only when the preferred object becomes the target for a saccade.

The purpose of this study was to determine the order of cognitive events leading to object recognition in a complex visual scene. We chose to record neuronal activity in FEF and IT because they are known to represent cognitive processes related to spatial selection and object identification, respectively. FEF forms an interconnected network with brain regions such as the lateral intraparietal area and the superior colliculus that exhibit similar patterns of spatially selective signals during visual search tasks (33–35). We hypothesize that the results would be similar if spatial selection times had been ascertained from neuronal activity recorded in these other structures.

Object identity and category can be read out from the activity of IT neurons by classifier models with great precision (36)—some-

thing not generally found for “earlier” visual areas (e.g., V1, V4). The active disambiguation of competing object representations is much more closely related to activity in IT than in earlier visual areas (37). More recent work has shown that choices in the face of uncertainty link IT activity to recognition decisions in a paradigm quite similar to the one used in this task (18). IT activity has been linked to the representation of knowledge in an object association task (38). When monkeys make saccades to a target to be identified in an active search task (17) neurons fire before the saccade is initiated signaling the noticing of the preferred stimulus. Nevertheless, we cannot be certain that the time of object identification measured in IT activity is the time at which the animal has consciously “recognized” the target object. Our paper sheds light on the order of cognitive events leading up to object recognition, but does not clearly link IT activity to conscious recognition of objects.

In summary, we show that during a visual search task in which the monkeys were rewarded for reporting target identity, not location, activity in the brain specifies where a target object is before identifying what it is. Our results are consistent with those of a recent psychophysical study that showed that attentional selection is a necessary prerequisite for target identification (39). The spatial selection process indexed by FEF activity could be guided by a preattentive low-level visual feature analysis or by an endogenous serial search process. Regardless of what guides the spatial selection process, our results provide a neural basis for early selection theories that hypothesize that the selection of an object’s location is necessary for its identification. The spatial attentional selection represented in FEF may facilitate the perceptual binding of features to object representations in IT (23–26).

Methods

General Procedures. Data were collected from two male monkeys (*Macaca mulatta*). All procedures were performed in accordance with the US Public Health Service policy on the humane care and use of laboratory animals and all protocols were approved by the National Eye Institute Animal Care and Use Committee. The activity of single neurons in FEF and IT were recorded simultaneously with tungsten microelectrodes. Behavioral control, stimulus presentation, monitoring of eye and lever position, and neuronal recording procedures were described previously (20).

For each experimental session a single microelectrode was advanced into FEF with a motorized microdrive, and a second microelectrode was advanced into IT with a second motorized microdrive. The microdrives were under computer control. Action potential waveforms were digitized and saved using a computer-based data acquisition system (Plexon). First, a single unit was isolated in FEF. Then while maintaining isolation on the FEF neuron, a single unit was found in IT. The search for a suitable IT neuron often took up to 90 min. For an FEF–IT neuron pair to be included in the analysis, the FEF neuron had to exhibit spatially selective activity for the target of the search array and the IT neuron had to exhibit object selective activity for one of the target objects. Once a suitable FEF–IT neuron pair was found, isolation on each of the neurons had to be maintained for about 1 h, during which time the monkey had to continue performing the task, to collect enough data for analysis. Approximately 250 recording sessions yielded 24 usable FEF–IT pairs.

The stimulus library and the localization of IT were described previously in detail (18). A passive viewing task in which stimuli were presented at the center of the screen was used to identify a preferred and a nonpreferred object for the IT neuron (18). All IT neurons that were activated by a target in the passive viewing task were also activated by that target in the cued covert search task. This is consistent with previous reports that showed that the object tuning of IT neurons remains relatively unchanged whether the objects are presented alone or in complex natural scenes (16, 17). The 24 IT neurons included in this report displayed strong object identification (t test, $P < 0.001$) in the visual search task.

Recording locations in FEF and IT were verified with MRI in M1 and histologically in M2. We sampled IT within a ~5-mm region along the posterior–anterior axis (Fig. S1). The time of object identification was not correlated with the recording site along the posterior–anterior axis.

