

# Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporal Cortex

Davide Zoccolan, Minjoon Kouh, Tomaso Poggio, and James J. DiCarlo

McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Center for Biological and Computational Learning, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

Object recognition requires both selectivity among different objects and tolerance to vastly different retinal images of the same object, resulting from natural variation in (e.g.) position, size, illumination, and clutter. Thus, discovering neuronal responses that have object selectivity and tolerance to identity-preserving transformations is fundamental to understanding object recognition. Although selectivity and tolerance are found at the highest level of the primate ventral visual stream [the inferotemporal cortex (IT)], both properties are highly varied and poorly understood. If an IT neuron has very sharp selectivity for a unique combination of object features (“diagnostic features”), this might automatically endow it with high tolerance. However, this relationship cannot be taken as given; although some IT neurons are highly object selective and some are highly tolerant, the empirical connection of these key properties is unknown. In this study, we systematically measured both object selectivity and tolerance to different identity-preserving image transformations in the spiking responses of a population of monkey IT neurons. We found that IT neurons with high object selectivity typically have low tolerance (and vice versa), regardless of how object selectivity was quantified and the type of tolerance examined. The discovery of this trade-off illuminates object selectivity and tolerance in IT and unifies a range of previous, seemingly disparate results. This finding also argues against the idea that diagnostic conjunctions of features guarantee tolerance. Instead, it is naturally explained by object recognition models in which object selectivity is built through AND-like tuning mechanisms.

**Key words:** inferotemporal cortex; monkey; object recognition; multiple objects; trade-off; tolerance

## Introduction

The key computational problem of object recognition is attaining both selectivity among different visual objects and tolerance to variation in each object’s appearance (e.g., as a result of changes in position, size, illumination, clutter, etc.). The primate visual system has solved this problem: primates robustly and effortlessly discriminate among visual objects over the wide range of images that each object produces during natural vision (Potter, 1976; Intraub, 1980; Rubín and Turano, 1992; Logothetis and Sheinberg, 1996; Thorpe et al., 1996; Edelman, 1999; Rousset et al., 2004). What neuronal architecture and computations create such a selective and tolerant representation of visual objects? Because previous work indicates that these properties are built by a hierarchy of cortical stages (the ventral visual stream) (Logothetis and Sheinberg, 1996; Tanaka, 1996; Rolls, 2000; Rousset et al., 2004), experimental studies can shed light on this question by

examining the strength, variation, and relationship of object selectivity and tolerance across the ventral visual stream.

Here we focus on the culmination of the ventral visual stream, the anterior portion of the inferotemporal cortex (IT). Although object selectivity and tolerance are found in IT, these properties are highly varied both within and across studies (Logothetis and Sheinberg, 1996; Tanaka, 1996; Rolls, 2000; Rousset et al., 2004), and their connection at the level of single IT neurons is not understood. One possibility is that, as signals propagate through the visual system, neurons become highly selective for unique combinations of features that also guarantee high tolerance to identity-preserving transformations of those features. This notion derives from the grandmother or gnostic cell concept of Lettvin and Konorski (Gross, 2002), and has been recently invigorated by the observation that some neurons in the human medial temporal lobe (Quiroga et al., 2005) respond only to specific visual objects regardless of size, pose, and visual clutter. This notion also seems consistent with previous physiological results: sharpness of object selectivity and tolerance (e.g., to position or size changes) both increase along the ventral visual stream (Kobatake and Tanaka, 1994; Logothetis and Sheinberg, 1996; Tanaka, 1996; Edelman, 1999; Rolls, 2000; Rousset et al., 2004). Nevertheless, however appealing, this notion cannot be taken as given; we do not yet know whether individual IT neurons attain high values of both selectivity and tolerance. More generally, we do not even know whether these properties are connected in any

Received April 26, 2007; revised Sept. 20, 2007; accepted Sept. 24, 2007.

This work was supported primarily by the National Institutes of Health. Additional support was provided by Defense Advanced Research Projects Agency, Office of Naval Research, The Pew Charitable Trusts, The McKnight Foundation, Eastman Kodak Company, Honda Research Institute, Siemens Corporate Research, Sony, and the McDermott chair (T.P.). D.Z. was supported by a Postdoctoral Fellowship of The International Human Frontier Science Program Organization and by a Charles A. King Trust Postdoctoral Research Fellowship. We thank J. Maunsell, N. Li, and N. Rust for comments on this manuscript; D. Cox, T. Serre, and G. Kreiman for discussion; C. Hung for help in stimulus preparation; and J. Deutsch for technical support.

Correspondence should be addressed to James J. DiCarlo, McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: dicarlo@mit.edu.

DOI:10.1523/JNEUROSCI.1897-07.2007

Copyright © 2007 Society for Neuroscience 0270-6474/07/2712292-16\$15.00/0

way, because no study has systematically measured both properties in the same IT neuronal population.

To address this issue, we systematically measured both object selectivity and tolerance to different identity-preserving image transformations within the same IT neuronal population in two monkey subjects engaged in a simple object detection task. For each neuron, we measured (1) how selectively it responded across a large object set (object selectivity) (see Fig. 1A, top); and (2) how well preserved its response was to a very effective reference object that underwent different transformations: position, size, and contrast changes and addition of visual clutter (absolute tolerance) (see Fig. 1A, bottom). Although another way to measure tolerance is to check how well relative object preference is preserved across transformations (relative tolerance) (Tovee et al., 1994; Ito et al., 1995), we focused here on absolute tolerance of neuronal responses because it provides a bound on other tolerance metrics and measures the ability of individual neurons to support recognition without further processing (see Discussion). Contrary to the appealing idea that IT contains neurons that are both highly shape selective and highly tolerant, we discovered that selectivity and tolerance trade off across the IT population: neurons with high object selectivity typically have relatively low tolerance, and vice versa.

## Materials and Methods

We used standard procedures for surgical preparation, behavioral task and training, eye position monitoring, and single-unit electrophysiological recording in awake monkeys, and details are described by Zoccolan et al. (2005). Here we briefly describe those methods that are most relevant to the present study. All animal procedures were performed in accord with National Institute of Health guidelines and the Massachusetts Institute of Technology Committee on Animal Care.

### Visual stimuli and behavioral task

All recorded neurons were probed with a fixed set of 213 grayscale pictures of isolated real-world objects, mostly modified from the Caltech 101 database (Fei-Fei et al., 2004), but including also (1) five fixed object prototypes from each of three spaces of parameterized objects (see below); (2) five patches of texture; (3) four low-contrast images of one of the objects; and (4) a blank frame (used to measure neuronal background rate). The full set is shown in supplemental Figure 1 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Some neurons were also tested using additional objects drawn from the three spaces [cars, faces, and two-dimensional (2-D) silhouettes] with parametrically controllable shape similarity within each space. Each object space consisted of 14 morph axes (for a total of 42 morph axes), and each morph axis was composed of five shapes resulting from blending two object prototypes (e.g., two car brands) in different proportions (see examples in Fig. 2B) (for details, see Zoccolan et al., 2005). All objects subtended  $\sim 2^\circ$ .

During recordings, both monkeys were engaged in a simple recognition task that required the detection of a fixed, red target shape (a red triangle) that was presented at the end of a temporal sequence of object conditions drawn from our stimulus set (see Fig. 1B). The total number of stimulus conditions presented on each behavioral trial ranged from 3 to 20. The target was always the last in the sequence, and each monkey was rewarded for maintaining fixation ( $\pm 1.5^\circ$  fixation window) until the appearance of the target and then making a saccade to a fixed visual field location ( $7^\circ$  eccentricity) within 800 ms after the appearance of the target.

The size of the fixation window was chosen to be small enough to guarantee that the monkeys could not make a saccade to any of the eccentric positions that were tested during the mapping of the receptive field (RF) (see below) without leaving the fixation window and, therefore, aborting the trial. This was guaranteed by the fact that the closest positions to fixation were  $\pm 2.5^\circ$  from the fixation spot (see below), i.e.,  $\pm 1^\circ$  beyond the edge of the fixation window.

Single objects from the fixed stimulus set and identity-preserving transformations of a very effective reference object (see Fig. 1A) were

pseudorandomly interleaved. Visual stimuli were presented at a rate of 5 per second; i.e., each stimulus condition was shown for 100 ms, followed by 100 ms of a gray screen (no stimulus), followed by another stimulus condition for 100 ms, etc. (see Fig. 1B). This task was meant to obtain a large amount of data, while still engaging the animal in a recognition task. We have previously shown that clutter suppression during such tasks is not simply explained by variation in spatial attention (Zoccolan et al., 2005).

