Behavioral/Systems/Cognitive

Neural Dynamics in Inferior Temporal Cortex during a Visual Working Memory Task

Luke Woloszyn and David L. Sheinberg
Department of Neuroscience, Brown University, Providence, Rhode Island 02912

Intelligent organisms are capable of tracking objects even when they temporarily disappear from sight, a cognitive capacity commonly referred to as visual working memory (VWM). The neural basis of VWM has been the subject of significant scientific debate, with recent work focusing on the relative roles of posterior visual areas, such as the inferior temporal cortex (ITC), and the prefrontal cortex. Here we reexamined the contribution of ITC to VWM by recording from highly selective individual ITC neurons as monkeys engaged in multiple versions of an occlusion-based memory task. As expected, we found strong evidence for a role of ITC in stimulus encoding. We also found that almost half of these selective cells showed stimulus-selective delay period modulation, with a small but significant fraction exhibiting differential responses even in the presence of simultaneously visible interfering information. When we combined the informational content of multiple neurons, we found that the accuracy with which we could decode memory content increased drastically. The memory epoch analyses suggest that behaviorally relevant visual memories were reinstated in ITC. Furthermore, we observed a population-wide enhancement of neuronal response to a match stimulus compared with the same stimulus presented as a nonmatch. The single-cell enhancement preceded any match effects identified in the local field potential, leading us to speculate that enhancement is the result of neural processing local to ITC. Moreover, match enhancement was only later followed by the more commonly observed match suppression. Altogether, the data support the hypothesis that, when a stimulus is held in memory, ITC neurons are actively biased in favor of task-relevant visual representations and that this bias can immediately impact subsequent recognition events.

Introduction

Visual working memory (VWM) is the cognitive capacity to temporarily represent, manipulate, and monitor visual information that is no longer readily available to the peripheral sense organs (Baddeley, 1996). The role of VWM ranges from aiding the construction of temporally contiguous perception to planning and guiding future actions. A vast network of cortical areas underlies VWM (Fuster and Alexander, 1971; Goldman-Rakic, 1987; Pessoa et al., 2002; Todd and Marois, 2004; Pasternak and Greenlee, 2005; Xu and Chun, 2006), with recent work stressing the importance of the prefrontal cortex (PFC) in reactivating the appropriate neural representations in posterior sensory areas (Ranganath and D'Esposito, 2005). In this framework, the contents of VWM are represented by neural activity in unimodal visual cortex, whereas what allows these activity patterns to persist throughout the memory interval is a top-down maintenance signal from PFC. This view has lead to a resurgence of studies reconsidering the function of sustained activity in high-level visual areas (Mikami, 1995; Nakamura and Kubota, 1995; Petrides, 2000; Druzgal and D'Esposito, 2003; Ranganath et al., 2004; Lee et al., 2005; Fiebach et al., 2006; Zaksas and Pasternak, 2006; Lewis-Peacock and Postle, 2008). For visual object WM specifically, the hypothesis predicts that neural activity within the inferior temporal cortex (ITC), a visual association area critical for the perception and recognition of visual objects (for review, see Logothetis and Sheinberg, 1996), temporarily represents complex object information that is no longer visible.

Indeed, early lesion studies had suggested that ITC contributes to the retention of visual information (Mishkin, 1982) and single-cell recordings in ITC in subjects performing delayed response tasks showed visually selective delay period activity (Fuster and Jervey, 1982; Miyashita and Chang, 1988). Despite these initial reports, additional work revealed the selective delay period activity in ITC to be susceptible to the presence of intervening sensory information (Miller et al., 1993); concomitantly, it was shown that PFC neurons were resistant to such interference (Miller et al., 1996). Nonetheless, the acknowledged involvement of ITC in VWM extends beyond mere encoding and into the memory epoch as it has been shown recently that a population of ITC neurons does regain category selectivity toward the latter stages of the delay interval (Meyers et al., 2008). In addition, ITC is involved in the matching phase of VWM because some ITC neurons respond more strongly to a stimulus if it matches the one actively held in mind or more weakly if a stimulus simply repeats itself, effects commonly referred to as match enhancement and match suppression, respectively (Baylis and Rolls, 1987; Eskandar et al., 1992; Miller et al., 1993; Miller and Desimone, 1994). In fact, match enhancement is hypothesized to reflect the biasing of visual activity in ITC by PFC feedback (Miller and Desimone, 1994; Miller et al., 1996). The precise cortical origins and circuitry underlying match effects, however, remain elusive.

Received Dec. 4, 2008; revised March 16, 2009; accepted March 18, 2009.

This work was supported by National Institutes of Health Grant RO1-EY014681 (D.L.S.), the James S. McDonnell Foundation (D.L.S.), and Interdisciplinary Vision Training Program Grant 5T32EY018080-02 (L.W.).

Correspondence should be addressed to David L. Sheinberg, Box G-LN, Brown University, Sidney E. Frank Hall, 185 Meeting Street, Providence, RI 02912. E-mail: david_sheinberg@brown.edu.

DOI:10.1523/JNEUROSCI.5785-08.2009

 0270-6474/09/295494-14$15.00/0

In this study, we reinvestigated the activity of ITC neurons during VWM by recording from highly selective single neurons in ITC as monkeys engaged in a more physically realistic same/different VWM task. Using an occlusion-based VWM paradigm, we could also explore neural processes associated with representing the implied presence of an object, which is believed to depend on the context in which it disappears from sight (Baker et al., 2001; Kaufman et al., 2003, 2005). Our careful examination of the temporal evolution of the single-cell responses, individually and jointly as a population, during the various phases of VMW leads us to conclude that ITC is an integral component of the working memory system for visual objects. Importantly, by comparing single-cell activity with local field potential (LFP) dynamics, we show that match enhancement is likely a consequence of neural processing occurring within ITC, supporting the view that it plays a central role not only in object perception but also in the matching component of VWM. Although our results clearly support the view that ITC activity depends on both the physical visual world and memory demands for objects within that world, data from our task and our nonhuman subjects provide no clear evidence that manipulations thought to affect perceived object permanence (e.g., mode of disappearance) have a direct impact on neural activity in this cortical region.

Materials and Methods

Subjects and surgery. Two male adult macaque monkeys (*Macaca mulatta*; monkey M and monkey R), weighing between 9.0 and 11.0 kg, participated in the experiment. The monkeys were first familiarized with the behavioral setup and then implanted with titanium head posts for head restraint. To gain access to ITC, we placed a custom chamber either over the left hemisphere (monkey M) or the right hemisphere (monkey R) (Horsley-Clark coordinates, +15 anterior, +20 lateral). All surgeries were performed under isoflurane anesthesia, in accordance with the guidelines published in the National Institutes of Health *Guide for the Care and Use of Laboratory Animals* and approved by the Brown University Institutional Animal Care and Use Committee.

Recording methodology. For monkey M, each recording session began with the insertion of a guide tube (25 gauge) just beyond the surface of the dura (~5 mm). Monkey R had a permanent, but adjustable, guide tube ending ~15 mm below the dura. Both approaches allowed us to explore an area approximately ±5 mm in radius in the *x–y* plane of the brain at the depth of our recordings. Using a hydraulic micropositioner (David Kopf Instruments), a single tungsten microelectrode [1.0–1.5 MΩ (Alpha Omega) or 7.0 MΩ (FHC)] was then advanced through the guide tube and into the cortical tissue. The neural signal was passed through a head stage, split into two independent channels, filtered, and amplified (model 15LT Bipolar; Grass Instruments). One channel was bandpass filtered between 0.3 and 300 Hz, and the other channel was bandpass filtered between 100 Hz and 6 kHz. The low- and high-frequency channels were digitized and stored at 2.5 and 34 kHz, respectively. Single units were sorted online using a threshold and dual amplitude windows (custom software) but were sorted offline using previously described clustering software (Quiroga et al., 2004). ITC was located on the basis of the overall electrode depth, the number of gray/white matter transitions, and the shape and magnitude of the visually evoked potential, as well as the visual object selectivity of the encountered neurons. In one monkey (monkey R), recording locations were confirmed with a magnetic resonance imaging scan (supplemental Fig. S1, available at www.jneurosci.org as supplemental material).

Eye signal. Eye movements were recorded using the EyeLinkII video tracking system (SR Research) running at 500 Hz. The analog output from the eye-tracking system was sampled at 1 kHz by our acquisition system and stored to disk as a running average at 200 Hz.

Stimulus presentation and behavioral control. The stimuli were presented on a dual-processor graphics workstation, which ran a custom-written, OpenGL-based visual stimulation program under WindowsXP (Microsoft). The screen had a resolution of 1024 × 768 and a vertical refresh rate of 100 Hz. The slave graphics system was controlled via a network of interconnected computers all running the QNX real-time operating system (QNX Software Systems). This hub of computers provided deterministic control and acquisition of button responses and eye position and communicated with the graphics workstation via an isolated high-speed Ethernet and digital input/output. Experimental control and data acquisition software consisted of custom-written C programs.

Behavioral paradigm and daily recordings. The stimulus set used in this study consisted of 125 full color images of everyday objects (Hemera Technologies). Each stimulus subtended ~2.0° × 2.0°. Example stimuli are presented in Figure 1*E*. Both monkeys were familiarized with the images before recording, but neither monkey had ever been trained to associate a particular image in this set with a specific eye or hand movement. Although the monkeys had extensive experience with the full set of images, in any given recording session, the monkey worked with only a subset of seven images. This ensured that we obtained a sufficient number of repeats per stimulus per condition. Two of the seven images were tailored to the response properties of the recorded neurons such that one of them drove the isolated cell much more reliably and robustly than the other. The other five stimuli were picked completely randomly and were intended to keep the monkey engaged (one of these stimuli served as the complex occluder; see below). The characterization of single neurons was done via online inspection of rasters and spike density functions (SDFs). Although all well isolated neurons were quickly inspected, only those that showed a differential response to any arbitrary effective and ineffective pair were chosen for actual data collection. Before and after the behavioral task, we collected neural responses to the seven stimuli during a viewing-only task (for examples, see Fig. 2*A*). In this task, images were centrally presented on a uniform gray background (luminance, 5 cd/m^2). Each trial consisted of five stimuli with a 200 ms presentation time and a 200 ms interstimulus time, and the monkey had to maintain fixation throughout the trial. Each of the seven stimuli was repeated 10 times, resulting in 14 trials per block (1 preblock + 1 postblock = 28 trials total).

The monkeys performed four variants of a same/different task. All trials started only after the monkey had acquired a central fixation spot for 300–450 ms. In the "disappear" condition, the monkey encoded a sample image for 1000 ms, retained its identity through a 1000 ms blank memory interval, and then compared the choice stimulus with the encoded sample (Fig. 1*A*). The monkey's task was to indicate with either a right or left button press whether the comparison stimulus was the same or different as the sample stimulus. In the "invisible occluder" condition, a square of the same color as the background was presented above the sample image (Fig. 1*B*). This invisible square was located 4° above the center of the display. During the ensuing 1000 ms sample presentation time, the invisible square gradually occluded the sample, reaching full occlusion at the onset of the delay interval; thus, the invisible square translated at a rate of 4°/s. At the end of the delay, the comparison stimulus was presented. The "square occluder" condition was very similar to the invisible occluder condition except that the occluder was black and therefore visible (Fig. 1*C*). Furthermore, the black square remained visible throughout the delay, giving an impression of object permanence. At the end of the delay, the comparison stimulus instantaneously replaced the black square. The "complex occluder" condition was identical to the square occluder condition except the black square occluder had a complex visual image embedded inside (Fig. 1*D*). This condition had the dual role of implying object permanence as well as introducing the presence of intervening sensory information during the delay period.