FEF was localized with low current microstimulation to evoke saccades and the presence of saccade-related neurons (40). The memory-guided saccade task was used to map the spatial extent of the FEF neuron’s RF (40). The search array was arranged so that at least one stimulus was in the FEF RF;

eccentricities ranged between 8° and 10° of visual angle. The 24 FEF neurons included in this study were visually responsive (20).

Spike density functions were calculated for each trial by convolving spike times with a Gaussian filter ($\sigma = 10$ ms). Spatial selection and object identification times were calculated using a *t* test every millisecond. Pearson correlations were used to test relatedness between measurements.

Details regarding the monkeys' task and training, the calculation of spatial selection and object identification times for single sessions, and the

calculation of spatial selection and object identification times for single trials can be found in [SI Text](#).

ACKNOWLEDGMENTS. We thank Drs. Ethan S. Bromberg-Martin, Bruce G. Cumming, Okihide Hikosaka, Masaharu Yasuda, Robert H. Wurtz, and Jeffrey D. Schall for helpful discussion and M. K. Smith, G. Tansey, D. Parker, B. Nagy, and J. W. McClurkin for technical assistance. This work was supported by the intramural research program at the National Eye Institute, National Institutes of Health.

1. Johnston WA, Dark VJ (1986) Selective attention. *Annu Rev Psychol* 37:43–75.
2. Treisman AM (1969) Strategies and models of selective attention. *Psychol Rev* 76: 282–299.
3. Broadbent DE (1982) Task combination and selective intake of information. *Acta Psychol (Amst)* 50:253–290.
4. Deutsch JA, Deutsch D (1963) Some theoretical considerations. *Psychol Rev* 70:80–90.
5. Duncan J (1980) The locus of interference in the perception of simultaneous stimuli. *Psychol Rev* 87:272–300.
6. Yantis S, Johnston JC (1990) On the locus of visual selection: Evidence from focused attention tasks. *J Exp Psychol Hum Percept Perform* 16:135–149.
7. Vecera SP, Farah MJ (1994) Does visual attention select objects or locations? *J Exp Psychol Gen* 123:146–160.
8. Evans KK, Treisman A (2005) Perception of objects in natural scenes: Is it really attention free? *J Exp Psychol Hum Percept Perform* 31:1476–1492.
9. Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.
10. Schall JD, Thompson KG (1999) Neural selection and control of visually guided eye movements. *Annu Rev Neurosci* 22:241–259.
11. Monosov IE, Thompson KG (2009) Frontal eye field activity enhances object identification during covert visual search. *J Neurophysiol* 102:3656–3672.
12. Riesenhuber M, Poggio T (2002) Neural mechanisms of object recognition. *Curr Opin Neurobiol* 12:162–168.
13. Logothetis NK, Sheinberg DL (1996) Visual object recognition. *Annu Rev Neurosci* 19: 577–621.
14. Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189.
15. Op De Beeck H, Vogels R (2000) Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426:505–518.
16. Rolls ET, Aggelopoulos NC, Zheng FS (2003) The receptive fields of inferior temporal cortex neurons in natural scenes. *J Neurosci* 23:339–348.
17. Sheinberg DL, Logothetis NK (2001) Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *J Neurosci* 21:1340–1350.
18. Mruczek RE, Sheinberg DL (2007) Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *J Neurosci* 27:2825–2836.
19. Mruczek RE, Sheinberg DL (2007) Context familiarity enhances target processing by inferior temporal cortex neurons. *J Neurosci* 27:8533–8545.
20. Thompson KG, Bischof KL, Sato TR (2005) Neuronal basis of covert spatial attention in the frontal eye field. *J Neurosci* 25:9479–9487.
21. Sato T, Murthy A, Thompson KG, Schall JD (2001) Search efficiency but not response interference affects visual selection in frontal eye field. *Neuron* 30:583–591.
22. Rensink RA (2002) Change detection. *Annu Rev Psychol* 53:245–277.
23. Wolfe JM, Cave KR, Franzel SL (1989) Guided search: An alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform* 15:419–433.
24. Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12:97–136.
25. Itti L, Koch C (2001) Computational modelling of visual attention. *Nat Rev Neurosci* 2: 194–203.
26. Treisman A (1998) Feature binding, attention and object perception. *Philos Trans R Soc Lond B Biol Sci* 353:1295–1306.
27. DiCarlo JJ, Maunsell JH (2005) Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. *J Neurophysiol* 93:2974–2986.
28. Matsumoto K, Tanaka K (2004) The role of the medial prefrontal cortex in achieving goals. *Curr Opin Neurobiol* 14:178–185.
29. Eifuku S, De Souza WC, Tamura R, Nishijo H, Ono T (2004) Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *J Neurophysiol* 91: 358–371.
30. Chelazzi L, Miller EK, Duncan J, Desimone R (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363:345–347.
31. 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.
32. Hamker FH (2005) The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cereb Cortex* 15:431–447.
33. Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315:1860–1862.
34. Shen K, Pare M (2007) Neuronal activity in superior colliculus signals both stimulus identity and saccade goals during visual conjunction search. *J Vis* 7(5):1–13.
35. Balan PF, Oristaglio J, Schneider DM, Gottlieb J (2008) Neuronal correlates of the set-size effect in monkey lateral intraparietal area. *PLoS Biol* 6:e158.
36. Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310:863–866.
37. Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci USA* 94:3408–3413.
38. 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.
39. Zenon A, Ben Hamed S, Duhamel JR, Olivier E (2008) Spatial and temporal dynamics of attentional guidance during inefficient visual search. *PLoS ONE* 3:e2219.
40. Bruce CJ, Goldberg ME (1985) Primate frontal eye fields. I. Single neurons discharging before saccades. *J Neurophysiol* 53:603–635.