### Neuronal recordings

During each recording session, a single extracellular metal electrode was advanced into IT. Over  $\sim 6$  months of daily recording sessions in the two monkeys, we sampled neurons over an  $\sim 5 \times 4$  mm area of the ventral superior temporal sulcus and ventral surface lateral to the anterior middle temporal sulcus (Horsey-Clark coordinates: anteroposterior, 13–17 mm; 18–21 mm mediolateral at recording depth).

**Screening procedures.** Each isolated neuron was tested for responsiveness across the fixed set of 213 stimuli plus 30 additional object prototypes belonging to the parameterized morphed spaces (Zoccolan et al., 2005) using a very inclusive criterion: a neuron was considered responsive if its mean firing rate was significantly higher than background rate for at least one of these stimuli ( $t$  test,  $p < 0.005$ , which implies  $p < 0.7$  corrected for multiple tests). All stimuli were presented at the center of gaze. Two to four presentation repetitions were collected for each stimulus. This screening procedure was used to identify a very effective reference object (the object that produced the strongest neuronal response, higher than background rate according to the  $t$  test) and six poorly effective flanking objects (the objects that produced the smallest response, not significantly higher than background rate according to the  $t$  test). Note, however, that, because of trial-by-trial noise, in the final testing in which more repetitions were used (see below), the chosen objects did not always turn out to be the most effective (and ineffective) objects. Only responsive neurons for which these reference and flanking objects could be found were selected for further testing and recordings. The screening procedure was also used to decide whether the neuron was responsive to any of the object prototypes belonging to the parameterized morphed spaces (a total of 45 prototypes were used, 15 of which belonged to the fixed set of 213 visual stimuli; see above).

If selected for recordings (see above), a neuron was further screened to identify its preferred receptive field location (RF center) within a narrow span of visual angle around the center of gaze. Our goal was not to map each neuron RF over the whole visual field, but rather to optimize the location in which to present our single object conditions during the main testing session (see below). To achieve this, the most effective (reference) object and the least effective object were presented over a span of  $2^\circ$  around the fixation spot (8–10 repetitions). More specifically, six visual field locations (in addition to fixation) were tested. Four of these locations were the extremes of a “cross” centered on fixation (i.e.,  $2^\circ$  above and below fixation and  $2^\circ$  left and right to fixation). The other two locations were  $\pm 2^\circ$  in elevation and  $2^\circ$  in azimuth, in the contralateral hemifield with respect to the recording chamber. The visual field location in which the response produced by the reference was higher was chosen as RF center of the neuron. Within this  $2^\circ$  span, most neurons (75/94) had their RF center at the center of gaze.

**Recording session.** Complete recordings were obtained from each neuron using pseudorandomly interleaved stimuli from our entire battery of stimulus conditions: (1) single objects belonging to the fixed object set and presented in the neuron’s RF center (see Fig. 1A, broad sampling); (2) identity-preserving transformations of the reference object to test clutter tolerance (CT), position tolerance (PT), size tolerance (ST), and contrast tolerance (CrT) (see Fig. 1A). If the neuron was responsive to any of the 45 morphed object prototypes that were probed during screening (see above), it was also tested with five objects belonging to the morph axis that included the effective object prototype (see Fig. 1A, local sampling).

Tolerance to position changes was assessed by mapping the response to the reference object across a vertical  $\sim 12^\circ$  span of visual field (see Fig. 3A). The reference object was presented in the RF center (see above) and in eight additional positions: (1)  $-2.5^\circ$  below the RF center; and (2)  $2.5^\circ$

3.5, 4.5, 5.5, 7.5, 9.5, and either 11 (73/94 cells) or 12° (21/94 cells) above the RF center. Position tolerance was not mapped in full 2-D space because it would have been too time consuming (see below). Based on results in the literature (Tovee et al., 1994; Ito et al., 1995; Op De Beeck and Vogels, 2000), our testing should provide a reasonable estimate of position tolerance, and we have no reason to believe that a full 2-D map of the RF field would lead to qualitatively different results.

Tolerance to size changes was measured by presenting the reference object at four different sizes (1, 2, 4, and 6°) at the RF center (see Fig. 3B). Object conditions to measure ST were tested for all neurons from monkey 2 ( $n = 34$ ) but only for a subset of neurons from monkey 1 (18/60).

Tolerance to contrast changes was assessed by presenting the reference object at three low contrasts (1.5, 2, and 3%), in addition to its default contrast (mean default contrast across reference objects  $\pm$  SD =  $33 \pm 12\%$ ), at 2.5° above the RF center (see Fig. 3C). The reason that contrast tolerance was not measured at the RF center is that the data presented in this study are part of a larger experimental design, in which, among other things, the relationship between clutter tolerance and the contrast of the flanker objects was tested. To measure such a relationship, flankers at different low contrasts were presented 2.5° above the RF center. To save time (see below), the same low-contrast conditions were not presented also at the RF center. Stimulus contrast was defined as follows:  $[\text{med}(\text{pix} > L_{\text{background}}) - \text{med}(\text{pix} < L_{\text{background}})] / [\text{med}(\text{pix} > L_{\text{background}}) + \text{med}(\text{pix} < L_{\text{background}})]$ , where med indicates the median taken over the pixel values either above or below the monitor background luminance level ( $L_{\text{background}}$ ).

To test clutter tolerance, the reference object was presented both in isolation and along with each of the six, poorly effective flanking objects (see above) (see Fig. 3D). Objects in the pairs used to test clutter tolerance were 2.5° apart (center to center), with the effective reference object in the neuron RF center and the flanking object located 2.5° either above or below the RF center.

As described above, given the time limitations of awake recording, only a limited amount of object conditions could be used to measure each tolerance property. Therefore, our goal was not to obtain the best possible estimate of each tolerance property per se, but rather to collect enough data to understand how these properties covaried with object selectivity and with each other.

The transformations of the reference object described above allow measuring the absolute tolerance of each neuron, i.e., how well preserved the response to the effective reference object across each transformation (see below for a definition of the tolerance metrics) is. Alternatively, tolerance can be defined in terms of how well preserved the rank order of object selectivity across the tested transformations (relative tolerance) is. Although our study was not meant to provide an accurate estimate of relative tolerance (see Introduction and Discussion), in addition to measuring how the response to the effective reference object changed by varying its position, size, and contrast, we also measured how these same transformations affected the response to a very ineffective object (one of the flanking objects). This allows a first-order assessment of relative tolerance across the recorded population (see below).

Five to thirty presentation repetitions were collected for each object condition (see supplemental material 1.B for details, available at www.jneurosci.org).

### Data analysis

**Choice of the spike count window.** To get the most statistical power from the data, average firing rates were computed over a time window (relative to stimulus onset) whose extremes were optimally chosen for each neuron by an apposite algorithm. Briefly, epochs of the neuronal response were identified in which the background corrected response was at least 20% of its peak value. If multiple such epochs existed, they were merged to the epoch containing the response peak if they were within 25 ms from it. This algorithm is completely described in supplemental material 1.A and supplemental Figure 2 (available at www.jneurosci.org as supplemental material). Analyses performed using either such neuron-specific spike count windows (mean onset across neurons  $\pm$  SD =  $101 \pm 18$  ms; mean offset =  $236 \pm 48$  ms) or count windows that were held fixed across the recorded neuronal population yielded very similar results (supple-

mental Tables 1, 2, available at www.jneurosci.org as supplemental material). The neuron-specific spike count windows were used to estimate latency and duration of the neuronal responses in the analysis shown in supplemental Table 6 (available at www.jneurosci.org as supplemental material). For such analysis, onset and offset of the response were computed considering epochs in which the background corrected response was at least either 20% (as for our standard spike count window) or 10% (for a better comparison with previous studies) (Brincat and Connor, 2006) of its peak value.

To assess the dependency of the selectivity and tolerance properties (see below) from the time course of the neuronal response, firing rates were also computed in overlapping time windows of 50 ms shifted in time steps of 25 ms (see Fig. 6).

Analyses were performed using absolute firing rates, but very similar results were obtained when driven (i.e., “background subtracted”) rates were used (supplemental Tables 1, 2, available at www.jneurosci.org as supplemental material).