During the sample, delay, and comparison epochs (until button response) of all conditions, the monkey was required to maintain its eye position within a bounding box subtending 2.5° × 2.5° of visual angle. This ensured that the monkey was looking at the black square occluder and the complex occluder during the delay period as well as at the comparison stimulus during the choice period. To guarantee a robust dataset, the two images that differentially drove the cell (effective/ineffective) were repeated slightly more frequently than the other randomly chosen images. In particular, each block of trials contained 96 trials split evenly

between the four conditions (24 trials/condition). Of these 24 trials, six were reserved for the effective stimulus, six were reserved for the ineffective stimulus, and three each were reserved for the four random stimuli, with one of the seven stimuli serving solely as the complex occluder; thus, the monkeys could actively ignore the complex occluder because the image embedded in it was never the sample or the choice. Of the 96 trials in a block, 48 were of the "same" type and 48 were of the "different" type. Only cells that could be held for at least three full blocks of behavioral trials plus two full blocks of viewing-only trials were analyzed further.

Each recording session refers to such a set of blocks. Because the monkeys performed the trials relatively quickly, on many occasions we were able to gather a full dataset for one set of seven images, move a few hundred micrometers (>200 μm), isolate another cell, and gather another dataset. Therefore, it was possible to have multiple "recording sessions" per day. In all, we had 69 (monkey M, 34; monkey R, 35) recording sessions during which we collected data from 134 cells. There were more cells than recording sessions because, on most days, we had multiple units on a single electrode, although in most situations only one of the cells on an electrode was analyzed. This is reflected in the fact that only 79 of the 134 cells met additional criteria and were analyzed (see below). Figure 1*F* shows the average performance of the two monkeys collapsed across all recording sessions. As can be seen, behavior was excellent (>90%) in the three simpler conditions and slightly lower but still well above chance in the complex occluder condition.

Data analysis. All data analysis was performed with custom-written scripts, the IMSLS C library (Visual Numerics), the R statistical computing environment (www.r-project.org) and Matlab (MathWorks). To analyze spike data, we first convolved the spike trains with a Gaussian kernel (resolution, 1 ms; $\sigma = 10$ ms). To compute firing rates, we integrated the resultant SDFs, divided by the appropriate time bin, and averaged over trials. To quantify the stability of our recorded cells, we computed Pearson's *r* coefficient between the responses to the seven images in the first and second viewing-only blocks (responses were integrated from 75 to 375 ms after stimulus onset). To quantify the selectivity of our recorded sample of cells, we computed the stimulus selectivity index (SSI) based on the firing rates of the cells from 75 to 375 ms after stimulus onset during the viewing-only task. The SSI has the formula $(R_{\text{eff}} - R_{\text{ineff}})/(R_{\text{eff}} + R_{\text{ineff}})$, where R_{eff} and R_{ineff} are the firing rates elicited by the effective and ineffective stimuli, respectively, and ranges between 1 and −1. A value of 1 signifies a total lack of response to the ineffective stimulus, whereas −1 indicates no response to the effective stimulus.

To quantify how reliably an individual neural response could discriminate between the effective and ineffective stimulus, we constructed receiver operating characteristic (ROC) curves from the neural firing rates (Green and Swets, 1966). For our purposes, it sufficed to summarize the whole curve by computing the area under the curve (AUC). The value of the AUC ranges between 0 and 1, with 0.5 equal to chance performance and values nearer 0 and 1 indicating more pronounced reliability in discriminating the two distributions of interest. Whether the AUC is above or below 0.5 depends on which distribution is given the "noise" label and which is given the "signal" label. The effective stimulus was designated the signal.

Depending on the specific analysis, the AUCs incorporated neural data segments of varying lengths. The following list summarizes the primary analysis epochs. (1) To compute the AUCs during the viewing-only task, firing rates obtained from 75 to 375 ms after stimulus onset were used. (2) For the sliding window analyses, we used step sizes equal to 10 ms and window sizes equal to 100 ms. In this case, the plotted AUC is centered on the window of interest (e.g., the value at 0 ms is the AUC calculated from the firing rates between −50 and 50 ms). (3) Delay period AUCs (see Fig. 4*C*) were based on the firing rates from the last 750 ms of the delay period. In cases in which different parts of the delay period were explored, the 750 ms epoch was divided into three equal and nonoverlapping segments (early, middle, and late). (4) In instances in which integration time was of interest, the ROC analysis included firing rates computed from longer

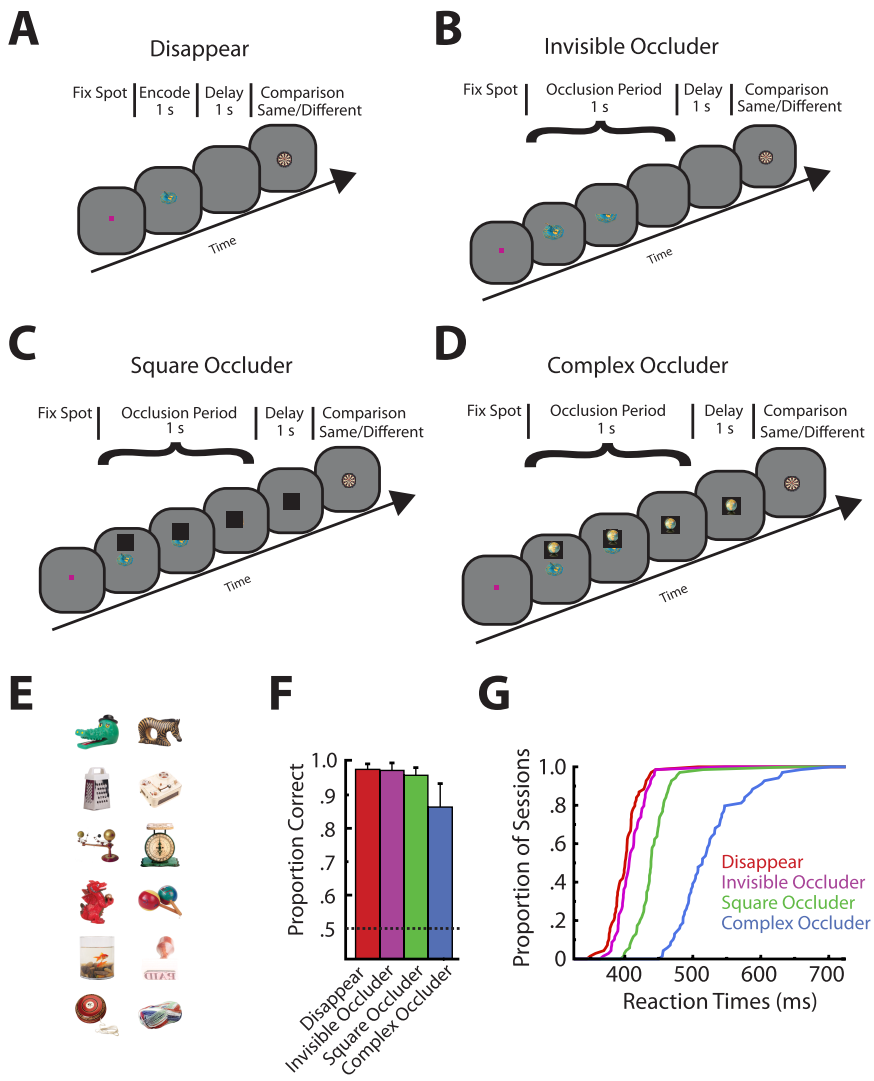

Figure 1. Task, stimuli, and behavioral performance. ***A***, Disappear condition. A sample was presented for 1 s, followed by a 1 s blank delay period, at the end of which a choice stimulus appeared. The monkey's task was to indicate with either a right or left button press whether the choice matched the sample. ***B***, Invisible occluder condition. Same as in ***A***, except that, during the 1 s sample presentation time, a square of the same color as the background gradually occluded the sample, reaching full occlusion at the onset of the delay interval. ***C***, Square occluder condition. Same as in ***B***, but the occluder was now a black square. ***D***, Complex occluder condition. Same as in ***C***, but there was a complex visual stimulus embedded inside the black square. All conditions (***A–D***) required the monkey to first fixate a small spot in the center of the display before trial onset. ***E***, Example stimuli. Twelve of the 125 stimuli used throughout the recording sessions. ***F***, Performance of the two monkeys averaged across all recording sessions and sorted by condition. Error bars indicate SD of performance across the 69 sessions. Notice the slightly worse and more variable performance in the complex occluder condition. ***G***, Cumulative distribution of the reaction times of all recording sessions sorted by condition. The more difficult conditions (square and complex occluder) led to longer reaction times ($F_{(3,272)} = 180.58$; $p \ll 0.0001$).

and longer time windows, beginning 25 ms before visual latency and ending at 125 ms after visual latency (see Fig. 7*D*,*E*). The windows were −25 to −20, −25 to −15,…, −25 to 120, −25 to 125 ms.

The AUC computed with the trapezoidal rule is related to the Mann–Whitney *U* test statistic via the following relationship: $AUC = U/(n_1 * n_2)$, where n_1 and n_2 are the sample sizes of the two distributions (Hanley and McNeil, 1982). Thus, the two measures are equivalent. This relationship allowed us to circumvent the need to do permutation tests to compute the significance values of individual AUCs; instead, we simply used the *p* values obtained from the Mann–Whitney *U* test (see Fig. 4*C*).

To quantify how well a population of neurons could discriminate between the effective and ineffective stimulus, we used a linear classifier known as the support vector machine (SVM) (Cristianini and Shawe-Taylor, 2000). If we assume that a dataset is linearly separable, then the SVM classifier can be defined as $H(z) = \text{sign}(\langle w, z\rangle + b)$, where the pair (w, b) defines a hyperplane that separates points belonging to one class from points belonging to the other class, and z is a feature vector in the original input space; in the case of neuronal population data, z is the set of firing rates across multiple neurons. The defining attribute of an SVM is that the pair (w, b) is chosen to maximize the minimum margin between the two classes. As this particular optimization problem is underconstrained, to make it computationally tractable one introduces the appropriate set of constraints, namely $C_k(\langle w,Z_k\rangle + b) \geq 1$ for $1 \leq k \leq N$, where C_k is the true class of the point being classified ($C_k \in \{-1, 1\}$). A given real-world dataset is not likely to be linearly separable; in this case, one can either perform a nonlinear transformation on the original input space and/or introduce slack variables s_k, where $C_k(\langle w,Z_k\rangle + b) \geq 1 - s_k$, $s_k \geq 0$, and λ is the cost parameter governing the penalty associated with each s_k. Throughout the text, we used a $\lambda = 1$, but performance was relatively unaffected through a rather wide range of λ. Because we sought to make the classifier more biologically plausible, we refrained from using nonlinear transformations and instead used a linear kernel only. All fitting was performed in R with the svm function found in the e1071 package (Dimitriadou et al., 2008). Data were scaled internally to have zero mean and unit variance.