Supporting Information

Monosov et al. 10.1073/pnas.1002870107

SI Results

Visual Response Latencies in FEF and IT. After the presentation of the search array, some of the IT neurons had an initial nonselective visual response. It could be argued that object identification occurs at the time of IT visual response onset. We determined the latency of the initial visual responses in FEF and IT with a *t* test at each millisecond following the time of the search array presentation. The comparison was made to baseline activity, which was defined as the average firing rate during the 100 ms before the presentation of the search array. The running *t* test estimated the probability (*P*) at each millisecond that the activity after the presentation of the search array did not differ from baseline. The statistical threshold was exactly the same as that used to find the times of object identification and spatial selection. On average, the initial visual response was detected at 78.9 ± 7.5 ms (mean \pm SE) in FEF and 138.9 ± 7.4 ms in IT. Response latencies in FEF and IT were significantly different (paired *t* test, $P < 0.001$). We compared the initial response latency in IT to the time of spatial selection in FEF. The average time of FEF spatial selection was 116 ± 3.7 ms (mean \pm SE). Spatial selection time in FEF was significantly earlier than the initial visual response in IT (paired *t* test; $P = 0.02$).

Performance Accuracy and Reaction Time During Control Trials. Performance accuracy and reaction times from the control trials with random target images did not differ significantly from those obtained from the *preferred* and *nonpreferred* target trials during each of the 24 recording sessions (performance accuracy: χ^2 test, $P > 0.1$; reaction time: *t* test, $P > 0.1$). Average performance accuracy was 94% on cued control trials, and 91% on control neutral cue trials. Average reaction time was 334 ± 58 ms (mean \pm SD) on cued control trials and 357 ± 61 ms on control neutral cue trials.