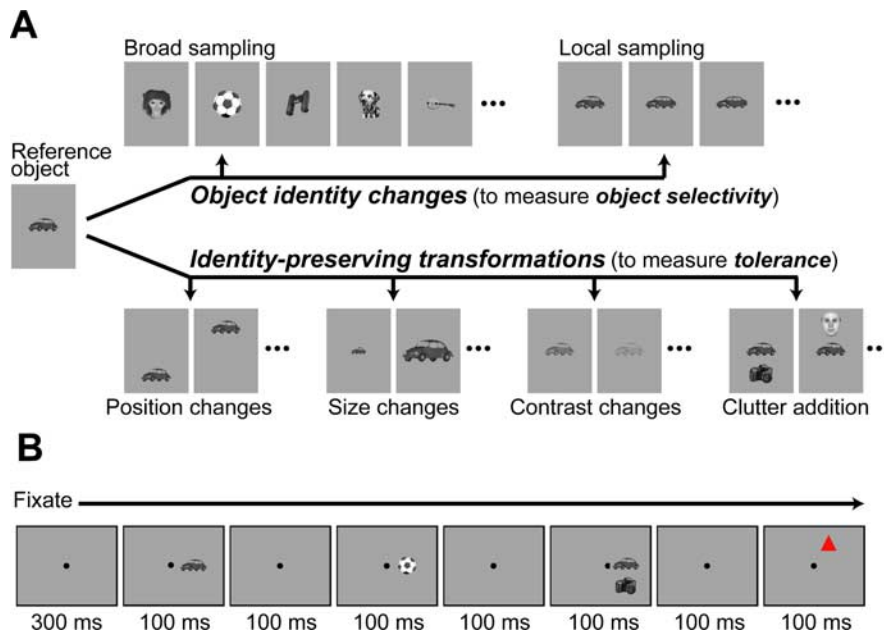
**Selectivity and tolerance metrics.** The selectivity of each neuron across the fixed set of 213 stimuli was quantified by the sparseness of its response (Rolls and Tovee, 1995a; Vinje and Gallant, 2000; Olshausen and Field, 2004):  $S = \{1 - [(\sum R_i/n)^2 / \sum (R_i^2/n)]\} / [1 - (1/n)]$ , where  $R_i$  is the neuron response to the  $i$ th stimulus and  $n$  is the number of stimuli in the set.  $S$  ranges from 0 (no object selectivity) to 1 (maximal object selectivity). Neuronal selectivity within each morphed set was quantified by the following morph tuning index (Rainer et al., 1998):  $MT = [n - (\sum R_i/R_{\text{max}})] / (n - 1)$ , where  $R_i$  is the neuron response to the  $i$ th morphed object,  $R_{\text{max}}$  is the maximal response within the morphed set, and  $n$  is the number of objects in the set. As sparseness, MT ranges from 0 (no shape selectivity) to 1 (maximal shape selectivity).

For each neuron, PT was defined as two times the SD of the Gaussian function that best fitted the normalized driven rates across the tested RF positions. Operationally, we subtracted the background activity from the neuron responses across the tested RF positions, and we normalized (1.0) the resulting RF profile. Then, we fit a Gaussian function to the RF profile, with the peak of the Gaussian centered on the peak of the RF. As a goodness-of-fit measure, we used the sum of the absolute residuals SR. Only neurons such that  $SR < 1.5$  were included in the analyses presented throughout the paper. For the analysis shown in Figure 5C (top left), only neurons such that  $SR < 0.8$  were included in the population-averaged RF profiles, to guarantee that only RFs with homogeneous shapes (i.e., all strictly Gaussian) contributed to the average. Note that in Figure 5C (see legend), before being averaged, the receptive field profiles were aligned to the position (elevation) of their peak values. Because the tested elevations were not located in an equally spaced grid (see above), aligning the peaks produces a misalignment of the elevations at which each neuron was tested. Thus the RF profiles in Figure 5C were averaged in overlapping windows of  $\sim 3^\circ$ , shifted in steps of  $\sim 1^\circ$ . Depending on what elevations fell in a given averaging window, the average elevation in that window may be different between subpopulations.

ST was quantified by normalizing (1.0) the size tuning curves to their maximal values (see Fig. 3B, bottom) and then averaging the resulting tuning curve values that were  $< 1$ , i.e.,  $ST = \langle R_{\text{test size}} / \max(R_{\text{test size}}) \rangle$ , where  $R_{\text{test size}}$  is the mean response to a given tested size of the reference object, and  $\langle \cdot \rangle$  is the average over the sizes such that  $R_{\text{test size}} < \max(R_{\text{test size}})$ . Contrast tolerance (CrT) was similarly measured, but after normalizing each contrast tuning curve to the original contrast of the reference object (see Fig. 3C, bottom, leftmost point), i.e.,  $CrT = \langle R_{\text{test contrast}} / R_{\text{ref contrast}} \rangle$ , where  $R_{\text{test contrast}}$  is the mean response to a given tested contrast of the reference object,  $R_{\text{ref contrast}}$  is the mean response to the reference object presented at its original contrast, and  $\langle \cdot \rangle$  is the average over the tested contrasts (3, 2, and 1.5%) (see Fig. 3D). ST and CrT measure the amount of response that is preserved, on average, when the reference object is transformed by changing, respectively, its optimal size or its reference contrast. Both metrics range between 0 (no tolerance) and 1 (complete invariance).

We quantified the clutter tolerance of each neuron with the following clutter tolerance (CT) metric:  $CT = \langle (R_{\text{ref \& flanker}} - R_{\text{flanker}}) / (R_{\text{ref}} - R_{\text{flanker}}) \rangle$ , where  $R_{\text{ref \& flanker}}$  is the mean response to a given pair reference/flanker,  $R_{\text{ref}}$  and  $R_{\text{flanker}}$  are the mean responses to the constituent





**Figure 1.** Rationale of the experimental design and behavioral task. **A**, For each neuron, we measured (1) its object selectivity, i.e., its sensitivity to changes in object identity (measured both across a large set of real-world objects, and, when possible, across local, parameterized sets of morphed shapes); and (2) its tolerance to different identity-preserving image transformations of an effective reference object. **B**, During recordings, monkeys were presented with rapid sequences of pseudorandomly interleaved grayscale objects used to measure selectivity and tolerance (see above). The monkey's task was to respond to the red triangle at the end of each sequence. The number of objects presented before the triangle was random (between 3 and 20).

objects of that pair (i.e., the reference and flanker objects presented alone), and  $\langle \cdot \rangle$  is the average over the included flanker conditions (for most analyses, only flankers such that  $R_{\text{flanker}} < 50\% R_{\text{ref}}$  or less were considered) (but see Fig. 7A). CT values near 0 indicate very poor clutter tolerance (strong suppression by flanker objects), whereas values near 1 indicate complete clutter invariance (CCI; no effect of flanker objects).  $CT > 1$  indicates facilitation by clutter. It is important to note that, because the response to the isolated flanker objects is subtracted from both  $R_{\text{ref} \& \text{flanker}}$  and  $R_{\text{ref}}$  in the definition of CT, the metric is unbiased with respect to either a CCI or a clutter averaging rule (Zoccolan et al., 2005) (see Fig. 7A for details).

Significance and SE of the correlation coefficients reported through this manuscript were computed, respectively, using a one-tailed permutation test (5000 samples) and bootstrap analysis (500 samples).

Presentation repetitions of the reference object were split to guarantee complete independence of the datasets over which selectivity and tolerance metrics were computed (see supplemental material 1.B, available at [www.jneurosci.org](http://www.jneurosci.org)). The number of neurons included in the different analyses presented throughout the paper (see Figs. 4–7; supplemental Tables 1–3, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) was selected according to specific criteria that are fully described in supplemental material 1.C (available at [www.jneurosci.org](http://www.jneurosci.org)).

To obtain a measure of the relative tolerance of object selectivity across the tested transformations, for each of the position, size, and contrast changes, we computed the difference between the response (average firing rate) to the effective reference object ( $R_{\text{ref}}$ ) and the response to one of the poorly effective flanker objects ( $R_{\text{flanker}}$ ), normalized by the mean of the standard deviations ( $\sigma_{\text{ref}}$  and  $\sigma_{\text{flanker}}$ ) of the two response rates. In other words, we computed a  $d'$  index, defined as follows:  $d' = (R_{\text{ref}} - R_{\text{flanker}})/\sigma$ , with  $\sigma = (\sigma_{\text{ref}} + \sigma_{\text{flanker}})/2$ . This index tells how far apart the responses to the effective and ineffective objects are. Negative values indicate that, for a given transformation, the tuning of the neuron has reversed (i.e., the ineffective flanker produced a response larger than the effective reference).