The procedure used to train and test the classifier was very similar to that outlined by Meyers et al. (2008). To obtain an adequate estimate of generalization performance, we restricted this analysis to cells that had at least 12 correct effective and 12 correct ineffective trials in the condition of interest, leaving 79, 79, 79, and 76 cells for the disappear, invisible occluder, square occluder, and complex occluder conditions, respectively. For each time window in the sliding window analysis (see Fig. 5*A*) (step, 10 ms; window, 100 ms), we randomly selected 64 cells and then from these cells randomly sampled 12 effective and 12 ineffective trials. We trained the classifier on 16 trials (eight effective and eight ineffective) and tested on the remaining eight trials (four effective and four ineffective). This step was repeated three times per random population of cells, with a different subset of trials serving as training (and testing) sets. We then averaged performance across the three cross-validation runs. To obtain a bootstrap-like estimate of the performance mean and SD at each time point, we repeated the entire procedure 100 times, selecting a new random population of cells and set of trials on every iteration. The error bars shown in the plots in Figure 5 and reported throughout the text are the SDs obtained across the 100 bootstrap-like iterations.

For the population delay period analysis (see Fig. 5*B*), we averaged spike counts over the entire last 750 ms of the delay. We performed the same training and testing procedures on these spike counts as outlined above, except we systematically varied the number of cells used to train and test the classifier's performance ($n = 1, 2, 4, 8, 16, 32,$ or 64). The data were then fitted with a function of the form proportion correct $\sim 1 - 0.5 * e^{-(n/\alpha)^{\beta}}$ (Quick, 1974), where n is the number of cells, α is the number of cells needed to reach 82% correct, and β is the slope parameter. This equation is derived from the cumulative distribution function of the Weibull distribution. Fitting was performed with the nls (nonlinear least squares) function in R using a Gauss-Newton algorithm.

To explore the relationship between visual responses and reaction times, we defined visual latency to be the first of 25 consecutive milliseconds to exceed 25% of the maximum visual response, with only responses to the effective stimulus included in the analysis. [The same visual latencies were used in the integration analyses (Fig. 7*D*,*E*)]. To obtain a similar latency measure for visual selectivity (match suppression), the same analysis was repeated on the difference SDFs obtained by subtracting the ineffective (match) SDF from the effective (nonmatch) SDF. The latencies in this case could not occur before visual response latency. Before all the latency computations, the responses were zeroed by subtracting the mean firing rate in the last 100 ms of the delay (1900–2000 ms absolute time). All latencies were computed separately for each condition, yielding four latencies per cell. If the condition latency of a particular cell could not be found as outlined above (approximately two cells), its latency was defined as the average of the other cells. To normalize latencies across days, we subtracted from the condition latency of each cell the average of its four conditions.

To obtain z-normalized reaction times, we first subtracted the mean and then divided raw reaction times by the SD observed on that day; we then averaged reaction times for each condition and for each day separately. Because on some days a single instance of reaction times corresponded to more than one cell (more than one cell/electrode), some reaction times were replicated so that lengths of the two vectors matched. To quantify correlation, we used Pearson's correlation coefficient. For Figure 6, *B* and *D*, reaction times were restricted to trials that had the effective or ineffective stimulus as the comparison stimulus.

To investigate match effects, we only included cells that had at least five match and five nonmatch trials. For the effective stimulus, this reduced our cell counts to 73, 77, 73, and 69 for the disappear, invisible occluder, square occluder, and complex occluder conditions, respectively. The ineffective stimulus yielded 76, 78, 78, and 74 cells, respectively. Although we observed quite robust visual selectivity in the LFP (supplemental Fig. S3, available at www.jneurosci.org as supplemental material), the LFP match effects were not stimulus dependent (supplemental Fig. S5, available at www.jneurosci.org as supplemental material). As such, the LFP analyses shown in Figure 8 were collapsed across all stimuli.

The statistical test used to determine when a particular effect became significant was based on the nonparametric test developed by Blair and Karniski (Blair and Karniski, 1993; Appelbaum et al., 2006). Briefly, if we have an $n \times m$ matrix where each of the m columns corresponds to a difference time series from a single cell or site, we derive a vector of t scores from this matrix by computing the $n \times 1$ mean difference waveform and dividing by the SEM. Now, the null hypothesis states that there is no distinction between conditions A and B; therefore, the difference SDF of each cell or the difference LFP of each site can be sign inverted with no effect. To get a true reference distribution for one's statistic of choice, one could in theory compute every single combination of real and inverted differences, but in this case that would be 2^{79} (or 2^{69}) combinations. To get an estimated reference distribution, for each of 10,000 random permutations, we computed a t score vector and extracted the maximum absolute t value. To achieve an α level of 0.05 for a two-sided test, the critical quantities become the t values that cut off the top and bottom 2.5% of the reference distribution. By comparing the actual vector of t scores to these critical values, we determined when the differences achieved significance. We verified the lack of spurious results by repeating the same analysis on the fixation period, defined as −310 to 0 ms relative to trial onset (supplemental Fig. S6, available at www.jneurosci.org as supplemental material). SDFs were normalized before the analysis according to the formula $(R - R_{\text{baseline}})/(R_{\text{effmax}} - R_{\text{baseline}})$, where R is the original response, R_{effmax} is the maximum response to the effective image obtained from the average SDF in the viewing-only blocks, and R_{baseline} is the average firing rate in the 200 ms before trial onset.

Results

Task and behavior

To investigate the contribution of ITC neural activity to VWM, we trained two monkeys to perform four variants of a same/different task. In all conditions, the monkeys were required to encode a sample stimulus, maintain its identity through a 1-s-long interval, and then report with either a right or left button press whether the comparison stimulus matched or differed from the sample (Fig. 1*A–D*). Importantly, with the aid of gradual

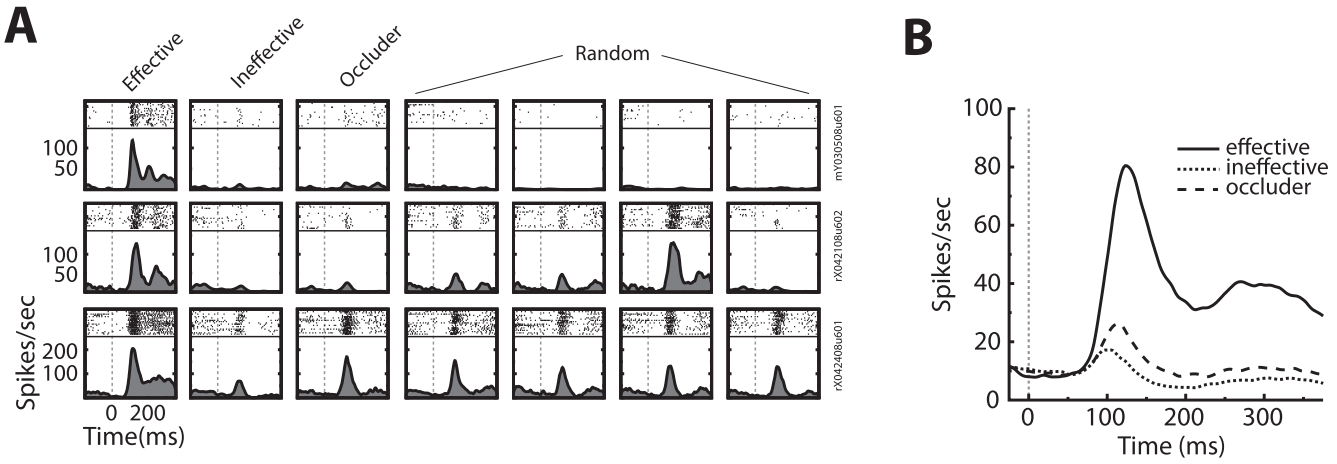

Figure 2. Selectivity of cells during a viewing-only task. ***A***, Three typical cells. Each row of plots depicts the responses (rasters and spike density functions) of one cell to the presentation of the effective, ineffective, and complex occluder stimuli as well as to four other randomly chosen objects. Notice the robust selectivity as pertains to the effective and ineffective pair. ***B***, Population averaged spike density functions sorted by effective, ineffective, and complex occluder labels. The plot includes data only from cells that had SSIs > 0.25, AUCs > 0.75, and Pearson's $r > 0.75$ (for additional details, see Materials and Methods). Note the large and extended differences between the effective and ineffective response time courses. Also, although the complex occluder was chosen randomly, it is evident that it rarely elicited a response much bigger than the ineffective stimulus.

occlusion, two of the manipulations, the square and complex occluder conditions (Fig. 1*C*,*D*), allowed us to imply the physical presence of the object during the memory period (Baker et al., 2001). Furthermore, the complex occluder condition (Fig. 1*D*) made it possible to examine the impact of intervening sensory information on stimulus-selective delay period responses. As is clear from Figure 1*F*, both monkeys performed well in all four conditions (>80%), with the complex occluder condition proving to be most difficult. The square and complex occluder conditions also led to slower reaction times, again highlighting the differences in difficulty among the conditions (Fig. 1*G*) ($F_{(3,272)} = 180.58$; $p \ll 0.0001$). Although we sought to establish a relationship between LFP oscillations and object permanence (Tallon-Baudry et al., 1998; Kaufman et al., 2003, 2005), defined as the cognitive ability to understand the continued physical presence of an object that occurs when objects do not simply vanish but instead disappear in a manner consistent with their continued existence (e.g., gradual occlusion), we found no systematic differences related to our permanence manipulations (see supplemental Results and supplemental Fig. S4, available at www.jneurosci.org as supplemental material).

Stimulus selectivity during passive fixation

We recorded from a total of 134 neurons (monkey M, 57 neurons; monkey R, 77 neurons). Because previous reports have suggested that highly selective ITC neurons may be preferentially involved in VWM (Mikami, 1995), we specifically sought neurons whose responses could reliably discriminate an arbitrarily chosen pair of effective and ineffective stimuli. Figure 2*A* shows the rasters and SDFs obtained during a viewing-only task for three such neurons. To quantify selectivity and reliability of the differential responses, we used the SSI and the area under an ROC curve (see Materials and Methods). Only neurons with SSI > 0.25 (SSI mean of 0.70), AUC > 0.75 (AUC mean of 0.98), and firing rates that remained consistent from the beginning to the end of the recording session (Pearson's $r > 0.75$; Pearson's r mean of 0.94) (see Materials and Methods) were included in additional analyses. This restricted our sample to 79 neurons (monkey M, 35 neurons; monkey R, 44 neurons). We feel confident that this subset of neurons contained some of the most differentiable responses (as pertains to the arbitrary effective/ineffective pair) to be found in ITC (Fig. 2*B*). All subsequent analyses concentrate on this selective pool of neurons.

Stimulus selectivity during encoding and delay periods of VWM

Single-cell analyses

The goal of the present study was to characterize the activity patterns of individual ITC neurons during the various phases of VWM. Figure 3, *A* and *B*, shows the firing rate modulations of two sample neurons during the encoding and memory periods of four variants of a same/different task. Both neurons exhibited pronounced bursts of activity to the presentation of the effective stimulus and much reduced increases to the ineffective stimulus. Sliding window ROC analyses (step, 10 ms; window, 100 ms) revealed that the AUCs of both neurons approached values of 1.0 during the encoding periods of all four conditions (Fig. 3*C*,*D*), indicating highly separable distributions of encoding-related neural responses.

