Object decoding with attention in inferior temporal cortex

Ying Zhang[a,1], Ethan M. Meyers[a,1,2], Narcisse P. Bichot[a], Thomas Serre[a,b], Tomaso A. Poggio[a], and Robert Desimone[a]

[a]Department of Brain and Cognitive Sciences, McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA 02139; and [b]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912

Edited by Charles G. Gross, Princeton University, Princeton, NJ, and approved April 11, 2011 (received for review January 20, 2011)

Recognizing objects in cluttered scenes requires attentional mechanisms to filter out distracting information. Previous studies have found several physiological correlates of attention in visual cortex, including larger responses for attended objects. However, it has been unclear whether these attention-related changes have a large impact on information about objects at the neural population level. To address this question, we trained monkeys to covertly deploy their visual attention from a central fixation point to one of three objects displayed in the periphery, and we decoded information about the identity and position of the objects from populations of ∼200 neurons from the inferior temporal cortex using a pattern classifier. The results show that before attention was deployed, information about the identity and position of each object was greatly reduced relative to when these objects were shown in isolation. However, when a monkey attended to an object, the pattern of neural activity, represented as a vector with dimensionality equal to the size of the neural population, was restored toward the vector representing the isolated object. Despite this nearly exclusive representation of the attended object, an increase in the salience of nonattended objects caused "bottom-up" mechanisms to override these "top-down" attentional enhancements. The method described here can be used to assess which attention-related physiological changes are directly related to object recognition, and should be helpful in assessing the role of additional physiological changes in the future.

macaque | vision | readout | population coding | neural coding

Previous work examining how attention influences the ventral visual pathway has shown that attending to a stimulus in the receptive field (RF) of a neuron is correlated with increases in firing rates or effective contrast, increases in gamma synchronization, and decreases in the Fano factor and noise correlation, compared with when attention is directed outside the RF (1–8). However, because these effects are often relatively modest, it has been unclear whether these effects would have a large impact on information contained at the population level when any arbitrary stimulus needs to be represented. Indeed, recent work has suggested that high-level brain areas can represent multiple objects with the same accuracy as single objects even when attention is not directed to a specific object (9), which raises questions about the importance of the attention-related effects that have been reported in previous studies.

Another feature of the previous neurophysiology work on attention has been that it has primarily focused on the neural mechanisms that underlie attention (i.e., what neural circuits/processing underlie the changes seen with attention). This approach, which is related to David Marr's implementational level of analysis (10), has been fruitful, as evidenced by the fact that several mechanistic models have been created that can account for a variety of firing-rate changes seen in a number of studies (11–20). Less work, however, has focused on Marr's "algorithmic/representational level," which in this context would address how particular physiological changes enable improvements in neural representations that are useful in solving specific "computational-level" tasks (such as recognizing objects).

To assess the significance of particular physiological changes associated with changes in attentional state, and to gain a deeper algorithmic/representational-level understanding of how attention impacts visual object recognition, we use a neural population decoding approach (21, 22) to analyze electrophysiological data. Our approach is based on the hypothesis that visual objects are represented by patterns of activity across populations of neurons. Thus, we assess how these representations change when the visual objects are displayed in clutter and when spatial attention is deployed. Our results show that even limited clutter decreases information about particular objects in inferior temporal cortex (IT), and that attention-related firing-rate changes significantly increase the amount of information about behaviorally relevant objects in IT. Additionally, by focusing on how information is represented by populations of neurons, we find that "competitive" effects that occur when two stimuli are presented within a neuron's RF, and global "gain-like" effects that occur when a single stimulus is presented within a neuron's RF, can both be viewed as restoring patterns of neural activity for object identity and position information, respectively. Future work using this approach should help assess whether other physiological changes apart from firing-rate changes have an important impact on information content of IT, and should further help illuminate the computations that underlie object recognition.

Results

We recorded the responses of IT neurons to either one or three extrafoveal stimuli in the contralateral visual field while monkeys fixated a spot at the center of a display (Fig. 1*A* and Fig. S1). The three stimuli were positioned so that each was likely to be contained within a different RF of cells in V4 and lower-order areas but within the same large RFs of IT cells. When one stimulus appeared in isolation, it was always the task-relevant target, but when three stimuli appeared, one was the target while the other two stimuli were distractors on a given trial. Approximately 525 ms after the stimuli onset, a directional cue (line segment) appeared that "pointed" to the target stimulus to attend. The monkey was rewarded for making a saccade to the target stimulus when it changed slightly in color, which occurred randomly from 518 to 1,260 ms after cue onset. On half of the trials, one of the distractor stimuli changed color before the target change (foils), but the monkey was required to withhold a saccade to it. Of trials that the monkeys fixated until the time of cue onset, correct saccades to the target color change occurred on ∼72% of trials, and incorrect saccades to

Author contributions: Y.Z., E.M.M., N.P.B., T.S., T.A.P., and R.D. designed research; Y.Z. and N.P.B. performed research; E.M.M. contributed new reagents/analytic tools; Y.Z. and E.M.M. analyzed data; and E.M.M., T.A.P., and R.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Y.Z. and E.M.M. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: emeyers@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1100999108/-/DCSupplemental.