**Spike waveform analysis.** Putative inhibitory and excitatory neurons were identified across the recorded neuronal population based on the waveform of the action potentials collected for each cell. Previous studies

performed in cortical slices and anesthetized animals (McCormick et al., 1985; Connors and Gutnick, 1990; Nowak et al., 2003) and additional studies based on cross-correlation analysis of spike trains (Bartho et al., 2004; Tamura et al., 2004) have shown the reliability of such an approach in identifying neuronal typing. As a consequence, many authors have distinguished between putative excitatory and inhibitory neurons recorded in different brain areas of awake animals (such as hippocampus, somatosensory cortex, lower and higher visual cortical areas, and rat barrel cortex) using only the temporal features of the spike waveforms (Mountcastle et al., 1969; Gur et al., 1999; Frank et al., 2001; Bruno and Simons, 2002; Andermann et al., 2004; Mitchell et al., 2007). In this study, following Bruno and Simons (2002), two components of the waveform were taken as distinctive features of the neuron type: (1) the duration (width) of the central peak (negative or positive); and (2) the width of the following trough (corresponding to the spike afterhyperpolarization). Figure 5A shows a scatter plot with the mean width of the peak and trough for our population of 94 neurons. The cluster of neurons with shorter mean widths (empty circles) was identified as putative inhibitory. The cluster with longer mean widths (filled circles) was designated as excitatory. Waveforms of one example neuron from each cluster are shown in Figure 5A (top).

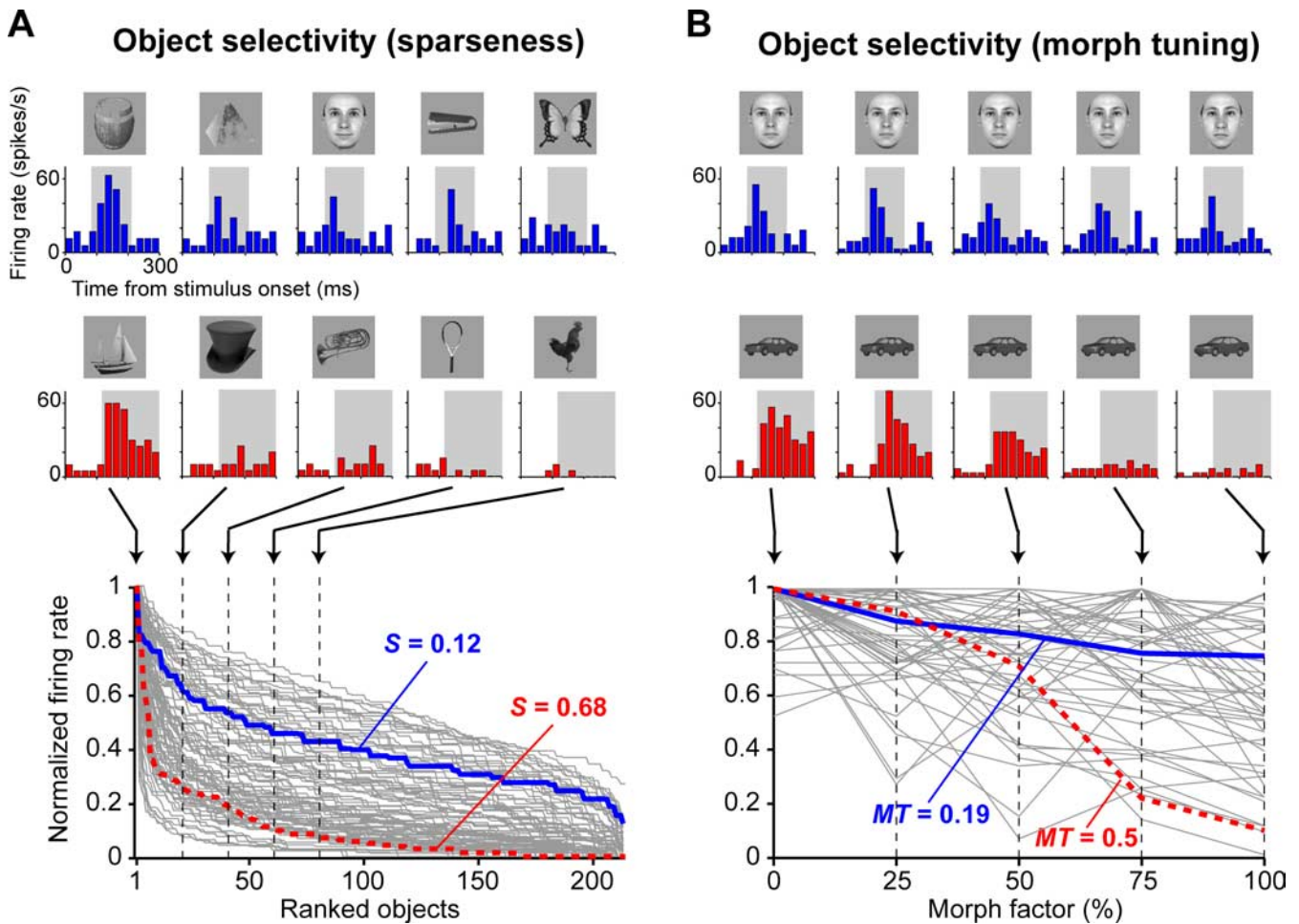
**Monte Carlo simulations to check that the sparseness and clutter tolerance metrics are not implicitly correlated.** Monte Carlo simulations were run to check that the inverse relationship between  $S$  and CT (see above) could not arise by chance as the result of an implicit correlation between these metrics computed over a population of noisy neuronal responses (see Fig. 7B). For any given neuron, responses to the single object conditions were simulated using Poisson spike generators with mean rate equal to the actual mean firing rate that was recorded for each single object condition (the same number of repetitions collected for each object condition were simulated). The selectivity (sparseness,  $S$ ) of each neuron was computed based on these Poisson simulated responses to single objects. A simulated response to clutter (i.e., to the tested object pairs) was implemented either assuming an averaging rule or assuming CCI. Two versions of the averaging rule were implemented: (1) one exact, i.e.,  $R_{\text{ref} \& \text{flanker}} = 0.5 (R_{\text{ref}} + R_{\text{flanker}})$ ; (2) the other approximate, i.e., with  $R_{\text{ref} \& \text{flanker}}$  randomly sampled between  $R_{\text{flanker}}$  and  $R_{\text{ref}}$  (in both cases, Poisson statistics of spike trains were assumed). The CT of each neuron was computed based on these Poisson simulated responses to object pairs. Each Monte Carlo simulation was repeated 500 times, yielding the null distributions of regression lines and correlation coefficients shown in Figure 7B.

#### Model simulations

The hierarchical object recognition model used to generate Figure 9E is fully described in previous reports (Riesenhuber and Poggio, 1999; Serre et al., 2005, 2007a,b). See supplemental material 2 and supplemental Figure 7 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) for details about model simulations and additional modeling results.

## Results

To examine the relationship between object selectivity and tolerance in IT, we performed extracellular microelectrode recordings in two monkeys that viewed grayscale images of real-world objects presented at a rate of 5 images/s, while the animals were engaged in a simple object detection task (Fig. 1B) (see Materials and Methods). Well isolated neurons were randomly sampled



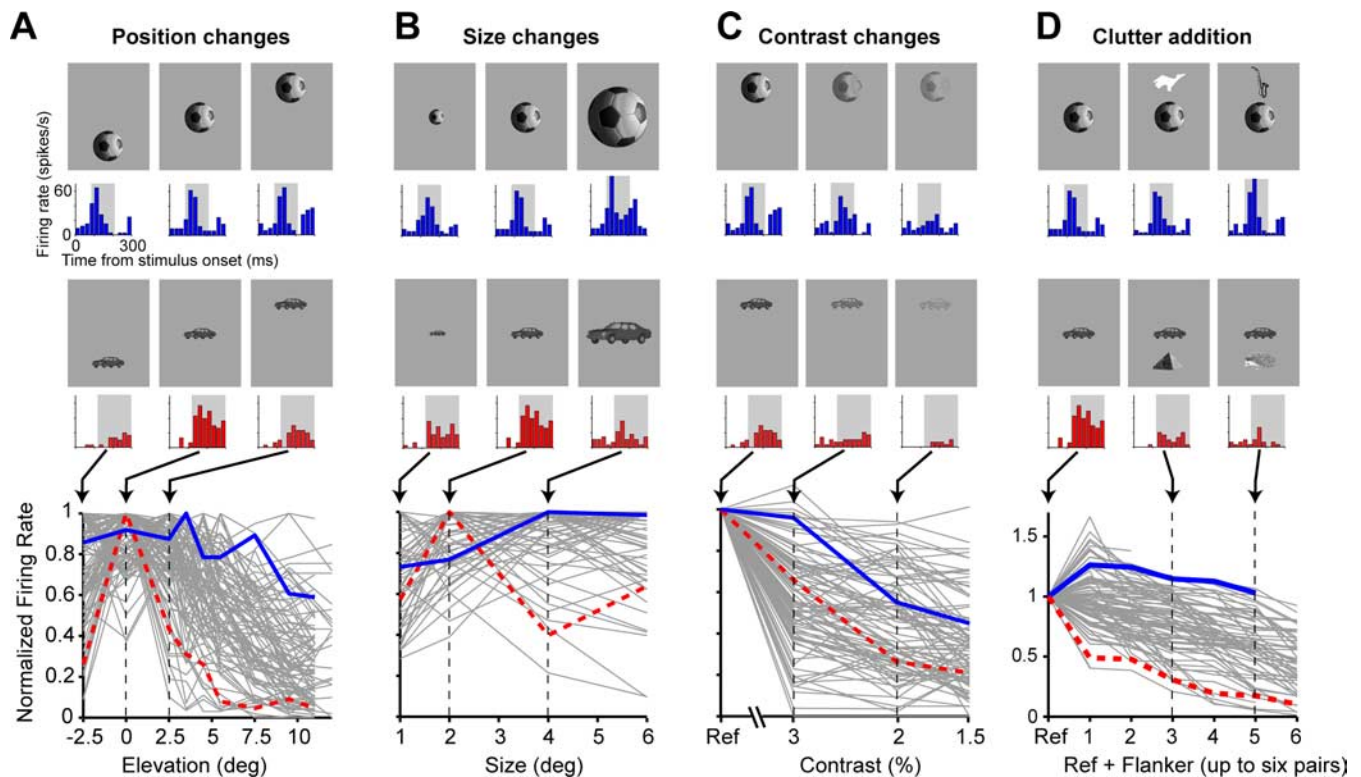
**Figure 2.** Broadness of object selectivity in IT. *A*, Normalized firing rate profiles (bottom) across a fixed set of 213 objects for a population of 94 neurons. For each neuron, objects in abscissa are ranked based on the mean response they evoked. The figure highlights data from two example neurons (blue and red curves) and their responses (blue and red PSTHs) to five objects chosen at equally spaced intervals along the abscissa (dashed lines). Gray boxes indicate the spike count window. *B*, Normalized tuning profiles (bottom) across five parametrically morphed objects (49 cells; each cell was tested using one of 42 possible morphed object sets). The same example neurons and color code are used as in *A*.  $S$  and  $MT$  values are indicated for the example cells.