During the delay itself, the two neurons showed strikingly different behaviors. A comparison of the firing rates in the last 500 ms of the delay period for the first cell (Fig. 3*A*) shows clear evidence for persistent stimulus selectivity in all four conditions. As a result, this particular neuron had delay period AUCs that were well above chance; crucially, this was the case even in the complex occluder condition, in which there was another complex image in view of the monkey (Fig. 3*C*, blue line). In the second example (Fig. 3*B*), the same comparison shows that, whereas the AUCs for the disappear and invisible occluder conditions do rebound toward the latter stages of the delay (Fig. 3*D*, red and magenta lines), those for the square and complex occluder conditions remained at chance throughout the whole memory interval (Fig. 3*D*, green and blue lines).

To examine the AUC time courses for all individual neurons, we constructed AUC color plots. In these plots (Fig. 4*A*), which have been split by condition, the *x*-axis specifies time, hue indicates AUC, and each row represents the AUC time course of one cell. For each condition separately, we have arranged the cells from top to bottom in descending order of delay period AUCs (see below). It appears that all conditions, including the complex

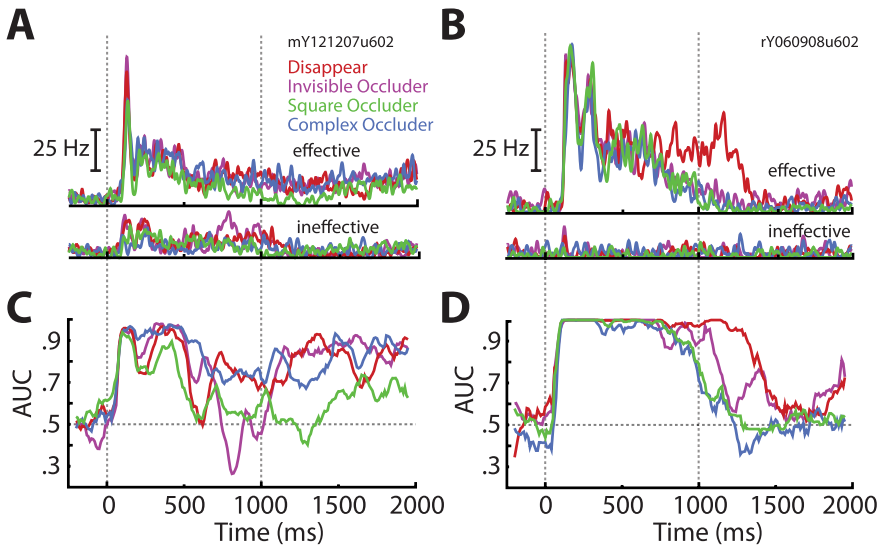

Figure 3. Encoding and delay period activities of two example neurons. ***A***, ***B***, Spike density functions of two example neurons sorted by image effectiveness and condition. Observe the robust response differences during the encoding epoch of all conditions in both ***A*** and ***B***. The differences become much smaller during the delay period, vanishing altogether in the square and complex occluder conditions in ***B***. ***C***, ***D***, Sliding window ROC analyses (step, 10 ms; window, 100 ms) for the same two neurons. Each AUC line compares the responses to the effective versus ineffective stimulus for one condition. Note that the neuron in ***C*** sustained its stimulus-selective response throughout the delay periods of all conditions. The neuron shown in ***D***, however, exhibited no stimulus selectivity during the delay periods of the square and complex occluder conditions.

occluder condition, contained a collection of neurons that exhibited stimulus-selective activity during the memory epoch. This is most evident by inspecting the top portions of the panels, in which individual cells have reddish AUCs in various parts of the delay (right of the second dashed line).

Figure 4*B* shows the average AUCs for the entire sample of neurons. The population average shows that there is an obvious loss of selectivity after the disappearance of the stimulus, but that the three simpler conditions partially regain their selectivity in the latter stages of the delay. [Because the disappear condition allowed the sample stimulus to be fully visible for the whole duration of the encoding epoch (Fig. 1*A*), its AUCs took longer to decay.] To quantify the recovery of stimulus selectivity, we divided the last 750 ms of the delay into three nonoverlapping segments of 250 ms each (early, middle, and late delay) and computed the AUC for each cell and for each segment separately and averaged across cells. For the disappear condition, we then compared the middle with late periods and for the three occluder conditions the early to late periods; indeed, the comparisons revealed significant increases for the three simpler conditions (data not shown; paired *t* tests; disappear, $\Delta AUC = 0.05$, $p < 0.0001$; invisible occluder, $\Delta AUC = 0.06$, $p < 0.0003$; square occluder, $\Delta AUC = 0.08$, $p < 0.0001$; complex occluder, $\Delta AUC = 0.00$, $p = 0.94$). This suggests that stimulus-selective delay activity in ITC is not the result of information passively spilling over from the encoding epoch but rather the neural correlate of the contents of the working memory store (Miyashita and Chang, 1988; Meyers et al., 2008).

By averaging the AUCs across cells (Fig. 4*B*), we implicitly assumed that all neurons represent the stimulus in the same short 100 ms intervals throughout the memory interval. This, however, need not be the case. For example, one neuron could represent the stimulus from 1500 to 1600 ms, whereas another neuron could represent the stimulus from 1600 to 1700 ms. Averaging across cells could obscure this information, underestimating the contribution of ITC to memory maintenance. To avoid this problem and get a more reliable estimate of delay period selectivity, we recomputed delay period AUCs based on the activity over the entire last 750 ms of the delay. Population histograms of these delay period AUCs, again sorted by condition, are shown in Figure 4*C*. The means of the distributions confirmed that the three simpler conditions exhibited population-wide delay period selectivity, whereas the complex occluder condition itself did not (Fig. 4*C*) (two-tailed *t* tests; disappear, 0.65, $p < 0.0001$; invisible occluder, 0.64, $p < 0.0001$; square occluder, 0.59, $p < 0.0001$; complex occluder, 0.52, $p = 0.24$).

Interestingly, 41.8% (33 of 79) and 45.6% (36 of 79) of cells during the disappear and invisible occluder conditions, respectively, had significantly selective responses to the effective stimulus ($r_{effective} > r_{ineffective}$) during the delay period (Fig. 4*C*, black bars to the right of 0.50). These proportions were higher than previously reported for ITC (Fuster and Jervey, 1982; Eskandar et al., 1992; Mikami, 1995; Miller et al., 1996) (but see Miyashita and Chang, 1988), likely reflecting the selective nature of our neurons. In addition, 12.7% (10 of 79) of cells showed significantly selective responses to the effective stimulus (again, $r_{effective} > r_{ineffective}$) during the memory epoch of the complex occluder condition. Despite our relatively small sample of cells, this percentage was greater than expected by chance, assuming one erroneously rejects the null hypothesis in 5% of cases (binomial test; $p = 0.006$). Furthermore, repeating the delay–rise analysis on this subset of selective cells yielded a robust AUC increase from the early to late delay periods (Wilcoxon's signed ranks test, $\Delta AUC = 0.09$; $p = 0.0098$) (see supplemental Fig. S2, available at www.jneurosci.org as supplemental material), again suggesting that stimulus selective delay period activity is functionally distinct from passive decay.

Population analyses

The ROC analyses reported above treated the contribution of each cell to coding stimulus identity individually. To investigate how well our selective population of cells could signal stimulus identity throughout the trial if their informational content were combined, we trained and tested the ability of a linear SVM to decode the effective and ineffective stimuli based on the joint activity of 64 randomly selected cells (see Materials and Methods). Crucially, the test data were always separate from the train data. The results of this sliding window population analysis (step, 10 ms; window, 100 ms) are shown in Figure 5*A*, split by condition. Similar to the ROC analyses, the population approach was essentially perfect at decoding stimulus identity throughout the duration of the encoding epoch, regardless of the condition. Furthermore, for the disappear and invisible occluder conditions, there was a significant amount of stimulus information throughout the majority of the delay. By comparing it with the average single-cell AUC values (dotted lines), it is evident, and not surprising, that the population of neurons considerably outperformed single cells. Importantly, we see a trend for the performance of the classifier to rise toward the latter stages of the delay.

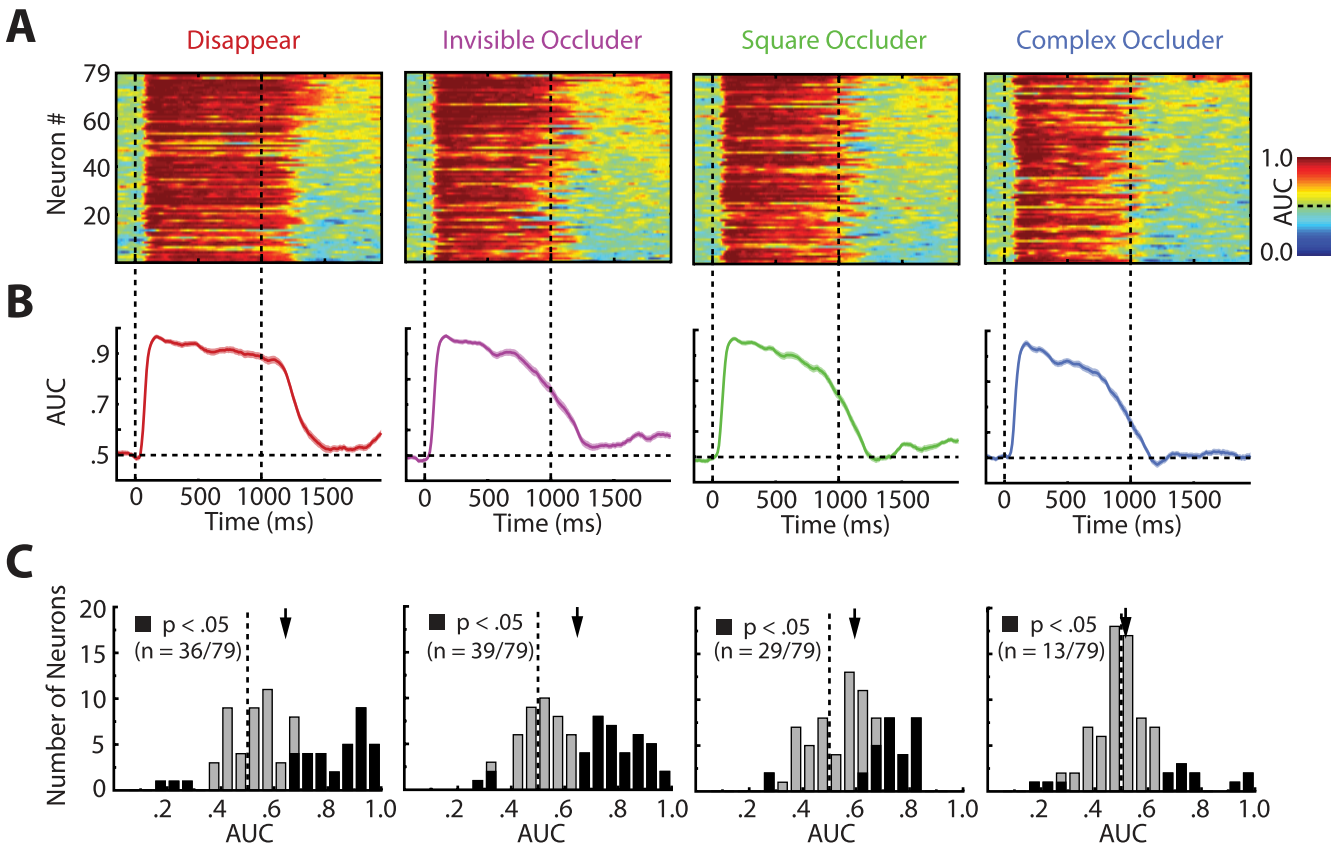

Figure 4. Analysis of encoding and delay period AUCs. ***A***, AUCs of individual neurons sorted by condition. The plots show the AUCs for all neurons (ordinate) as a function of time (abscissa) and condition (panel). AUC values are color coded, with hotter colors illustrating values closer to 1.0. For every condition independently, we have sorted the neurons in descending order of delay period AUCs (as computed in ***C***). Note, by inspecting the top parts of the panels, that every condition had a subpopulation of neurons that exhibited stimulus-selective responses during the delay. ***B***, Population averaged AUCs. Each curve depicts the sliding window AUCs averaged across the entire population of neurons for one condition. Shaded regions indicate ±1 SEM. The two vertical dotted lines in ***A*** and ***B*** indicate sample and delay onset, respectively. ***C***, Population histograms of delay period AUCs. The AUC values were computed from the entire last 750 ms of the delay period. The black bars indicate individual cells that were significant at the $\alpha = 0.05$ level (Mann–Whitney *U* test). The arrows denote the population means of the distributions. Notice that the population means for the three simpler conditions were all significantly above 0.50 ($p < 0.0001$ in all cases). Although the complex occluder condition did not have a population mean significantly different from 0.50, 10 individual cells (of 79) did have significantly increased responses to the effective stimulus, more than expected by chance (binomial test, $p < 0.006$).