SI Methods

Calculation of Spatial Selection and Object Identification Times for Single Sessions. Fig. 2 illustrates the selection time analysis used to determine selection times for single sessions. The single session selection times were used in the analyses shown in Figs. 3*G* and 4*A* and *B* of the main text. To calculate the selection time for a single recording session, we identified the preferred location for both the FEF and the IT neuron, the target location that produced the greatest and earliest activity. Spatial selection time for an FEF neuron was calculated by comparing the average activity on trials in which the preferred target (based on IT neuron activity) was in the FEF neuron's receptive field to the average activity when the target was at the opposite location and a distractor was in the FEF neuron's receptive field (Fig. 2*E*). Object identification time for an IT neuron was calculated by comparing the average activity on trials in which the preferred target was at its preferred location to the average activity when the nonpreferred target was at its preferred location (Fig. 2*F*). This method of determining object identification time was necessary for two reasons. First, it controls for activity differences related to target location. It reveals differential activity related solely to object identity because the activity for both conditions is from trials when the target was presented at the same location in the search array. Second, a target- vs. distractor-related activity comparison was not possible for most of the IT neurons because the preferred target elicited a response at all search array locations (Fig. S2).

We first generated a spike density function for each trial by convolving spike times with a Gaussian filter ($\sigma = 10$ ms). Spatial

selection and object identification times on single sessions were determined with a *t* test at each millisecond following the time of search array presentation. The running *t* test estimated the probability (*P*) at each millisecond that the observed difference between the two conditions was due to chance. The spatial selection and object identification times were defined as the first millisecond that the *P* value dropped below the 0.05 level only if it continued past the 0.001 level and remained below the 0.05 level for the next 25 ms. To obtain the earliest possible spatial selection and object identification times, a threshold of $P = 0.05$ was used. A threshold of $P = 0.01$ did not alter the results.

Calculation of Spatial Selection and Object Identification Times for Single Trials. For each FEF–IT neuron pair we calculated the spatial selection time from the FEF neuron and the object identification time from the IT neuron during simultaneously recorded single trials in which the preferred target (based on the IT neuron's activity) was at the preferred location (based on the FEF neuron's activity). The results of this analysis are shown in Fig. 4*C*. For this analysis spike density functions were first calculated for each trial by convolving spike times with a Gaussian filter ($\sigma = 10$ ms).

For the FEF neuron, single trial activity was compared with the average activity on trials when the target was at the opposite location and a distractor was at the preferred location. For the IT neuron, single trial activity was compared with the average activity on trials when the nonpreferred target was at the preferred location. A running *z* test estimated the probability (*P*) at each millisecond that the activity on the single trial did not differ from the baseline average activity. Single trial spatial selection and object identification times were defined as the first millisecond that the *P* value dropped below the 0.05 level only if it continued past the 0.001 level and remained below the 0.05 level for the next 25 ms. To obtain the earliest possible spatial selection and object identification times, a threshold of $P = 0.05$ was used. A threshold of $P = 0.01$ did not alter the results. Note that our variance-based analysis compares activity for two task conditions for the same neuron. The same thresholds were used for both signals.

Neuronal activity on single trials is noisy. Therefore, selection times from single trials are noisy and are less reliable than those obtained from comparing activity across multiple trials. In addition, the firing rate on many trials did not reach significance. This may be a particular concern for neurons with very low firing rates. An example of one such neuron recorded in IT is shown in Fig. S3. Although the average firing rate is low, the preferred target often elicited a brief burst of activity and, therefore, we were able to obtain an object identification time for a majority of the trials.

A total of 532 trials met the requirement that the preferred target was at the preferred location. We were able to obtain spatial selection times for 367 trials (69%) from FEF neurons and object identification times for 412 trials (77%) from IT neurons.

Although single trial analysis is inherently noisy, we still found a strong and significant correlation between the times of spatial selection and object identification (Fig. 4*C*).

Task and Training. Initially, the monkeys were trained to identify 20 target objects presented at the center of the screen with a lever turn, 10 of which required a left lever turn and 10 of which required a right lever turn (Fig. 1*A*). A set of 50 other objects served as distractors.