throughout anterior IT and tested for responsiveness across a fixed, large set of visual objects (see Materials and Methods). Each responsive neuron was then tested with a battery of object conditions to measure (1) its object selectivity, i.e., its sensitivity to changes in object identity (Fig. 1*A*, top); and (2) its tolerance to different identity-preserving transformations of a reference object, including position, size and contrast changes, and presence of clutter (Fig. 1*A*, bottom). Complete recordings were obtained from 94 IT neurons (60 in monkey 1, 34 in monkey 2). We took special care to obtain independent data and design independent metrics for selectivity and tolerance so as to guarantee no implicit relationship among these properties (see Materials and Methods and supplemental material 1*B*, available at [www.jneurosci.org](http://www.jneurosci.org)).

### Broad range of object selectivity across the IT neuronal population

To get a first-order measure of each neuron's object selectivity, we estimated the fraction of objects in a large, fixed set of 213 stimuli (see supplemental Fig. 1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material, Fig. 2*A*) that produced a response (sparseness, described immediately below). For most neurons (see Materials and Methods), we also measured shape selectivity within several predefined parameterized object shape spaces (Fig. 2*B*, morph tuning, described in detail later in Results) (see also

Zoccolan et al., 2005). Both methods uncovered a remarkably broad spectrum of sensitivity to shape changes within IT. For example, although many neurons responded strongly to their preferred object within the fixed shape set (population mean  $\pm$  SD =  $49.5 \pm 21.2$  spikes/s) (see supplemental Fig. 5, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material), the population was highly varied in the number of objects that elicited a strong response. This can be visually appreciated by ranking, for each neuron, the 213 test stimuli based on the response they evoked (Fig. 2*A*, bottom). For example, some neurons were weakly selective in that they showed a strong response to many objects [Fig. 2*A*, blue curve and peristimulus time histograms (PSTHs)]. Other neurons were highly selective in that they responded well to only a handful of objects (Fig. 2*A*, red curve and PSTHs). These example neurons illustrate the broad range of selectivity seen across the IT population (Fig. 2*A*, gray lines). Note however, that even the least-selective neurons within our IT population typically showed some (nonzero) object selectivity in that some objects elicited little or no response (Fig. 2*A*). In fact, all neurons that fired at least 10 spikes/s to the effective reference object (see Materials and Methods) and were included in further analyses (91/94 cells), responded significantly more to the reference object than to one of the weakly effective flanker objects chosen during



**Figure 3.** Broad ranges of tolerance properties in IT. Normalized tuning curves produced by four different identity-preserving transformations of neuron-specific reference objects are shown for the recorded neuronal population (gray) and for two example neurons (red and blue; same neurons and color code as in Fig. 2). Image transformations include changes in object position (**A**), size (**B**), and contrast (**C**), as well as addition of clutter (**D**). For the two example neurons, some raw neuronal responses (PSTHs) to the tested transformations of the reference object (a soccer ball and car, respectively) are also shown in matching colors. The contrast in **C** is for display purpose only and does not match the actual contrast of stimuli used during recordings.

the screening procedure (see Materials and Methods; one-tailed  $t$  test,  $p < 0.05$ ).

We quantified each neuron's selectivity across the stimulus set by the sparseness (Rolls and Tovee, 1995a; Vinje and Gallant, 2000; Olshausen and Field, 2004) of its response ( $0 < S < 1$ ; see Materials and Methods), which is a well established metric to quantify the fraction of stimuli in a given stimulus set that produce a response. A value of  $S$  near 0 indicates that a neuron responds nearly equally to many objects in the stimulus set (low object selectivity), whereas a value near 1 indicates that a neuron responds well to only a few objects (high object selectivity). The sparseness distribution across the neuronal population extended over a very broad range of values (from 0.05 to 0.94; mean  $\pm$  SD =  $0.4 \pm 0.22$ ,  $n = 94$ ) (see Fig. 5B). This systematic quantification of the object selectivity of each IT neuron allowed us to look across the population for any relationship between object selectivity and tolerance to identity-preserving transformations.

### Broad ranges of tolerance properties across the IT neuronal population

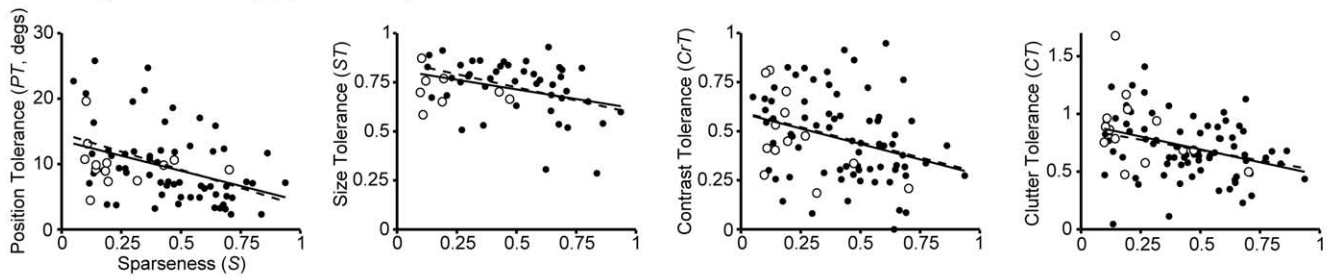
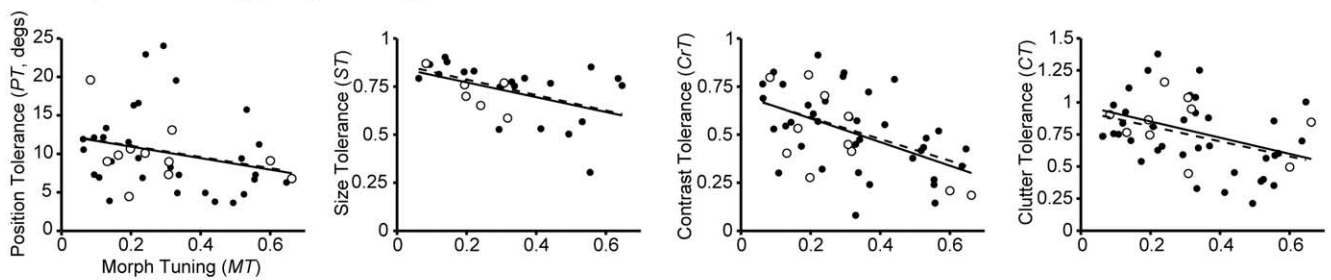
To quantify each neuron's tolerance to identity-preserving image transformations of its highly preferred objects, we measured its change in firing rate in response to identity-preserving transformations of a reference object. This reference object was chosen for each neuron from among the objects that most effectively drove the cell during a screening procedure preceding the recording session (see Materials and Methods). The tested image transformations of the reference object were pseudorandomly interleaved with the testing of selectivity (Fig. 1B), and they included changes in object position and size (position and size tolerance),

changes in object contrast (contrast tolerance), and the addition of other objects (a test of tolerance to visual clutter) (Fig. 1A).