Although the square and complex occluder conditions had noticeably lower performance during the delay compared with the other conditions, we again see better performance for the population than for the single cell; moreover, the performance of the classifier exhibited the same climbing pattern as the delay epoch neared its end.

The sliding window SVM analyses are subject to the same considerations as outlined for the sliding window ROC analyses; namely, we assumed that all cells signal stimulus identity in the same short epochs throughout the memory interval. To remove this assumption, we repeated the SVM analysis but, just as for the analogous ROC analysis, integrated spike counts over the entire last 750 ms of the delay (1250–2000 ms). To examine how well populations of various sizes could do, this analysis was run separately for population sizes of 1, 2, 4, 8, 16, 32, and 64. The results are shown in Figure 5*B*. We subsequently fit these points with a function of the form proportion correct $\sim 1 - 0.5 * e^{-(n/\alpha)^{\beta}}$, where n is the number of cells, α is the number of cells required to reach 82% correct, and β is the slope parameter (see Materials and Methods).

Three remarks can be made with regard to the points and their corresponding fits. First, performance with 64 cells was well above chance in all four conditions (disappear, 0.995 ± 0.013; invisible occluder, 0.994 ± 0.016; square occluder, 0.914 ± 0.065; complex occluder, 0.803 ± 0.088). Thus, combining the activities of multiple cells leads to, on average, much better decoding performance than reading out the activity of only one cell. Second, the thresholds of the fitted functions were smaller for the disappear, invisible occluder, and square occluder conditions (disappear, $\alpha = 8.41$ cells; invisible occluder, $\alpha = 9.24$ cells; square occluder, $\alpha = 26.69$ cells) than for the complex occluder condition (complex occluder, $\alpha = 71.25$ cells). The slopes of the Weibull fits exhibited the converse pattern, namely the complex occluder condition had the shallowest slope (disappear, $\beta = 0.74$; invisible occluder, $\beta = 0.68$; square occluder, $\beta = 0.66$; complex occluder, $\beta = 0.64$). This particular combination of parameters for the complex occluder condition suggests that more cells would be needed to achieve the same level of performance as in the simpler conditions. This is consistent with the notion that, when there is no intervening stimulus or the stimulus is not optimal for ITC, such as the black square, the contents of working memory can be more easily decoded from a population of ITC cells. Third, extrapolating the fits to $n = 100$ yielded performance values of 1.00, 1.00, 0.95, and 0.86 for the disappear, invisible occluder, square occluder, and complex occluder conditions, respectively. These values were very similar to the average performance of the two monkeys (0.98, 0.97, 0.96, and 0.86 for the disappear, invisible occluder, square occluder, and complex oc-

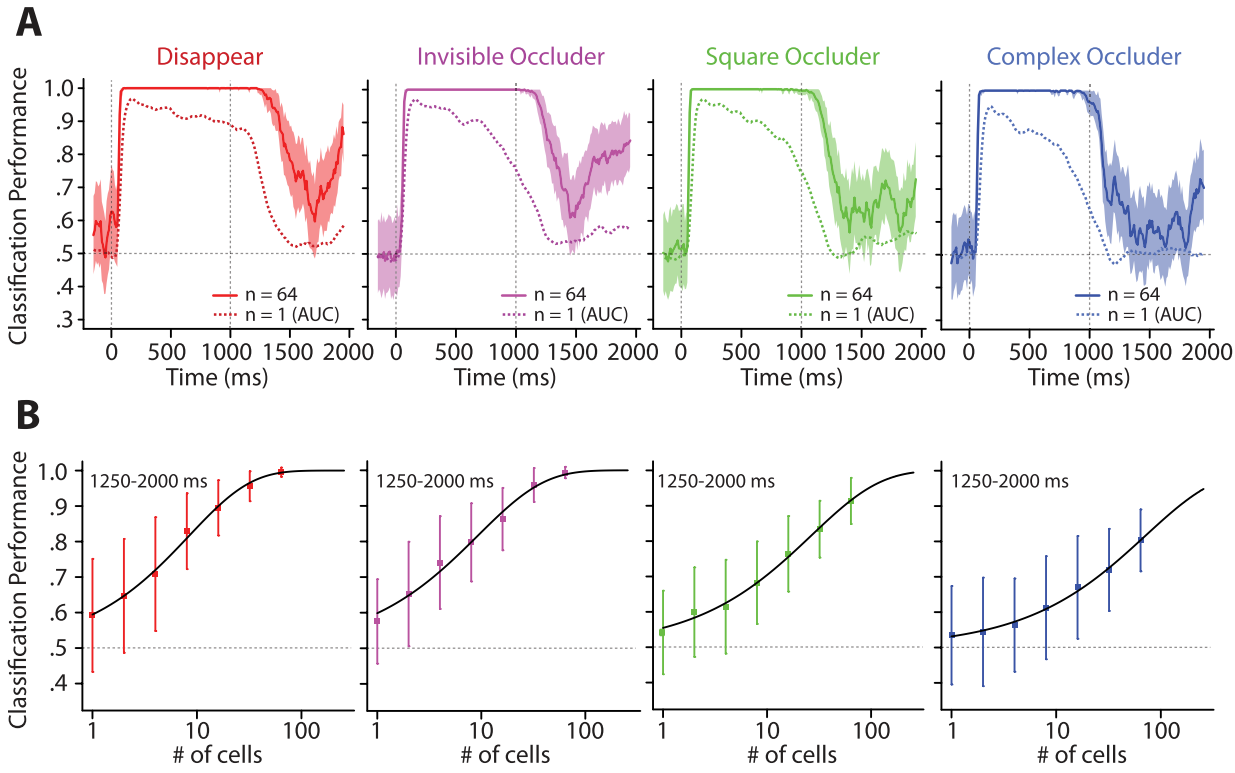

Figure 5. Population analyses using linear SVM. ***A***, Sliding window SVM analyses. Each panel plots the results of the sliding window SVM analysis (step, 10 ms; window, 100 ms) for one condition. In every case, 64 randomly selected cells were used to train and test the classifier. Shaded regions indicate ±1 SD across 100 bootstrap-like iterations of the training and testing procedure. For every single time point and for every single iteration of the resampling procedure, a new set of 64 randomly selected cells was used. For comparison, we have replotted the average AUC values from Figure 4*B* (dotted line). Notice the consistently better performance for the population of cells compared with the single cell. Also, note the rise in classification performance as the delay epoch draws to a close. ***B***, Performance of the linear SVM classifier for the entire delay period (1250–2000 ms) as a function of the number of cells and condition. Error bars indicate ±1 SD across 100 bootstrap-like iterations. Lines indicate best-fit Weibull functions. Note the well above chance performance for 64 cells in every single condition (rightmost points). Also, observe that the complex occluder Weibull fit had the highest threshold and shallowest slope, indicating that more cells would be needed to achieve the same level of classifier performance as in the other conditions.

cluder conditions, respectively). This suggests that ~100 highly selective (with respect to the encoding epoch) cells would be sufficient to mirror the behavioral performance of the subjects.

In summary, we believe the collection of delay period analyses, both at the single-cell and population level, support the hypothesis that ITC activity contains information about the visual properties of items currently represented in working memory.

Visual latency and stimulus selectivity during comparison period of VWM

We next shifted our focus to the comparison epoch. During this epoch, the subject must encode the choice stimulus, compare it with the remembered sample, and make a decision as to whether the two stimuli match. Figure 6*A* shows the population averaged SDFs sorted by condition and image effectiveness. Sliding window ROC analyses (step, 10 ms; window, 100 ms) confirmed the high discriminability of the neural responses elicited by the effective and ineffective stimuli (Fig. 6*B*). Thus, as expected, ITC neurons are likely critical to the encoding of the choice stimulus.

Visual inspection of Figure 6*A* exposed a forward masking effect not present in the sample encoding phase. Specifically, the visual response was slower to develop in the conditions in which some stimulus was present during the delay (square and complex occluder). To explore this in more detail, we quantified the visual latency for each neuron and for each condition separately (see Materials and Methods). Indeed, visual latencies varied considerably across conditions ($F_{(3,312)} = 32.11, p \ll 0.0001$; disappear, mean of 117.91 ms; invisible occluder, 118.63 ms; square occluder, 134.38 ms; complex occluder, 149.56 ms). Such an ordering of latencies led us to consider its connection with reaction times (compare Figs. 1*G*, 6*A*). To make this relationship explicit, we correlated the two measures; as anticipated, we observed a strong association between the onset of visual responses and reaction times (Fig. 6*C*) ($r = 0.66$; $p \ll 0.0001$).

Although the above results indicate that the time of arrival of visual information into ITC is strongly related to the subjects' decisions, we sought to strengthen the behavioral link by correlating reaction times with the latency of visual selectivity (Fig. 6*D*) (see Materials and Methods). As expected, the correlation between selectivity latency and reaction times ($r = 0.71$; $p \ll 0.0001$) was even greater than that between visual latency and reaction times. Taken as a whole, the robust and significant correlation analyses suggest that neural activity in ITC is involved in linking the encoding of visual objects with behavioral decisions about the identities of the objects.

Match enhancement and match suppression effects in single cells

To address the role of ITC neurons in the matching phase of VWM, we closely examined match effects, which are differences in neural activity that depend on whether the comparison stimulus is a match or nonmatch. In general, these analyses (except when otherwise noted) were restricted to the responses to the effective stimulus. As can be seen in Figure 7*A*, after the initial burst of activity, all conditions exhibited a clear match suppres-

sion effect wherein responses to the nonmatch stimulus were stronger than to the match stimulus. What is less evident at that scale, however, is a small, but remarkably consistent, match enhancement effect apparent very early in the visual response (Fig. 7*B*). We note that this enhancement could also be viewed as a decrease in response latency, but we refer to it as enhancement because, at any one time point during this epoch, the response to the match stimulus is elevated compared with that of the nonmatch. To quantify the direction and reliability of the match effects, we performed ROC analyses in which AUCs below 0.5 indicated match suppression and AUCs above 0.5 match enhancement (step, 10 ms; window, 100 ms). Averaging the AUCs across cells revealed that all conditions exhibited a reliable early match enhancement that later reversed and became match suppression (Fig. 7*C*). Thus, at the onset of the visual response, a match stimulus elicited earlier/greater activity than the exact same stimulus when it was a nonmatch.