After the object-lever turn association training, the monkeys learned the cued covert search task (Fig. 1*B*). In this task, after

grasping the lever and positioning it within 10° of vertical, a small (0.3°) central fixation spot appeared on a black background. The monkeys were required to maintain eye position within 2° of the central fixation cross until the reward. After the monkeys fixated the central cross, a cue array of four rings appeared. The cue array rings were isoecentric and were 3.5° in diameter. In about half of the trials one of the four rings was red and the rest were green (cued trials). In the other half of the trials all of the rings were green (neutral cue trials). After a 235-ms cue-target onset asynchrony (CTOA) an object appeared inside each ring. One of the objects was sampled from the target object pool (Fig. 1A) and the remaining objects were sampled from the distractor pool. In cued trials, the target object appeared in the red ring. While maintaining fixation on the central cross, the monkeys reported the identity of the target with the associated left or right lever turn for a juice reward. Reaction time was measured from the time of the presentation of the search array objects to the beginning of the lever turn, defined as a turn $>15^\circ$ from vertical.

1. Mruczek RE, Sheinberg DL (2007) Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *J Neurosci* 27: 2825–2836.

At the beginning of each recording session a passive viewing task (1) was used to identify two target objects associated with opposite lever turns such that the IT cell was consistently activated by one target (the preferred target) more than by the other target (the nonpreferred target). Four distractor objects were also chosen that evoked similar activity from the IT neuron as the nonpreferred target object. See the population response shown in Fig. 2E; the IT neurons were mainly driven by the presence of the preferred target.

On each trial, a target was chosen from the target pool. There was a 37.5% chance that the target was the preferred object, a 37.5% chance that it was the nonpreferred object, and a 25% chance that it was randomly chosen from the 18 remaining objects in the target pool (control trials). The three distractors on each trial were chosen at random from the four previously identified distractor objects. The locations of the target and each of the distractors were randomized from trial to trial.

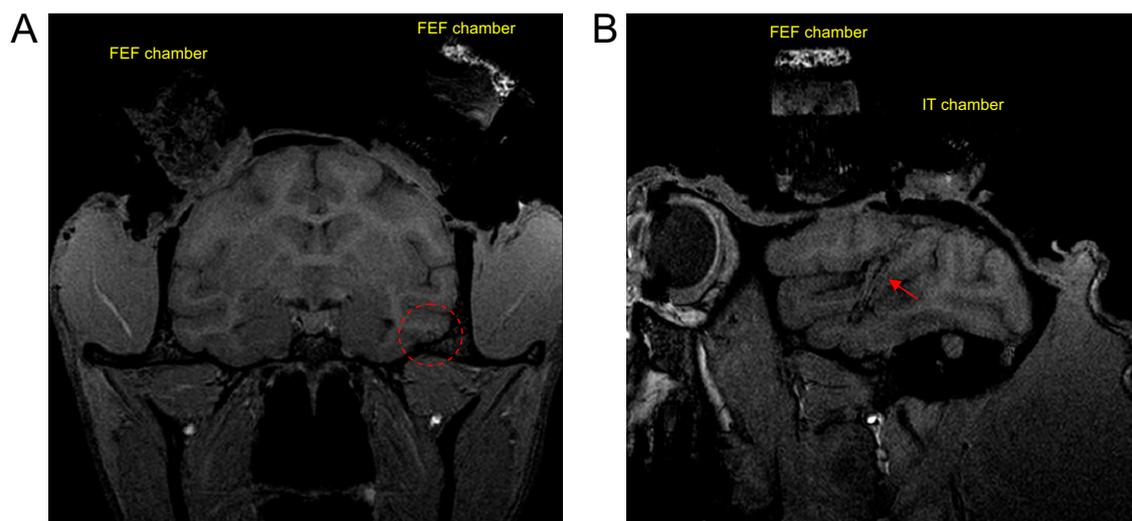


Fig. S1. Two MRI images of electrode penetration damage in monkey 1 showing the approximate trajectory of electrodes. Recording sites range from approximately +15 to +20 anterior in monkey 2, and approximately +17 to +21 anterior in monkey 1. Damage caused by the electrode is visible at the center of the red circle in *A*. This damage in IT is approximately +18 mm anterior of the interaural position. The arrow in *B* points to damage caused by the guide tube directed toward anterior IT. The IT chamber is not visible in *A* because it is aimed at anterior IT at an angle (from behind, see *B*). The visible bilateral chambers were for FEF recordings.