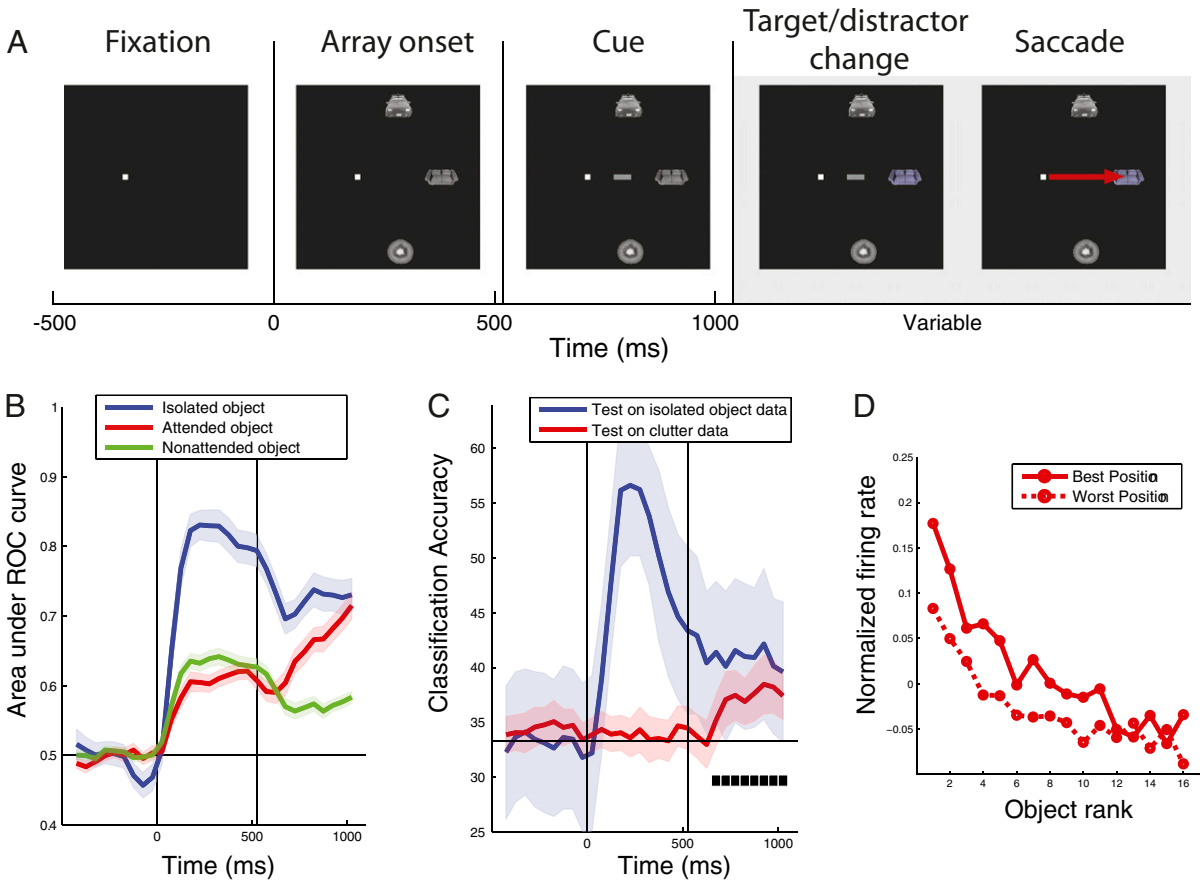

Fig. 1. Effects of attention on decoding accuracy. (*A*) Timeline for three-object trials. Single-object trials had the same timeline, except only one object was displayed. It should be noted that the attentional cue was shown for both isolated- and three-object trials, and once the cue was displayed it remained on the screen for the remainder of the trial (which could lead to potential visual–visual interactions). (*B*) Decoding accuracies for which object was shown on isolated-object trials (blue traces), and the attended object (red trace) and nonattended objects (green trace) in the three-object displays. Vertical lines indicate the times of stimulus onset, and cue onset. Colored shaded regions indicate ±1 SE of the decoding results (*Methods*). (*C*) Decoding accuracies for the position of the isolated stimulus (blue trace) and the attended stimulus (red trace). Black square boxes indicate times when the decoding accuracy for the position of the attended object was above what would be expected by chance (chance performance is 33%). (*D*) Z-score–normalized population firing rates to cluttered-display images ranked based on their isolated-object preferences. The data from isolated-object trials were first used to calculate each neuron's best and worst position and the ranking of its best to worst stimuli. The firing rates to these stimuli on cluttered trials were then calculated and averaged over all neurons, and are plotted separately for attention to the best versus worst position. Attending to the neuron's preferred position led to a relatively constant offset in the neuron's object tuning profile.

a distractor color change were made on only ~1% of trials. On the other 27% error trials, 36% of those were due to early saccades to the target location before the color change, and the remaining errors were simply random breaks in fixation.

To understand how information about objects is represented by populations of IT neurons, we applied population decoding methods (21, 22) to the firing rates of pseudopopulations of 187 neurons from two monkeys on a first stimulus set (similar results were obtained from each monkey, so the data were combined; Fig. S2) and on a second stimulus set shown to monkey 2 (Fig. S3). (By "pseudopopulation" response, we mean the response of a population of neurons that were recorded under the same stimulus conditions but the recordings were made in separate sessions, i.e., the neurons were not recorded simultaneously but treated as though they had been.) We trained a pattern classifier on data from isolated-object trials and then made predictions about which objects were shown on either different isolated-object trials or on trials in which three objects had been shown (*Methods*). Fig. 1*B* shows that information about the identity of isolated objects (blue trace) rose rapidly after stimulus onset, reaching a peak value for the area under the receiver operating characteristic (AUROC) curve of 0.83 ± 0.022 at 225 ms after stimulus onset, whereas information about the objects in the multiple-object displays also rose after the onset of the stimuli (red and green traces) but only reached a peak value of 0.62 ± 0.014 before the onset of the attentional cue. An AUROC of 0.5 represents chance performance. Thus, 75 ± 75 ms after the onset of the stimulus array, the amount of information about the objects in the three-object displays was greatly reduced compared with when these objects were shown in isolation ($P < 0.01$, permutation test; see *SI Text* for more details), showing that clutter has a significant impact on the amount of information about specific objects in IT (also see Fig. S2).