To gain a better understanding of the match effects, for each cell and for each condition separately, we aligned match/nonmatch responses on the time point 25 ms before the visual latencies and recomputed the AUCs in increasing steps of 5 ms (Fig. 7*D*). In other words, the step at 5 ms contained spike data from −25 to −20 ms, the step at 10 ms contained data from −25 to −15 ms, and so on, until −25 to 125 ms. The analysis revealed that, for all conditions, what was at first a significant match enhancement effect was quickly obscured by a robust match suppression effect. Specifically, average AUCs over the first 50 ms were significantly greater than 0.5 (two-tailed *t* tests; disappear, $p < 0.0001$; invisible occluder, $p < 0.0001$; square occluder, $p = 0.00017$; complex occluder, $p = 0.001$), whereas those over the last 50 ms were significantly less than 0.5 (two-tailed *t* tests; disappear, $p \ll 0.0001$; invisible occluder, $p < 0.0001$; square occluder, $p < 0.0001$; complex occluder, $p = 0.0014$). The match enhancement effect was stimulus specific as responses to the ineffective stimulus exhibited solely match suppression (Fig. 7*E*). Thus, whether the match effect is interpreted as match enhancement or match suppression depends not only on the integration window but also on the specific inputs driving individual cells.

From the collection of match analyses, we conclude that the population of ITC cells retained enough information about the encoded sample to initially respond more strongly/quickly and to later dampen its response to a matching comparison stimulus. Crucially, this memory trace persisted through the presentation of intervening sensory information, a topic to which we return in Discussion.

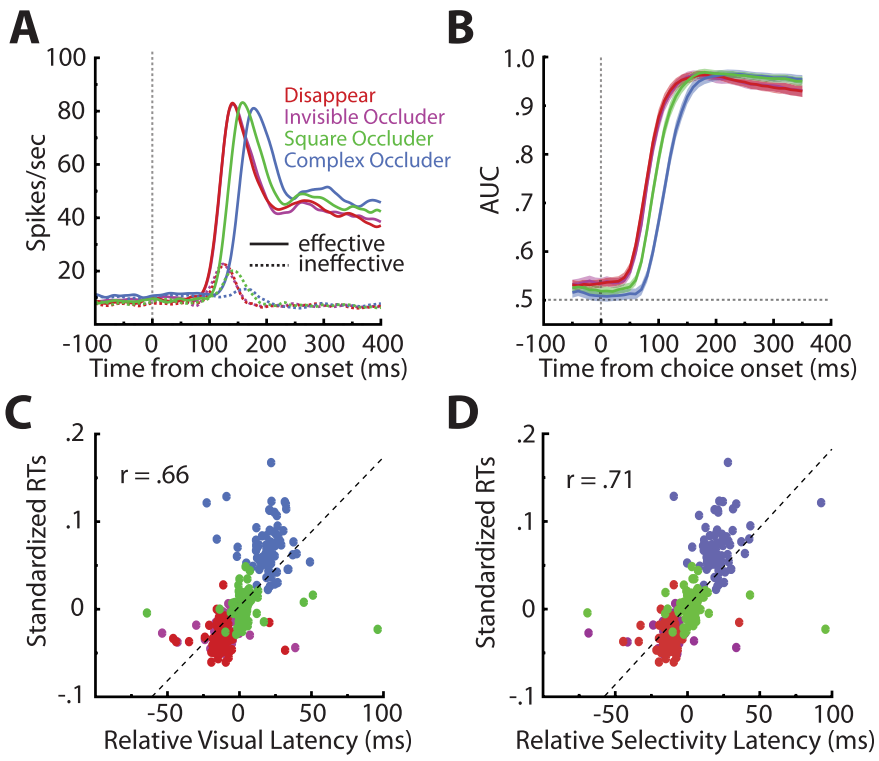

Figure 6. Analysis of choice epoch responses. ***A***, Population averaged spike density functions sorted by image effectiveness and condition. Note the temporal staggering of the responses reflecting the forward masking effect from the square and complex occluders. ***B***, Population averaged sliding window ROC analyses sorted by condition (step, 10 ms; window, 100 ms). Shaded regions indicate ±1 SEM. The graph confirms the high discriminability of the responses elicited by the effective and ineffective stimuli. ***C***, Scatter plot of the visual latencies of individual cells and the monkey's average standardized reaction times (RTs). Observe the highly significant correlation between the two measures. ***D***, Scatter plot of the selectivity latencies of individual cells and the monkey's average standardized reaction times. Note that the relationship is even stronger than between visual latency and reaction times. In both scatter plots, the colors indicate from which condition the points came and the line is the best least-squares fit to the data.

Latencies of single-cell and LFP match effects

To gain additional insight into the potential origin of the observed match effects, we computed visual selectivity and match suppression latencies for each cell and for each condition independently (Fig. 7*F*) (see Materials and Methods). If match suppression were simply fed into ITC, a scatter plot of the two measures should have most of its points around the unity line; however, most of the points lie above the line. In fact, the emergence of match suppression lagged visual selectivity by, on average, 35.37 ms (two-tailed paired *t* test, $p \ll 0.0001$), consistent with recent ITC studies exploring the dynamics of repetition suppression (Sawamura et al., 2006; McMahon and Olson, 2007; Liu et al., 2009). This suggests to us that match suppression could be a marker of finished cortical processing, which occurs faster for the matching compared with the nonmatching stimuli.

Another means of exploring whether the neural processing steps required to produce the match effects take place before or within ITC is to compare single-cell activity with LFP modulation (Monosov et al., 2008). Because the LFP is thought to reflect the synaptic input to and local processing within a cortical area, if the match effect is first evident at the single-cell level and only later shows up in the LFP, one can hypothesize that ITC underlies the generation of the match effects. Of course this analysis hinged on actually observing match effects in the LFP; indeed, match and nonmatch stimuli elicited reliably different LFP waveforms (Fig. 8*A*, collapsed across all stimuli), which is more readily appreciated by inspecting the difference waveforms (Fig. 8*B*) (for LFP difference waveforms sorted by stimulus, see supplemental Fig. S5, available at www.jneurosci.org as supplemental material).

To see when the differences achieved significance, we con-

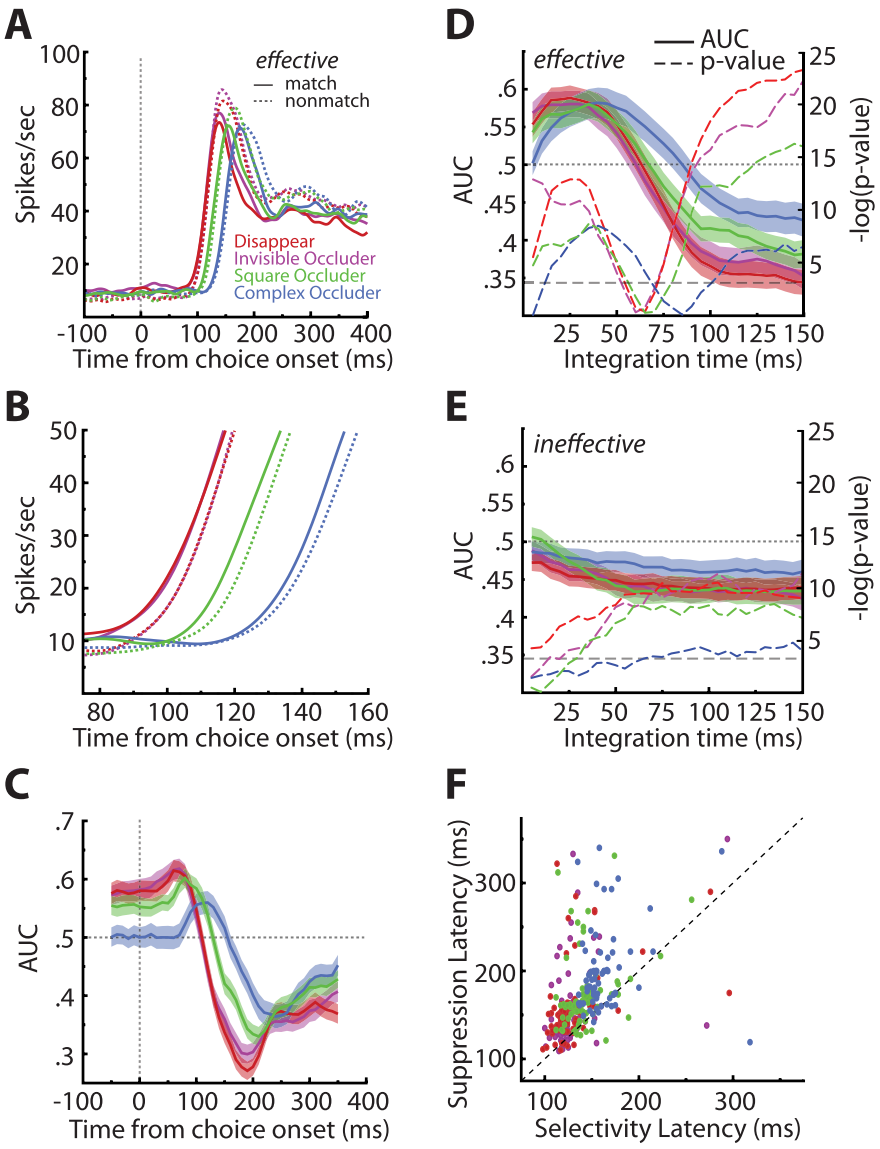

Figure 7. Single-cell match enhancement and suppression effects. ***A***, Population averaged spike density functions sorted by match status and condition, restricted to the effective stimulus. Note the large but delayed match suppression in all conditions. ***B***, Zoomed in view of the spike density functions sorted by match status and condition. This plot reveals a small but reliable match enhancement in all conditions. ***C***, Sliding window ROC analyses comparing match with nonmatch responses (step, 10 ms; window, 100 ms). Values above 0.5 indicate enhancement, whereas values below 0.5 indicate suppression. Shaded regions indicate ±1 SEM. The analysis confirmed an early match enhancement effect that was succeeded by match suppression. ***D***, Match enhancement and suppression as a function of counting window. Shaded regions indicate ±1 SEM. In all conditions, we see a reliable match enhancement that is obscured by match suppression if the integration time is allowed to be long enough. The −log(p value) of Wilcoxon's signed ranks tests done on all data points indicate that indeed both enhancement and suppression were significant [the bottom dashed line indicates −log(0.05)] (for more details, see Results). ***E***, The match enhancement was stimulus specific, because repeating the same analysis outlined in ***D*** on the ineffective stimulus yielded only match suppression. ***F***, Scatter plot of match suppression latencies as a function of visual latencies. The dashed line is what would be expected if selectivity and suppression arose simultaneously. Match suppression is unlikely to be purely bottom up because, on average, it was delayed by 35 ms relative to the first sign of visual selectivity ($p \ll 0.0001$). Colors denote from which conditions the points came.

verted the difference waveforms into t scores and found the times at which these exceeded the t scores expected by chance (two-sided permutation test, $p < 0.05$) (see Materials and Methods). Figure 8 shows the t-transformed differences separately for each condition (Fig. 8*C–F*, dark lines), with the dark bars at the bottom of the plots indicating when the differences achieved statistical significance. To directly compare the LFP differences with spike differences, we repeated the same analysis on the difference SDFs (Fig. 8*C–F*, light lines and bars). In all four conditions, the earliest significant differences (in this case match enhancement) always arose in the single-unit activity. This finding supports the conclusion that the observed match enhancement is the result of neural processes local to ITC.