Tolerance to position changes was assessed by mapping the response to the reference object across a vertical  $12^\circ$  span of visual field (Fig. 3A). Tolerance to size changes was measured by presenting the reference object at four different sizes (1, 2, 4, and  $6^\circ$ ) at the RF center (Fig. 3B). Tolerance to contrast changes was assessed by presenting the reference object at three low contrasts (1.5, 2, and 3%), in addition to its default contrast (mean default contrast across reference objects  $\pm$  SD =  $33 \pm 12\%$ ), at  $2.5^\circ$  above the RF center (Fig. 3C). As a first-order test of clutter tolerance, the reference object was presented both in isolation and along with a single, poorly effective flanker (clutter) object (Fig. 3D). Six such flanker objects were tested for each neuron. They were chosen, from the fixed set of 213 stimuli, among those objects that least effectively drove the neuron (see Materials and Methods). Only those flankers that produced little or no response ( $< 50\%$  of the response to the reference object) were included in further analysis (but see also Fig. 7A).

Like object selectivity (Fig. 2), we found that IT neurons varied greatly in their amount of tolerance to identity-preserving image transformations. This can be visually appreciated in Figure 3 (bottom panels), which plots the normalized responses of each neuron across each set of transformations of the reference object. For some IT neurons, the response to the reference object was only minimally altered when its position or size were varied, its contrast was lowered, or flanker objects were added, resulting in broad, slow-varying response profiles across the transformation axes (Fig. 3A–D, blue curves and PSTHs). Other neurons were much less tolerant to these same image transformations; their



**A Object selectivity (sparseness) vs. tolerance****B Object selectivity (morph tuning) vs. tolerance****C**

Object selectivity measure	Position tolerance	Size tolerance	Contrast tolerance	Clutter tolerance
<b>Sparseness</b> ( $R_{\text{ref}} \geq 10$ spikes/s)	<b>-0.39***</b> $\pm$ 0.09 (77)	<b>-0.33**</b> $\pm$ 0.13 (52)	<b>-0.32**</b> $\pm$ 0.1 (85)	<b>-0.35***</b> $\pm$ 0.09 (88)
<b>Sparseness</b> ( $R_{\text{ref}} \geq 30$ spikes/s)	<b>-0.36**</b> $\pm$ 0.11 (56)	<b>-0.36**</b> $\pm$ 0.17 (42)	<b>-0.30*</b> $\pm$ 0.11 (60)	<b>-0.46***</b> $\pm$ 0.08 (63)
<b>Morph tuning</b> ( $R_{\text{ref}} \geq 10$ spikes/s)	<b>-0.26*</b> $\pm$ 0.12 (41)	<b>-0.47**</b> $\pm$ 0.16 (26)	<b>-0.52***</b> $\pm$ 0.09 (48)	<b>-0.38**</b> $\pm$ 0.09 (48)
<b>Morph tuning</b> ( $R_{\text{ref}} \geq 20$ spikes/s)	<b>-0.29*</b> $\pm$ 0.1 (38)	<b>-0.43*</b> $\pm$ 0.16 (25)	<b>-0.52***</b> $\pm$ 0.1 (45)	<b>-0.32*</b> $\pm$ 0.1 (45)

**Figure 4.** Trade-off between object selectivity and tolerance to identity-preserving transformations in IT. **A**, The scatter plots show the inverse relationship between sparseness (as a measure of object selectivity; see Fig. 2A) and each of the tested tolerance properties. Open and filled circles refer, respectively, to putative inhibitory and excitatory neurons, according to the analysis shown in Figure 5A. Regression lines through all data points (solid) and only putative excitatory neurons (dashed) are also shown. **B**, The scatter plots show the inverse relationship between shape selectivity measured within a set of parametrically morphed objects (see Fig. 2B) and each of the tested tolerance properties. The same symbols as in **A** are used. Neurons that fired at least 10 spikes/s to the reference object were included in the plots shown in **A** and **B**. **C**, Correlation coefficients  $\pm$  SE between object selectivity (either sparseness or morph tuning) and each of the tolerance properties ( $*p \leq 0.05$ ;  $**p \leq 0.01$ ;  $***p \leq 0.001$ ; one-tailed permutation test; SE computed by bootstrap). Subsets of neurons with different levels of minimal response to the effective reference object ( $R_{\text{ref}}$ ) are considered. Only flanker object conditions in which  $R_{\text{flanker}} < 0.5 R_{\text{ref}}$  contributed to the clutter tolerance metric. The number of neurons contributing to each correlation is reported in parentheses.

response was drastically reduced by altering the position, size, and contrast of the reference object and by adding clutter shapes (Fig. 3A–D, red curves and PSTHs). These two example neurons (Fig. 3, blue and red) illustrate the broad ranges of tolerances seen across the IT population (Fig. 3, bottom panels, spread of gray lines).

We quantified the tolerance to position changes of a neuron (PT) by computing the size (in degrees) of its RF. This was done by fitting a Gaussian function to the RF profile and by taking twice the SD of the fitted Gaussian as a measure of the diameter of the neuron's RF (see Materials and Methods for details). To quantify the tolerance to the other identity-preserving transformations, we used a different approach. Under the premise that the neuron's response signals the presence of a preferred object, the mean decrease in neuronal response caused by size or contrast changes or by addition of clutter objects was taken to be an inverse measure of the neuron's tolerance to each of these transformations. Equations defining the ST, CrT, and CT metrics are provided in Materials and Methods, and important controls are provided for the clutter tolerance metric later in the Results. For each of these metrics, values near 0 indicate very poor tolerance (i.e., strong response reduction caused by the corresponding im-

age transformation), whereas values near 1 indicate maximal tolerance (i.e., "invariance"; the corresponding image transformation does not reduce the response to the reference object). ST and CrT ranges between 0 and 1, whereas CT can assume values  $> 1$  if the response to some of the object pairs (reference and flanker objects shown together) is higher than the response to the isolated reference object (see Materials and Methods). We found that all four tolerance metrics spanned a broad range of values across the population (see the spread of points along the ordinate axes in Fig. 4A).

**Trade-off between object selectivity and tolerance to identity-preserving image transformations**

By obtaining independent, quantitative measures of both object selectivity and tolerance to identity-preserving transformations for each IT neuron (above), we could examine their relationship. We found that object selectivity and tolerance were negatively correlated across the anterior IT population (Fig. 4A). That is, whereas high tolerance values were typically found in weakly shape-selective neurons (left-hand side of the sparseness axis), tolerance became proportionally lower for more sharply shape-selective neurons, dropping to very small values for some of the

most shape-selective neurons (right-hand side of the sparseness axis). The correlations between sparseness (shape selectivity) and each of the tolerance properties were all negative (range:  $-0.3$  to  $-0.46$ ) and significant (Fig. 4C, first two rows) (one-tailed permutation test). These negative relationships did not depend on how well the neurons responded to the reference objects used to measure their tolerance properties (two neuronal subpopulations with response to the reference object  $R_{ref} > 10$  or 30 spikes/s were considered in Fig. 4C). In sum, the highest levels of shape selectivity and tolerance observed in IT neurons are not both found in individual IT neurons. Instead, selectivity and tolerance trade off across the IT population, suggesting that individual IT neurons gain object selectivity only at the expense of tolerance, and vice versa.

This trade-off result cannot be explained by differences in background firing rate across the population. First, the trade-off did not crucially depend on whether raw or driven firing rates (i.e., background corrected rates) were used to compute the tolerance metrics (supplemental Tables 1, 2, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Second, when both the sparseness and the tolerance metrics were computed after subtracting the minimal response across the 213 stimuli (which includes the “background” blank image; see Materials and Methods), we found inverse correlations nearly identical to those obtained using raw rates (see supplemental Table 4 and supplemental Fig. 6, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

As expected, given the negative correlation between selectivity and tolerance, the pairwise correlations between each tolerance property were positive, although not large and not always significant (see supplemental Table 5, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). This weaker correlation may reflect a poor estimate of each tolerance property, given the relative small number of object conditions that were tested to estimate each property (see Materials and Methods), or, instead, may indicate that tolerance properties are built, at least at some extent, independently from one another along the ventral stream (Riesenhuber and Poggio, 1999; Serre et al., 2005, 2007b).