Approximately 150 ± 75 ms after the attentional cue was displayed, information about the attended object (red trace) rose significantly above the amount of information seen in the non-attended object ($P < 0.01$, permutation test). By 400 ms after cue onset, information about the attended object had reached an AUROC value of 0.64 ± 0.017, which was similar to the value of 0.68 ± 0.024 for decoding isolated-object trials during the same trial period. At the same time, information about the non-attended stimuli (green trace) decreased to a value of 0.56 ± 0.010. Thus, location-directed attention can have a significant impact on the amount of information about specific objects in IT.

These attention-related changes can also be observed in the firing rate of the population of neurons to preferred and nonpreferred stimuli (Fig. S4).

In addition to identity information, position information was also enhanced (Fig. 1*C*). When this position enhancement was examined using more conventional analyses that create tuning curves for each neuron (by ordering each neuron's responses from the best to the worst stimulus) and then plotting the population average tuning curves separately for the best versus worst location (Fig. 1*D*), this position enhancement with attention appeared as an upward shift in the population tuning curves, as has previously been reported (1). However, this upward shift is a consequence of aligning all neurons' responses to their preferred and nonpreferred locations, rather than a result of all neurons increasing their firing rates with attention. When attention is allocated to a particular location, increases and decreases in activity occur in different neurons (depending on a neuron's RF structure), which creates a distributed pattern of activity that contains information about the location of where the monkey is attending.

In the cluttered decoding results described above, we trained the classifier with data from isolated-object trials and then tested the classifier with data from three-object trials. This allowed us to test whether one of the effects of attention was to restore the pattern of neural activity to a state that was similar to when an object was shown in isolation. However, it is possible that attention could have additional effects on neural representations that modify the representation of each object to make them more distinct from one another (and thus increase the amount of information about the objects), but in a way that is not related to the neural representations that are present when the objects are shown in isolation. To test this possibility, we compared the decoding accuracies when training the classifier on cluttered-display data (Fig. 2, dashed traces) to the decoding accuracy when training with isolated-object data (Fig. 2, solid traces). If attention added additional information, then there should be higher accuracy when training with cluttered data from the cue period (Fig. 2 *Right*) than when training with isolated-object data. The results show that training on isolated-object trial data was better than training on cluttered trial data during the array period before the attentional cue (Fig. 2 *Left*). This result is consistent with the idea that clutter decreases information about specific objects. Most importantly, training the classifier with cluttered data in the array period (right subplots) did not lead to better performance than training the classifier with isolated-object data (in particular, the dashed line is not higher than the solid line in the right plot of Fig. 2*B* at any point in time). Thus, it appears that attention's main effect was to restore the pop-