The comparison of significant differences in single-unit activity and local field potentials revealed very early differences in the spike signal, particularly in the case of the invisible occluder condition in which the spike match enhancement reaches significance only ~20 ms after stimulus onset (Fig. 8*D*). This would imply that the match enhancement occurs before the onset of the induced visual response. Indeed, we do not deny this possibility but certainly do not think that this is attributable to chance. Because the single-cell match analysis was restricted to only the effective image, match trials always contained the effective image as the remembered sample. Nonmatch trials, conversely, contained any of the other, usually not as effective, images as the remembered sample. This means that match enhancement, in its current definition, extends the delay period selectivity analysis into the comparison epoch (with the caveat that the "ineffective" trials contain not only the ineffective image but also some random images). As such, match enhancement could arise before visual response onset. The fact that the complex occluder condition, which did not exhibit consistent single-cell delay period selectivity, still demonstrated match enhancement indicates that match enhancement is in fact real and can be partially dissociated from delay period selectivity. We speculate on the possible relationship between delay period selectivity and match enhancement below.

Discussion

For more than three decades, it has been known that neurons in prefrontal cortex exhibit persistent activity during the delay periods of working memory tasks (Fuster and Alexander, 1971). The view that this activity is the neural correlate of short-term storage has come under close scrutiny recently (Rushworth et al., 1997; Rowe et al., 2000; Lebedev et al., 2004), with some proposing that the major role of PFC in VWM maintenance is not storage per se but rather the reactivation of the appropriate visual representations (Ranganath and D'Esposito, 2005). In this latter framework, activity in both the PFC and ITC is necessary to accurately recall visual memories, but the actual memory reinstatement occurs within ITC. Here, we have shown that the reliability with which individual ITC neurons signal stimulus identity increased toward the latter stages of the delay period, suggesting that the relevant memory was reactivated just before the subject was to be faced with a same/different decision. This rise in stimulus selectivity

was also apparent when the activities of multiple neurons were combined, consistent with a previous population analysis showing the same phenomenon but with regard to category instead of stimulus identity information (Meyers et al., 2008). It is appealing to speculate that feedback projections from PFC, known from anatomical studies to be abundant (Pandya and Yeterian, 1990), were responsible for reinstating stimulus-specific neural activity within ITC. This conjecture is further supported by studies showing anticipatory delay period activity within PFC itself (Rainer et al., 1999; Rainer and Miller, 2002).

Another possible source of feedback to ITC is the medially adjacent perirhinal cortex (PRh). Although both ITC and PRh have the ability to respond robustly to a non-optimal stimulus if it signals the presentation of the preferred stimulus, this retrieval signal appears first in the PRh and only later in ITC, suggesting that it flows backward from PRh to ITC (Sakai and Miyashita, 1991; Naya et al., 1996, 2001). Most germane to this discussion is the observation that delay period selectivity in both PRh and ITC persists through distractors only for the sought after target and not for the stimulus that cued the retrieval of that target (Takeda et al., 2005). In our task, the cue and target were the same stimulus so it is difficult to tease apart cue- and target-related activities. However, if for the moment we assume that the rise in delay period selectivity is target related and PRh neurons have the ability to retrieve the same item that served as the cue, then it is possible that PRh neurons participate in reinstating working memory contents in ITC. Other medial temporal lobe structures, for example, the entorhinal cortex (Suzuki et al., 1997), could also contribute to this process, but their relevance to visual object retrieval specifically is less well understood.

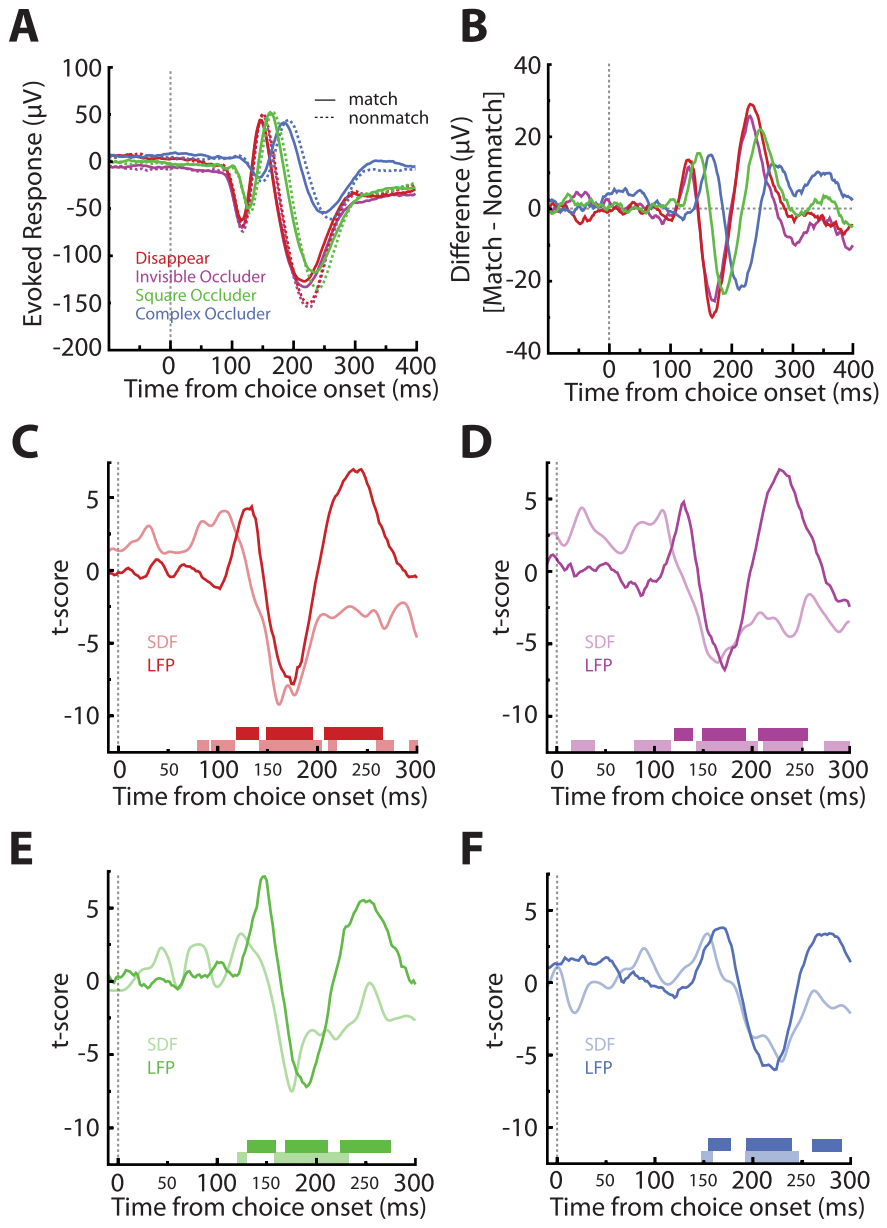

Figure 8. LFP match effects and their relation to spike match effects. ***A***, Raw LFP waveforms averaged across all stimuli and sessions and sorted by match status and condition. Note the small but reliable differences in the match/nonmatch waveforms. ***B***, LFP difference waveforms (match − nonmatch). Notice the consistent differences across the four conditions. ***C–F***, *t*-transformed LFP differences (dark lines) and SDF differences (light lines). The times at which the LFP or SDF differences became significant is indicated by the dark or light bars, respectively, at the bottom of the plots. Note that, in all cases, the very first significant match effect was present in the single-cell responses. Also, observe that the LFP differences seem to be lagging the SDF differences.

The question remains as to why we see fewer neurons regaining their stimulus-selective discharge during the delay of the complex occluder condition than in the other three conditions. We consider three alternative possibilities. First, we note that the monkeys' performance was most affected by this condition, which could be the result of a decreased proportion of ITC neurons exhibiting delay period selectivity. In other words, the subjects had a more difficult time reinstating visual memories when viewing other irrelevant visual information. This would be consistent with the notion that, when bottom-up input and short-term memory meet in ITC, bottom-up input dominates, allowing for a more veridical representation of the sensory environment.

Second, we showed that using the combined activity of up to 64 neurons dramatically improved the reliability with which we could decode stimulus identity during the delay. Thus, by usefully incorporating small amounts of information from many neurons, the monkey could do much better than by listening to only a single neuron, obviating the need for ITC to have large numbers of highly selective delay period neurons. Indeed, by extrapolating our population analyses, we showed that the behavior of the monkeys could be well matched with ~100 cells. These cell numbers, although obtained using a significantly longer time bin (~750 ms), are in line with previous studies using linear classifiers to decode stimulus information from a population of ITC neurons (Hung et al., 2005; Meyers et al., 2008). Although caution should be exercised in interpreting these results because the linear classifier does not care whether the effective stimulus elicits more or less activity than the ineffective stimulus during

the delay interval, as long as the responses are different, it is possible that ITC neurons convey different stimulus information during the encoding and delay epochs.

Third, and perhaps most importantly, the lack of firing by a neuron does not mean that the cell does not contribute to storing the contents of working memory. For example, recent computational work by Mongillo et al. (2008) has shown that synaptic calcium kinetics in a recurrent network of neurons can be used to store traces of past spiking activity. Additional modeling work by Sugase-Miyamoto et al. (2008) has focused on the ability of ITC neurons to store past visual inputs by "remembering" specific patterns of synaptic activity that occur at the time of encoding. Both of these models hypothesize that memory traces are implicitly stored in synapses and only explicitly read out in spiking form at the time of a memory recall signal and/or new visual input. Such quiescent storage models are compatible with the evidence presented in this paper, wherein individual neurons during the complex occluder condition do not show widespread delay period selectivity but do exhibit match enhancement that occurs essentially simultaneously with the arrival of new visual information. In other words, it could be the case that, if ITC spiking delay selectivity is washed out by interfering visual input, an implicit synaptic memory trace persists that manifests itself in the form of match enhancement at the time of recall. In fact, delay period selectivity itself could be reflecting synaptically stored memories, which would suggest a tight link between the neural processes underlying delay period selectivity and match enhancement.

Match enhancement has been hypothesized previously to reflect the augmentation of task-relevant visual representations (Miller and Desimone, 1994; Miller et al., 1996). In agreement with the original proposal (Desimone and Duncan, 1995; Miller and Cohen, 2001), we believe that, by biasing information flow in posterior visual areas, feedback from PFC and/or PRh, as well as intrinsic ITC circuitry, allows ITC neurons to respond more strongly and/or more quickly to the remembered stimulus. The present study has clarified the nature of this augmentation by showing that it specifically targets those neurons essential for the representation of the stimulus, because nonpreferred stimuli did not elicit an enhanced response. Furthermore, we have shown it to be more evident in ITC than previously thought, likely a result of the selectivity of our sample of neurons. Finally, our comparison of the single unit and LFP response differences leads us to believe that the processes underlying the generation of match enhancement may be localized to ITC.