### Does the trade-off depend on how object selectivity is determined?

Because object selectivity might be defined in many ways, we wanted to see whether the trade-off result depended on our particular choice of object selectivity metric or object test set. First, in addition to sparseness (above), we considered a number of different selectivity metrics computed on the responses to the set of 213 stimuli. In each case, we found the same result: a negative correlation between the object selectivity metric and all four types of tolerance (see supplemental Table 3, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

We also considered the possibility that the inverse relationship between sparseness and some of the tolerance properties was the result of measuring selectivity over a set of objects that differed in low-level visual properties (i.e., nonshape properties), such as area and contrast. To test whether the trade-off between selectivity and size (contrast) tolerance was an artifact of highly selective neurons being more sensitive to area (contrast) variations over the stimulus set, we measured how well each neuronal response profile correlated with variation in these low-level properties across the object test set. We considered the neuronal subpopulations within the first third (38 poorly selective cells) and the last third (17 highly selective cells) of the sparseness range (for additional details, see Fig. 5) and we compared the average sensitivity

of each neuronal response to stimulus area and contrast (i.e., the average correlation between the property and the neuronal response). On average, the responses of poorly selective neurons were positively (but very weakly) correlated with stimulus area (average correlation =  $0.05 \pm 0.04$  SE) and contrast (average correlation =  $0.08 \pm 0.02$  SE). A still weaker and negative correlation was observed between the responses of highly selective neurons and stimulus area (average correlation =  $-0.02 \pm 0.05$  SE) and contrast (average correlation =  $-0.03 \pm 0.01$  SE). Thus, IT neuronal responses are only minimally affected by area and contrast of the tested objects. Even more importantly, poorly selective neurons are more sensitive than highly selective neurons to these low-level properties over the object set, which would tend to make poorly selective neurons less tolerant to object size and contrast (by definition), the opposite of the trade-off we observed. This conclusion was confirmed by comparing the squares of the correlation coefficients ( $r^2$ , explained variance) between neuronal responses and low-level stimulus properties for the two populations of highly and weakly selective cells. On average, the amount of variance of the neuronal response explained by variations of stimulus contrast and area was small and larger for weakly than highly selective cells (weakly selective cells:  $r^2_{area} = 0.075 \pm 0.015$  SE;  $r^2_{contrast} = 0.023 \pm 0.005$  SE; highly selective cells:  $r^2_{area} = 0.039 \pm 0.022$  SE;  $r^2_{contrast} = 0.003 \pm 0.001$ ). Overall, this rules out the possibility that the trade-off between object selectivity and size/contrast tolerance is produced by variations in neuronal sensitivity to these low-level properties (also see next).

Finally, we also tested a subpopulation of 49 IT neurons using additional sets of test objects and an associated selectivity metric. Specifically, each of these neurons showed a response to any of 45 objects belonging to three sets of parameterized shapes [cars, faces, and two-dimensional silhouettes; see Materials and Methods and Zoccolan et al. (2005)] and could therefore be tested for tuning along a continuous shape dimension (morph axis) that included the effective object (see Materials and Methods). Each of the 49 neurons fired at least 10 spikes/s to the most effective object within the tested morph axis. Selectivity within each morph axis (five continuously morphed shapes per morph axis) (see examples in Fig. 2B) was quantified by a morph tuning index [MT; see Materials and Methods and Rainer et al. (1998)]. MT ranges from 0 (the neuron responds equally to every shape along the morph axis) to 1 (the neuron responds only to one shape). Morph tuning and sparseness ( $S$ , above) provide two complementary measures of neuronal selectivity for visual objects. Whereas sparseness quantifies neuronal responsiveness across a broad set of natural objects (that may vary in global shape, number and complexity of features and textures, and low-level visual properties such as area, luminance, and contrast), morph tuning quantifies neuronal sensitivity to small, controlled shape transformations of an effective object (Fig. 1A). Therefore, the morph tuning allowed us to assess the relationship between almost “pure” shape selectivity and tolerance properties, independent of potential confounds of low-level stimulus properties such as area and contrast.

As shown in the top of Figure 2B (same example cells and color code as in Fig. 2A), the sensitivity to small shape changes of the preferred prototype (a face and a car, respectively, for the two example neurons) can be widely different for neurons within anterior IT. In fact, like sparseness, morph tuning spanned a large range of values across the recorded neuronal population (from 0.06 to 0.66; mean =  $0.31 \pm 0.17$  SD,  $n = 49$ ), and it was well correlated with sparseness ( $r = 0.51 \pm 0.12$  SEM,  $p = 0.0002$ ,  $n =$

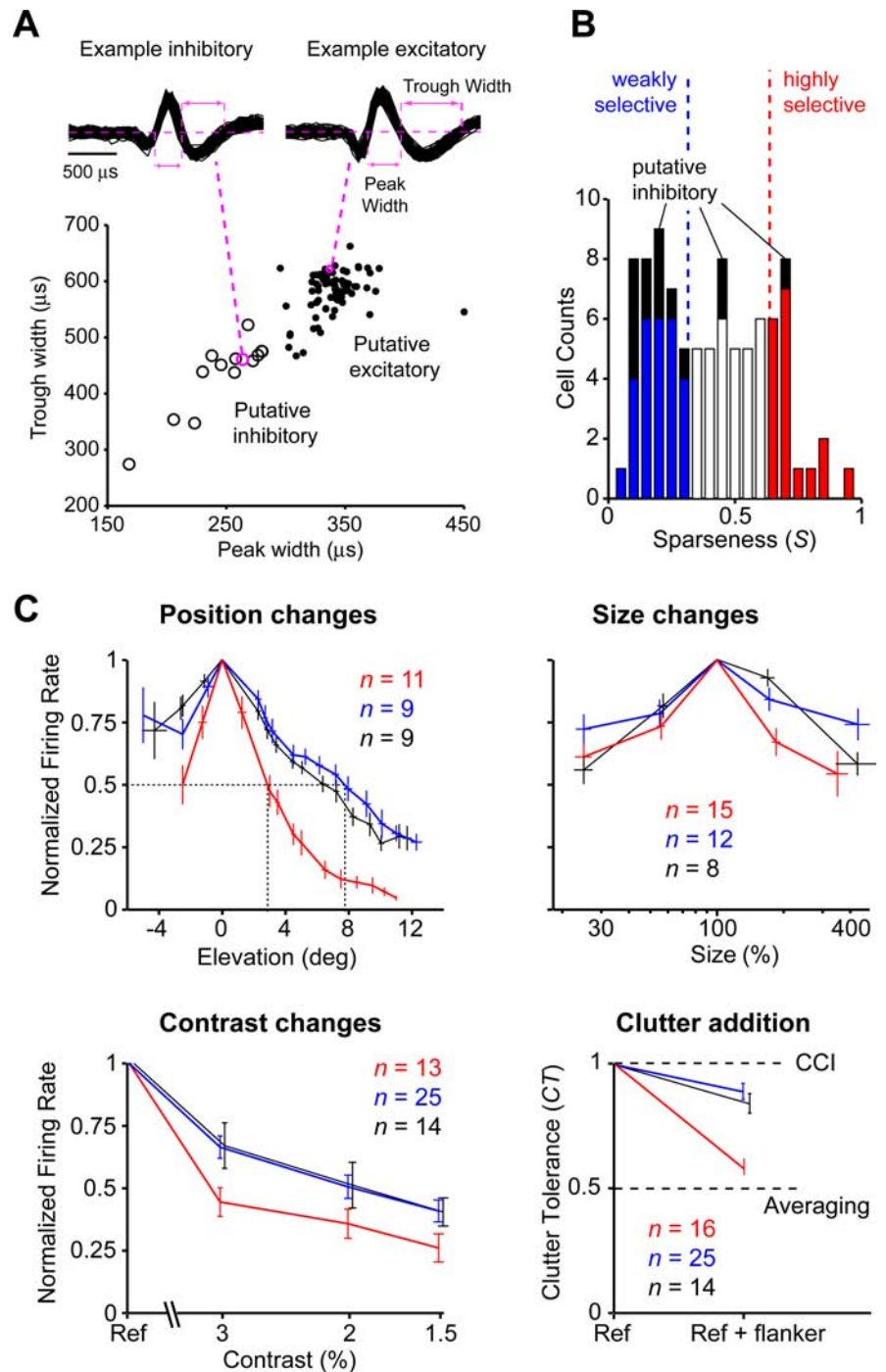


49, one-tailed permutation test), suggesting that both measures tap into each neuron's underlying shape selectivity. Moreover, like the sparseness metric, the morph tuning reveals a broad range of variation across the recorded IT population and allows an assessment of its correlation with the tolerance properties. Similar to what we found for sparseness, we also observed a trade-off between morph tuning and each of the four tolerance properties (Fig. 4B): morph tuning was negatively and significantly correlated with all the tolerance metrics (Fig. 4C, last two rows).