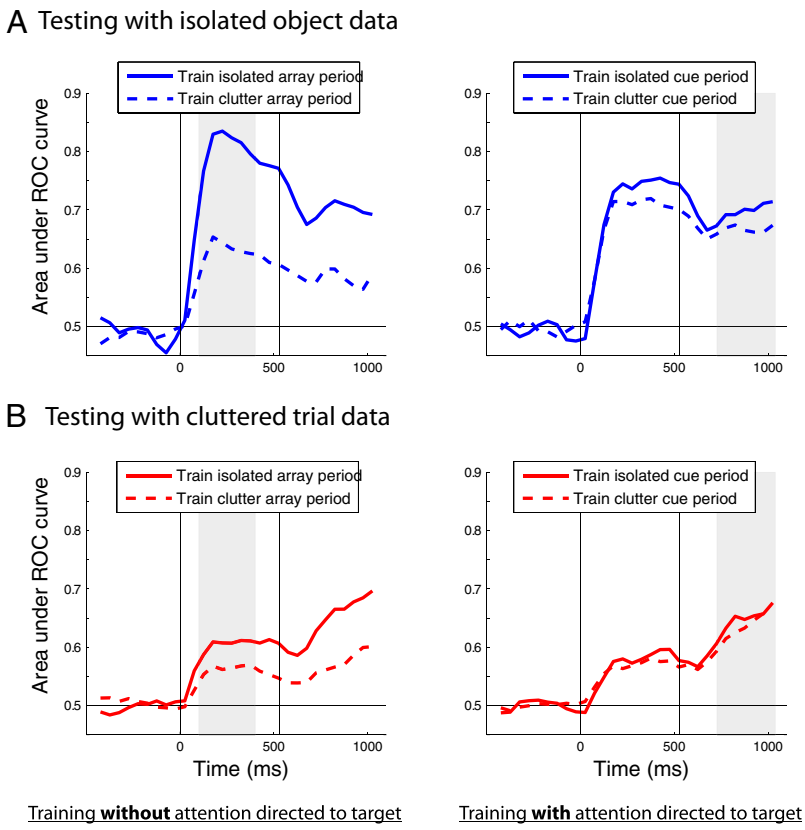

Fig. 2. Attention restores neural activity to a state that is similar to when the attended object is presented alone. Results are based on training a classifier with either 11 examples of isolated objects (solid lines) or with 11 examples of the same (attended) objects in cluttered displays (dashed line). The classifier was trained with either 300 ms of data from the array period (*Left*, gray shaded region) before the attentional cue or 310 ms of data from the period following the onset of the attentional cue (*Right*, gray shaded region). The classifier's performance was tested with either isolated-object trials (*A*) or with cluttered array data (*B*), using firing rates from 150 ms of data sampled in 50-ms sliding intervals. The results show that when training with data from the array period, better performance is achieved when training with isolated-object data compared with clutter data (left two plots), confirming that clutter reduces information about particular objects. Approximately equal levels of performance were obtained when training with either isolated or cluttered data from the period following the attentional cue (right two plots), indicating that attention caused the neural representation to enter a state that was similar to when objects were presented in isolation. It should be noted that for the isolated-object trials there is slightly more information about the stimuli when they are first shown, and hence the decoding accuracies are higher when training with data from the array period compared with training with data from the cue period (i.e., the solid blue trace in *A Left* is slightly higher than the solid blue trace in *A Right*).

ulation of neural activity to a state that was similar to that when the attended object was shown in isolation.

The above results show that top-down attention had a large impact on the object information represented in IT. We then asked how resistant these object representations would be to salient changes in the distractor. To test this, we aligned the data to the time when a distractor underwent a color change, and we decoded the identity of both the target and the distractor stimuli. The results, plotted in Fig. 3, show that before the distractor change there was a large improvement in decoding with attention (red trace) as seen before. However, when the distractor changed color, the dominant representation in IT switched transiently to the distractor object (light green trace), before returning to the attended-object representation (red trace). Thus, bottom-up, or stimulus, changes in the saliency of the distractor objects overrode the top-down attention-induced enhancements of particular objects. An examination of behavioral data (Fig. S5) revealed that reaction times were longer when the target changed soon after the distractor change, suggesting that the monkeys transiently switched their attention to the salient distractor, which impaired their ability to detect a target color change.