An objection could be raised that we did not observe match enhancement per se but rather a speeding of the neural response that underlies perceptual priming, a phenomenon wherein behavioral responses to a repeat presentation of a stimulus are quicker and/or more accurate (Tulving and Schacter, 1990). Indeed, it could be that the monkeys relied on some sort of behavioral priming strategy to solve the task. However, the priming observed here cannot be an entirely implicit process; rather, it would have to operate on a trial-by-trial basis because all the images appeared frequently and repeatedly in a single experimental session. We do not think that this alternative interpretation is at odds with our main conclusion that a memory trace of the encoded sample is reinstated in ITC. The fact that single-cell match enhancement consistently preceded any LFP match effects argues further against a purely bottom-up explanation of match enhancement.

Interestingly, the presence of match enhancement did not preclude the appearance of, in the same population of cells, the much more frequently observed match suppression effect. In fact, it has been proposed that the relative prevalence of match suppression (Baylis and Rolls, 1987; Eskandar et al., 1992; Miller et al., 1993; Sawamura et al., 2006; Liu et al., 2009) reflects its role as the neural correlate of perceptual priming, although direct evidence linking ITC suppression to priming is scarce (Sayres and Grill-Spector, 2006; McMahon and Olson, 2007; Schacter et al., 2007). One specific hypothesis relating match suppression to priming is the sharpening of neuronal responses. In this view, activity patterns are pruned via suppression, with the repeat encoding of a stimulus carried by a smaller, but more selective, pool of neurons (Desimone, 1996; Wiggs and Martin, 1998). Given our sample of selective cells, this model would have predicted that we should see minimal suppression. Contrary to this prediction, we observed robust suppression. Thus, it is unlikely that the sharpening model can fully account for our data (McMahon and Olson, 2007).

Instead, we believe the late suppression reflects a more general phenomenon in which a matching or familiar stimulus is more quickly and efficiently processed (Freedman et al., 2006; Anderson et al., 2008). From this perspective, suppression is a marker of process completion, supporting, to some extent, the facilitation model of suppression (James et al., 2000; Grill-Spector et al., 2006; James and Gauthier, 2006). This interpretation is bolstered by three results. First, match suppression did not appear until ~35 ms after visual selectivity, and this presuppression epoch may be sufficient for the signaling of behaviorally relevant information (Allred and Jagadeesh, 2007). Second, suppression reliably followed enhancement, with enhancement speculated to be the more pertinent of the two signals to solving the task (Miller and Desimone, 1994; Desimone, 1996; Miller et al., 1996). Third, we believe that the initial significant differences in the LFP difference waveforms (Fig. 8*C–F*) are a consequence of the matching waveforms advancing more rapidly than their nonmatching counterparts. This is most easily seen by comparing the solid and dashed blue curves in Figure 8*A*.

We also believe that the last of the three significant LFP match effects (Fig. 8*C–F*) is the LFP signature of suppression. Because this particular effect arose after spike suppression, we hypothesize that match suppression may also first be present at the level of ITC. If indeed suppression reflects finished computations, this finding would support the hypothesis that ITC activity, and not the entire visual cortex, is responsible for the more efficient encoding of a repeated object's identity.

In conclusion, by recording from extremely selective neurons, we have shown that delay period selectivity in ITC, an area thought to be most important for visual object encoding, is quite prevalent, reaching significance in an unprecedented fraction of neurons (45.6 and 49.4% in the disappear and invisible occluder conditions, respectively, if we include all significantly modulated cells, not just $r_{\text{effective}} > r_{\text{ineffective}}$). This extensive delay period selectivity is unlikely to simply be a consequence of passive, residual encoding activity because its reliability significantly increased in the latter stages of the memory epoch. Although the magnitude of stimulus-selective delay responses was attenuated in the presence of a distracting stimulus, there remained a small but significant fraction of cells that represented the relevant object identity. This delay selectivity was augmented considerably if we simultaneously used the activity of multiple neurons, even in the complex occluder condition. Furthermore, in all conditions, we observed an early match enhancement effect, which could ultimately be the signal of most relevance to the monkey for the behavioral task we have explored. Coupled with the fact that activity in ITC was so intimately linked with reaction times, we

believe that the data support the hypothesis that neural activity in ITC, with the help of modulatory feedback most likely arising from PFC and/or PRh, contributes not only to the encoding phase of VWM but also to the temporary representation of no longer available complex visual input.

References

Allred SR, Jagadeesh B (2007) Quantitative comparison between neural response in macaque inferotemporal cortex and behavioral discrimination of photographic images. J Neurophysiol 98:1263–1277.

Anderson B, Mruczek RE, Kawasaki K, Sheinberg D (2008) Effects of familiarity on neural activity in monkey inferior temporal lobe. Cereb Cortex 18:2540–2552.

Appelbaum LG, Wade AR, Vildavski VY, Pettet MW, Norcia AM (2006) Cue-invariant networks for figure and background processing in human visual cortex. J Neurosci 26:11695–11708.

Baddeley A (1996) The fractionation of working memory. Proc Natl Acad Sci USA 93:13468–13472.

Baker CI, Keysers C, Jellema T, Wicker B, Perrett DI (2001) Neuronal representation of disappearing and hidden objects in temporal cortex of the macaque. Exp Brain Res 140:375–381.

Baylis GC, Rolls ET (1987) Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp Brain Res 5:614–622.

Blair RC, Karniski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30:518–524.

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge, UK: Cambridge UP.

Desimone R (1996) Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci USA 93:13494–13499.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222.

Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A (2008) e1071: miscellaneous functions of the department of statistics (e1071), Technische Universität Wien. R package version 1.5–18.

Druzgal TJ, D'Esposito M (2003) Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. J Cogn Neurosci 15:771–784.

Eskandar EN, Richmond BJ, Optican LM (1992) Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. J Neurophysiol 68:1277–1295.

Fiebach CJ, Rissman J, D'Esposito M (2006) Modulation of inferotemporal cortex activation during verbal working memory maintenance. Neuron 51:251–261.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2006) Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cereb Cortex 16:1631–1644.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Fuster JM, Jervey JP (1982) Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. J Neurosci 2:361–375.

Goldman-Rakic PS (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Handbook of physiology: the nervous system (Plum F, Mountcastle VB, eds), pp 373–417. Bethesda, MD: American Physiological Society.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Huntington, NY: Krieger.

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23.

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36.

Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. Science 310:863–866.

James TW, Gauthier I (2006) Repetition-induced changes in BOLD response reflect accumulation of neural activity. Hum Brain Mapp 27:37–46.

James TW, Humphrey GK, Gati JS, Menon RS, Goodale MA (2000) The effects of visual object priming on brain activation before and after recognition. Curr Biol 10:1017–1024.

Kaufman J, Csibra G, Johnson MH (2003) Representing occluded objects in the human infant brain. Proc Biol Sci 270 [Suppl 2]:S140–S143.

Kaufman J, Csibra G, Johnson MH (2005) Oscillatory activity in the infant brain reflects object maintenance. Proc Natl Acad Sci USA 102:15271–15274.

Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. PLoS Biol 2:e365.

Lee H, Simpson GV, Logothetis NK, Rainer G (2005) Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. Neuron 45:147–156.

Lewis-Peacock JA, Postle BR (2008) Temporary activation of long-term memory supports working memory. J Neurosci 28:8765–8771.

Liu Y, Murray SO, Jagadeesh B (2009) Time course and stimulus dependence of repetition-induced response suppression in inferotemporal cortex. J Neurophysiol 101:418–436.

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

McMahon DB, Olson CR (2007) Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. J Neurophysiol 97:3532–3543.

Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T (2008) Dynamic population coding of category information in inferior temporal and prefrontal cortex. J Neurophysiol 100:1407–1419.

Mikami A (1995) Visual neurons with higher selectivity can retain memory in the monkey temporal cortex. Neurosci Lett 192:157–160.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

Miller EK, Desimone R (1994) Parallel neuronal mechanisms for short-term memory. Science 263:520–522.

Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J Neurosci 13:1460–1478.

Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci 16:5154–5167.

Mishkin M (1982) A memory system in the monkey. Philos Trans R Soc Lond B Biol Sci 298:83–95.

Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331:68–70.

Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319:1543–1546.

Monosov IE, Trageser JC, Thompson KG (2008) Measurements of simultaneously recorded spiking activity and local field potentials suggest that spatial selection emerges in the frontal eye field. Neuron 57:614–625.

Nakamura K, Kubota K (1995) Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. J Neurophysiol 74:162–178.

Naya Y, Sakai K, Miyashita Y (1996) Activity of primate inferotemporal neurons related to a sought target in pair-association task. Proc Natl Acad Sci USA 93:2664–2669.

Naya Y, Yoshida M, Miyashita Y (2001) Backward spreading of memory-retrieval signal in the primate temporal cortex. Science 291:661–664.

Pandya DN, Yeterian EH (1990) Prefrontal cortex in relation to other cortical areas in rhesus monkey: architecture and connections. Prog Brain Res 85:63–94.

Pasternak T, Greenlee MW (2005) Working memory in primate sensory systems. Nat Rev Neurosci 6:97–107.

Pessoa L, Gutierrez E, Bandettini P, Ungerleider L (2002) Neural correlates of visual working memory: fMRI amplitude predicts task performance. Neuron 35:975–987.

Petrides M (2000) Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. J Neurosci 20:7496–7503.

Quick RF Jr (1974) A vector-magnitude model of contrast detection. Kybernetik 16:65–67.

Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput 16:1661–1687.

Rainer G, Miller EK (2002) Timecourse of object-related neural activity in the primate prefrontal cortex during a short-term memory task. Eur J Neurosci 15:1244–1254.

Rainer G, Rao SC, Miller EK (1999) Prospective coding for objects in primate prefrontal cortex. J Neurosci 19:5493–5505.
Ranganath C, D'Esposito M (2005) Directing the mind's eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. Curr Opin Neurobiol 15:175–182.
Ranganath C, Cohen MX, Dam C, D'Esposito M (2004) Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. J Neurosci 24:3917–3925.
Rowe JB, Toni I, Josephs O, Frackowiak RS, Passingham RE (2000) The prefrontal cortex: response selection or maintenance within working memory? Science 288:1656–1660.
Rushworth MF, Nixon PD, Eacott MJ, Passingham RE (1997) Ventral prefrontal cortex is not essential for working memory. J Neurosci 17:4829–4838.
Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. Nature 354:152–155.
Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. Neuron 49:307–318.
Sayres R, Grill-Spector K (2006) Object-selective cortex exhibits performance-independent repetition suppression. J Neurophysiol 95:995–1007.
Schacter DL, Wig GS, Stevens WD (2007) Reductions in cortical activity during priming. Curr Opin Neurobiol 17:171–176.
Sugase-Miyamoto Y, Liu Z, Wiener MC, Optican LM, Richmond BJ (2008) Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. PLoS Comput Biol 4:e1000073.
Suzuki WA, Miller EK, Desimone R (1997) Object and place memory in the macaque entorhinal cortex. J Neurophysiol 78:1062–1081.
Takeda M, Naya Y, Fujimichi R, Takeuchi D, Miyashita Y (2005) Active maintenance of associative mnemonic signal in monkey inferior temporal cortex. Neuron 48:839–848.
Tallon-Baudry C, Bertrand O, Peronnet F, Pernier J (1998) Induced gamma-band activity during the delay of a visual short-term memory task in humans. J Neurosci 18:4244–4254.
Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428:751–754.
Tulving E, Schacter DL (1990) Priming and human memory systems. Science 247:301–306.
Wiggs CL, Martin A (1998) Properties and mechanisms of perceptual priming. Curr Opin Neurobiol 8:227–233.
Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.
Zaksas D, Pasternak T (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. J Neurosci 26:11726–11742.