Discussion

Previous work has shown that several different physiological changes are correlated with changes in attentional state. It has been unclear, however, which of these physiological changes are important for object recognition. In this work, we show that visual clutter does indeed reduce the amount of information about specific objects in IT, and that attention-related *firing-rate changes* do indeed have a large impact on the amount of information present about particular objects. By applying these same methods to simultaneously recorded neural activity in future studies, it should be possible to assess whether changes in noise correlations or synchrony also have an impact on information at the population level.

Across the studies that have examined attention-related firing-rate changes in neurons in the ventral visual pathway (and in area V4 in particular), two seemingly distinct effects have been widely reported. The first effect occurs when a preferred stimulus and a nonpreferred stimulus are presented simultaneously in a neuron's RF and the monkey must pay attention to either the preferred or the nonpreferred stimulus depending on a cue that varies from trial to trial. (By "preferred stimulus," we mean a stimulus that elicits a high firing rate from the neuron when it is presented in isolation, and by "nonpreferred stimulus," we mean a stimulus that elicits a low firing rate when it is presented in isolation.) Results from these studies show that firing rates increase when the monkey attends to the preferred stimulus and that firing rates decrease when the monkey attends to a nonpreferred stimulus. The second attention-related firing-rate change occurs when a single stimulus is shown in a neuron's RF, and orientation tuning curves for this single stimulus are mapped out when the monkey attends either inside the neuron's RF or outside the neuron's RF. Under these conditions, the tuning curve for the neuron is scaled upward at all orientations when the monkey attends inside the neuron's RF, in a way that is consistent with the tuning curve being multiplied by a constant. Together, these effects are consistent with "biased competition" and closely related normalization models (1, 3, 11, 13, 14). By analyzing our IT data in a similar way to these previous studies, we were able to see similar attention effects on IT responses (Fig. S4 and Fig. 1*D*). However, from a population coding perspective, these effects appear to be very similar because they both create distinct patterns of neural activity that contain information about attended objects (with the patterns of activity for identity and position information overlapping one another within the same population). A consequence of this viewpoint is that the limited spatial nonuniformity/extent of a neuron's RF is not a deficit in terms of achieving complete position invariance but rather a useful property that enables more precise signaling of position information.

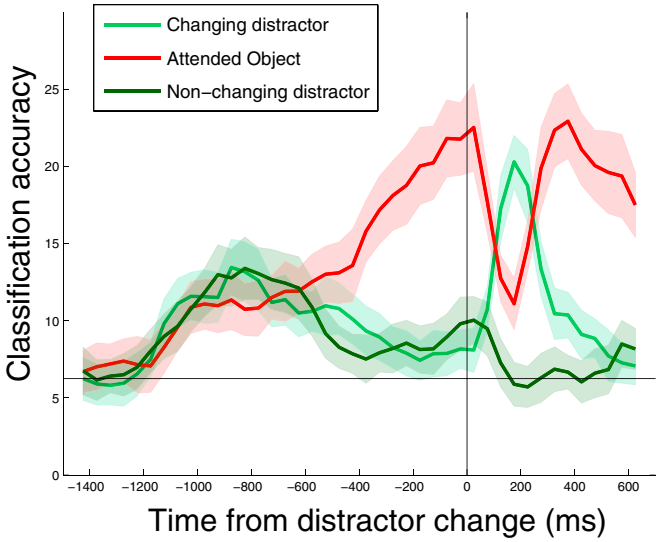

Fig. 3. Changes in the salience of distractor stimuli dominate over attention-related enhancements. A comparison of the decoding accuracies for the attended stimulus (red trace) to the distractor that underwent a color change (light green trace) and the distractor that did not undergo a color change (dark green trace). The data are aligned to the time when one of the distractors underwent a color change (black vertical bar). Chance decoding accuracy is 1/16 or 6.25%.

One discrepancy between our results and previous findings is represented by a study by Li et al. (9), which used similar population decoding methods and reported that clutter does not affect the amount of information about particular objects in IT. Although differences in stimulus parameters might be able to partially account for these effects (the stimuli used by Li et al. were smaller and presented closer to the fovea), we think that the largest factor contributing to the difference in the results was the way the classifiers were trained and tested. In particular, Li et al. trained and tested their classifier using the exact same cluttered scenes. Thus, it is possible that their classifier relied on the exact configuration in the images (by perhaps relying on visual features that spanned multiple objects) to achieve a high level of classification performance in the cluttered condition. In our study, we trained the classifier either on isolated objects (Fig. 1*B*) or using different cluttered scenes (Fig. 2) so that we would capture what is the more behaviorally relevant condition, namely being able to learn an object in isolation or on a particular background, and then being able to recognize it when seen in a different context. Indeed, when Li et al. replicated our analysis by training on isolated objects, they also found a similar decrease in classification accuracy for the cluttered conditions.

It is also important to note that the effects reported here may underestimate the impact that attention has on neural representations in IT. If the monkeys' attentional state was under stronger control by using a more difficult task (23, 24) or if the task the monkey engaged in more closely matched the information that was to be decoded (e.g., if the monkey was doing a shape discrimination task rather than a color change detection task), the effects of attention might have been even stronger. Additionally, we have seen in this study (Fig. 1*B*), and in a number of analyses of different datasets from IT, that the largest amount of information occurs when the stimuli first appear (21, 22, 25). Thus, we might also see larger attentional effects using a precuing attentional paradigm. However, we should note that even with these limitations, IT object representations with clutter

were restored by attention to nearly the same level of accuracy found with isolated objects in the visual field. Finally, it should be noted that we might be able to find stronger attention-related effects by decoding data from cells that were recorded simultaneously. Indeed, a recent study by Cohen and Maunsell (4) has suggested that one of the primary ways that attention improves the signal in a population is through a decrease in noise correlations. We briefly tried to address this issue by adding noise correlations to our pseudopopulation vectors, and found that the decoding accuracies were largely unchanged (Fig. S6). However, a more detailed examination of these effects using actual simultaneously recorded data is needed before we can draw any strong conclusions.

From an algorithmic-level viewpoint, the results seen in our study are consistent with the following interpretation. Spatial attention gates signals from a retinotopic area (say V4, in which RFs are smaller than the distance between the objects in our stimuli) to IT so that the responses to clutter stimuli do not interfere with the activity elicited by the attended object in IT. Recent experimental results (26) and computational models of attention (11–20, 27, 28) are consistent with this interpretation and furthermore suggest that the clutter interference has the form of a normalization operation similar to the biased competition model. Overall, our results support the view that the main goal of attention is to suppress neural interference induced by clutter to allow higher modules to recognize an object in context after learning its appearance from presentations in isolation (or on a different background).

Methods

Experimental Procedures. Procedures were done according to National Institutes of Health guidelines and were approved by the Massachusetts Institute of Technology Animal Care and Use Committee. All unit recordings were made from anterior IT.

Visual Stimuli. The visual stimuli consisted of 16 objects from four categories (cars, faces, couches, and fruit), and are shown in Fig. S1. The stimuli were 2.3° × 2.3° in size and were shown at an eccentricity of 5.5° from fixation at angles of +60°, 0°, and −60° relative to the horizontal meridian. The stimulus sizes/locations were chosen such that there would be little overlap between the three simultaneously presented stimuli in terms of most V4 neurons' RFs (29). For the three-object displays, 864 configurations were chosen (out of the possible 3,360 permutations of three unique objects). The cluttered displays could potentially consist of either one, two, or three objects belonging to the same category. To have a variety of hard and easy displays, we selected the displays such that two-thirds of the displays (576 displays) consisted of all three objects belonging to the same category and one-third of the displays (288 displays) consisted of all three objects belonging to different categories. After analyzing the data, we did not find a large difference between these display types, and so we grouped the results from both types of categories equivalently. A second set of seven stimuli was also shown to the second monkey for additional results presented in Fig. S3 all 630 configurations of three stimuli were used for the three-object displays in this second set of experiments.

Data Selection. A total of 98 and 139 neurons were recorded from monkey 1 and monkey 2, respectively. All of the recorded neurons were used for the individual-neuron analyses (Fig. 1*D* and Figs. S3*D* and S4). For the population decoding analyses, all neurons that had at least 12 presentations of the isolated objects and 800 trials with three-object displays were included. This resulted in 75 neurons from monkey 1 and 112 neurons from monkey 2. Because the monkeys did not always complete the full experiment, not all of the neurons had recordings from all 864 three-object images. Consequently, for the decoding analyses, we only used data from the three-object images that had been shown to all of the 75/112 neurons listed above, which gave 635 three-object trials. For the data recorded on the second stimulus set, we used all neurons that had been shown 60 repetitions of the isolated-object stimuli and all 630 three-object images, which gave us 87 usable neurons of the 132 recorded.

Data Analyses. The decoding results are based on a cross-validation procedure that has previously been described (22). The decoding method works by training a pattern classifier to "learn" which patterns of neural activity are indicative that particular experimental conditions are present (e.g., which visual stimulus has been shown) using a subset of data (called the "training set"). The reliability of the relationship between these patterns of neural activity and the different conditions (stimuli) is then assessed based on how accurately the classifier can predict which conditions are present on a separate "test set" of data.

To assess how well we could decode which stimulus was shown on the isolated-object trials (Fig. 1*B*, blue trace, and Fig. 2*A*, solid blue traces), we used a cross-validation procedure that had the following steps. (*i*) For each neuron, data from 12 trials from each of the 16 stimuli were randomly selected. For each of these trials, data from all of the neurons were concatenated to create pseudopopulation response vectors (i.e., "population" responses from neurons that were recorded under the same stimulus conditions on separate trials/sessions but were treated as though they had been recorded simultaneously). Because there were 187 neurons used, this gave 12 × 16 = 192 data points in 187-dimensional space. (*ii*) These pseudopopulation vectors were grouped into 12 splits of the data, with each split containing one pseudopopulation response vector to each of the 16 stimuli. (*iii*) A pattern classifier was trained using 11 splits of the data (176 training points), and the performance of the classifier was tested using the remaining split of the data (16 test points). Before sending the data to the classifier, a preprocessing normalization method was applied that calculated the mean and SD of each feature (neuron) using data from the training set, and a z-score normalization was applied to the training data and the test data using these means and SDs. This normalization method prevented neurons with high firing rates from dominating the outcome of the classifier. (*iv*) This procedure was repeated 12 times, leaving out a different test split each time (i.e., a 12-fold leave-one-split-out cross-validation procedure was used). (*v*) The classification accuracy from these different splits was evaluated using a measure based on the area under an ROC curve (see *SI Text* for a more detailed description of this measure). Different measures of decoding accuracy gave similar results (Fig. S7). (*vi*) The whole procedure [steps (*i*)–(*v*)] was repeated 50 times (which allowed us to assess the performance for different pseudopopulations and data splits), and the final results were averaged over all 50 repetitions.

To generate the SEs of the decoding accuracy, we used a bootstrap method that applied the above decoding procedure but created pseudopopulation vectors that sampled the neurons with replacement (being careful not to include any of the same data in the training and test sets). The SE was then estimated as the standard deviation of the mean decoding accuracy over the 50 bootstrap runs, which gave an estimate of the variability that would be present if a different subset of neurons had been selected from a similar population. Unless otherwise specified below, the decoding results in this paper are based on using a correlation coefficient classifier that was trained on the mean firing rate from 500 ms of data that started 85 ms after the onset of the stimulus, and the classifier was tested using the mean firing rates in 150-ms bins that were sample at 50-ms intervals (this created smooth curves that estimated the amount of information present in the population as a function of time). In contrast to some of the results of Meyers et al., (21), we found the neural representations in this study to be largely stationary (Fig. S8), which allowed us to use data from one training time period to decoding information at all other time points.

A similar method was used for the other decoding results reported here. To obtain the decoding accuracies for the cluttered-display objects (red and green traces in Fig. 1*B*, and also solid red traces in Fig. 2*B*), the classifier was trained on isolated-object trials using 11 repetitions of each of the 16 objects exactly as described above, but the classifier was then tested using the clutter-display trials, and the accuracies for the attended and nonattended objects were measured separately [also, because the data in the test set came from a completely different set of trials there was no need to divide the data into separate splits, so all test points were evaluated in one step (i.e., step [*iv*] was omitted)]. For the dashed traces in Fig. 2, we trained the classifier using data from the cluttered trials (again using 11 trials from each of the 16 stimuli to make a fair comparison), and the classifier was then tested using either isolated-object data (blue dashed lines in Fig. 2*A*) or the remaining cluttered-display trials that had not been used to train the classifier (dashed red lines in Fig. 2*B*). The training data for Fig. 2 were from the mean firing rates of either 300 ms of data that started 100 ms after stimulus onset (left plots) or 310 ms of data that started 200 ms after cue onset (right plots).

For the decoding of position information (Fig. 1*C*), the classifier was trained using the firing rates from isolated-object trials from a 300-ms bin that started 100 ms after the stimulus onset (thus avoiding the possibility that any visual information in the cue itself could influence the results). The results were based on a threefold cross-validation scheme, where each split contained each stimulus at all three locations (i.e., 96 total training points, and

48 test points on each split). The results from decoding the location of attention in Fig. 1*C* were based on using the same isolated-object training paradigm but the classifier was tested with cluttered displays that had 12 repetitions of each of the 16 attended stimuli at all 3 locations (576 test points). The results in Fig. 3 were based on training the classifier on isolated-object trials using 500 ms of data and 11 training points (i.e., the same paradigm used for Fig. 1*B*). The classifier was tested on 288 data points (16 stimuli $\times$ 18 repetitions) using data from the cluttered trials that were aligned to the time that the distractor underwent a color change (using 150-ms bins sampled at 50-ms intervals), and the results were compiled separately for the attended stimulus (red trace), the distractor stimulus that underwent a color change (light green trace), and the other distractor stimulus that did not undergo a color change (dark green trace).

As described in more detail in the *Discussion*, we used the simpler and more commonly used zero-one loss decoding measure (21, 22) for Figs. 1*C* and 3 because we did not need to compare attended and nonattended conditions (although the results were very similar when an AUROC measure was used). The methods used to calculate the AUROC and zero-one loss values, evaluate the statistical significance of the results, and create the supplemental figures are described in *SI Text*.

Fig. 1*D* was created by using isolated-object trials to find the position that elicited the highest and lowest firing rate for each neuron and then assessing the best to worst stimulus using 500 ms of data from the array period. These tuning curves were then plotted for the attended object using 300 ms of data from cluttered trials when attention was directed to either the best or worst position. Each neuron's firing rate was z-score–normalized before being averaged together, so that neurons with higher firing rates did not dominate the population average.

ACKNOWLEDGMENTS. We thank Jim DiCarlo and Nuo Li for their comments and for conducting additional analyses on data they collected. This research was sponsored by Defense Advanced Research Planning Agency grants (Information Processing Techniques Office and Defense Sciences Office), National Science Foundation Grants NSF-0640097 and NSF-0827427, and National Eye Institute Grant R01EY017292. Additional support was provided by Adobe, Honda Research Institute USA, and a King Abdullah University Science and Technology grant to B. DeVore. E.M.M. was supported by a National Defense Science and Engineering Graduate Research Fellowship and by a Herbert Schoemaker fellowship.

1. McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19:431–441.
2. Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291:1560–1563.
3. Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784.
4. Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12:1594–1600.
5. Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol* 70: 909–919.
6. Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19:1736–1753.
7. Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63:879–888.
8. Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55:131–141.
9. Li N, Cox DD, Zoccolan D, DiCarlo JJ (2009) What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? *J Neurophysiol* 102:360–376.
10. Marr D (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Henry Holt, New York).
11. Lee J, Maunsell JHR (2009) A normalization model of attentional modulation of single unit responses. *PLoS One* 4:e4651.
12. Chikkerur S, Serre T, Tan C, Poggio T (2010) What and where: A Bayesian inference theory of attention. *Vision Res* 50:2233–2247.
13. Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
14. Reynolds JH, Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. *Neuron* 24:19–29.
15. Lee DK, Itti L, Koch C, Braun J (1999) Attention activates winner-take-all competition among visual filters. *Nat Neurosci* 2:375–381.
16. 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.
17. Ardid S, Wang XJ, Compte A (2007) An integrated microcircuit model of attentional processing in the neocortex. *J Neurosci* 27:8486–8495.
18. Börgers C, Epstein S, Kopell NJ (2008) Gamma oscillations mediate stimulus competition and attentional selection in a cortical network model. *Proc Natl Acad Sci USA* 105:18023–18028.
19. Tiesinga PH, Fellous JM, Salinas E, José JV, Sejnowski TJ (2004) Inhibitory synchrony as a mechanism for attentional gain modulation. *J Physiol Paris* 98:296–314.
20. Tsotsos JK (1988) in *Computational Processes in Human Vision: An Interdisciplinary Perspective* (Ablex, Norwood, NJ), pp 286–338.
21. Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310:863–866.
22. 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.
23. Boudreau CE, Williford TH, Maunsell JHR (2006) Effects of task difficulty and target likelihood in area V4 of macaque monkeys. *J Neurophysiol* 96:2377–2387.
24. Spitzer H, Desimone R, Moran J (1988) Increased attention enhances both behavioral and neuronal performance. *Science* 240:338–340.
25. Gochin PM, Colombo M, Dorfman GA, Gerstein GL, Gross CG (1994) Neural ensemble coding in inferior temporal cortex. *J Neurophysiol* 71:2325–2337.
26. Buffalo EA, Bertini G, Ungerleider LG, Desimone R (2005) Impaired filtering of distracter stimuli by TE neurons following V4 and TEO lesions in macaques. *Cereb Cortex* 15:141–151.
27. Deco G, Rolls ET (2004) A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res* 44:621–642.
28. Spratling MW (2008) Predictive coding as a model of biased competition in visual attention. *Vision Res* 48:1391–1408.
29. Gattass R, Sousa AP, Gross CG (1988) Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8:1831–1845.