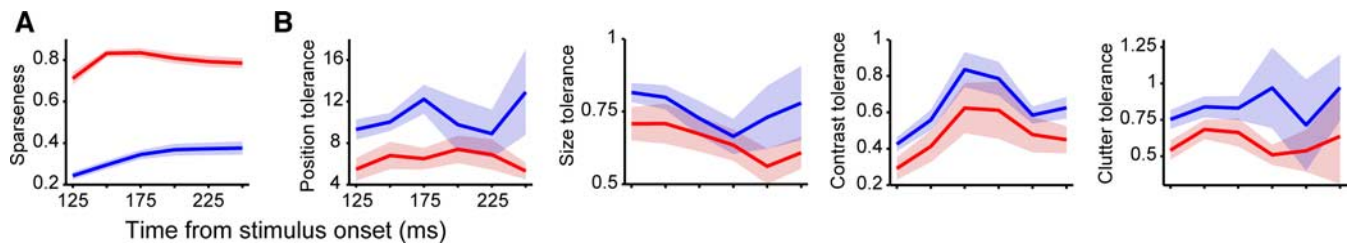
Although these results do not allow us to claim that we have precisely measured the shape tuning of any individual IT neuron, they show that the uncovered trade-off between object selectivity and tolerance across the IT population is highly robust to the metric and stimulus set used to quantify object selectivity and holds also when selectivity to almost pure shape changes is considered. Together, the results shown in Figure 4 and supplemental Table 3 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) show the existence of a trade-off between object selectivity and a wide range of tolerances to identity-preserving image transformations in IT.

#### Contribution of putative inhibitory and excitatory neurons to the observed trade-off between selectivity and tolerance

We considered the possibility that the observed trade-off between object selectivity and clutter tolerance was caused by a mixture of excitatory and inhibitory neurons in the population (e.g., perhaps with inhibitory neurons being less shape selective). To examine this, the recorded neuronal population was divided into putative inhibitory and excitatory neurons based on extracellular measures of excitatory and inhibitory neuronal typing suggested previously (Mountcastle et al., 1969; Gur et al., 1999; Frank et al., 2001; Bruno and Simons, 2002; Constantinidis and Goldman-Rakic, 2002; Swadlow, 2003; Andermann et al., 2004; Hasenstaub et al., 2005; Mitchell et al., 2007) (Fig. 5A). As indicated in the scatter plots of Figure 4, A and B, inhibitory neurons had a marked tendency to be both less shape selective than excitatory neurons (lower sparseness and morph tuning values) (see also Fig. 5B, black bars) and more tolerant to identity-preserving image transformations. Although the putative inhibitory population is small, it showed a trend consistent with the trade-off between selectivity and tolerance observed over the whole neuronal



**Figure 5.** Population-averaged tolerance profiles for inhibitory and for highly and weakly shape-selective excitatory neurons. **A**, Two components of the action potentials produced by a given cell were taken as distinctive features of the neuron type (top): (1) the duration (width) of the central peak; and (2) the width of the trailing trough. Across the recorded population, the cluster of neurons with shorter mean widths (open circles) was designated as putative inhibitory. The cluster with longer mean widths (filled circles) was designated as excitatory. **B**, Distribution of sparseness values observed across the neuronal population. Blue and red bars show, respectively, putative excitatory neurons in the first and last third of the sparseness range. Black bars show putative inhibitory neurons. **C**, Population-averaged tolerance profiles for excitatory weakly shape-selective (in blue) and highly shape-selective (in red) neurons and inhibitory neurons (in black), as defined in **B**. Before averaging, position and size tuning curves were aligned to the location of their peak values. For position, neuronal responses were averaged in overlapping windows of  $\sim 3^\circ$ , shifted in steps of  $\sim 1^\circ$  (only curves that were best fitted by Gaussian functions were averaged; see Materials and Methods). Size tuning curves were plotted as a function of the percentage of size change with respect to the most effective object size (100%) and then averaged in four nonoverlapping windows (approximately equally spaced on a logarithmic scale). For clutter tolerance, CT values in the first (in blue) and last (in red) third of the sparseness range were averaged. The small displacement of back curves in the bottom panels is for display purpose only.



**Figure 6.** Time course of object selectivity (sparseness) and tolerance properties for the two populations of weakly and highly selective neurons. **A, B.** Sparseness and tolerance properties were computed in spike count windows (time slices) of 50 ms, shifted in steps of 25 ms. In each time slice, the average value (solid line) and the SE (shaded regions) of the sparseness and tolerance properties were computed for the weakly selective neurons (blue;  $n = 38$ ) and the highly selective neurons (red;  $n = 17$ ) defined in Figure 5 (putative inhibitory neurons were included in the two sets). For each time slice, outliers were removed before averaging (i.e., tolerance metric values larger than 95th percentile in the time slice).

population. More importantly, the full spectrum of selectivity and tolerance values was observed even among only putative excitatory neurons (see the spread of filled circles in Fig. 4*A, B*), as well as a significant negative correlation between selectivity and each tolerance (Fig. 4*A, B*) (correlation values computed using only excitatory neurons are reported in supplemental Tables 1, 2, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). In summary, the observed trade-off between selectivity and tolerance does not depend on the putative subtype of the neurons whose responses were recorded (see also Fig. 5*C*, population average plots).

#### Population-averaged tolerance profiles for highly and weakly shape-selective IT neurons

To further examine the dependence of the amount of tolerance on each neuron's object selectivity, we considered the least shape-selective neurons and the most shape selective neurons in our population. Specifically, we focused on the neuronal subpopulation within the first third ( $0.05 < S < 0.35$ ) (Fig. 5*B*, blue bars) and the last third ( $0.64 < S < 0.94$ ) (Fig. 5*B*, red bars) of the observed sparseness range. For clarity, putative inhibitory neurons were excluded from both subpopulations and analyzed as a third, separate subpopulation (Fig. 5*B*, black bars), but results were nearly identical if this was not done (see below). The normalized neuronal responses across each set of transformations of the reference object (Fig. 3, bottom) were aligned according to their peak values (see legend for details) and averaged within each of the three neuronal subpopulations, yielding the population-averaged tolerance profiles shown in Figure 5*C* for weakly shape-selective (in blue), highly shape-selective (in red), and inhibitory neurons (in black).

This analysis confirmed that weakly shape-selective excitatory neurons were, on average, more tolerant than highly shape-selective excitatory neurons to each of the four identity-preserving transformations (Fig. 5*C*, compare blue and red curves). It also shows that the amount of tolerance of putative inhibitory neurons was nearly identical to that of the weakly shape-selective excitatory neurons (Fig. 5*C*, compare black and blue). This means that the difference in the amount of tolerance between the populations of weakly and highly selective neurons is not affected by whether the putative inhibitory neurons are treated as a separate population or not.

Figure 5 also allows estimating the difference in the average amount of tolerance in weakly versus highly shape-selective neurons. For example, the average RF width (measured at one-half the peak response) (Fig. 5*C*, top left, dashed lines) was more than twice as large for weakly shape-selective neurons (relative to highly shape-selective neurons). Similarly, whereas weakly shape-selective neurons showed, on average, almost no effect of

clutter (CCI;  $CT = 1$ ), highly shape-selective neurons showed strong suppression by clutter ( $CT \sim 0.5$ ). This  $CT$  value means that, for more highly shape-selective IT neurons, the response to a pair of simultaneously presented reference and flanker objects ( $R_{ref \& flanker}$ ) is close to the average of the responses to the reference ( $R_{ref}$ ) and flanker ( $R_{flanker}$ ) objects presented in isolation, which is consistent with the averaging rule reported in some recent studies (Zoccolan et al., 2005; De Baene et al., 2007) (described further below).

#### Latency of neuronal responses and time course of selectivity and tolerance properties

We asked whether the subpopulations of weakly and highly selective IT neurons show any significant difference in the latency and duration of their responses. In agreement with previous findings (Brincat and Connor, 2006), we found a weak but significant positive correlation between object selectivity (i.e., sparseness) and (1) latency of response onset and (2) response duration (see supplemental Table 6*A*, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). On average, weakly selective neurons fired  $\sim 10$  ms before highly selective cells, and their responses were  $\sim 30$  ms shorter (see supplemental Table 6*B*, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Although the latency of the peak of the response had a tendency to be longer for highly selective neurons than weakly selective neurons, this difference was not significant (see supplemental Table 6*B*, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material), and no significant correlation was found between peak latency and sparseness (see supplemental Table 6*A*, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

Previous studies reported that object selectivity can substantially change (e.g., increase) as a function of time during the response epoch (Sugase et al., 1999; Matsumoto et al., 2005; Brincat and Connor, 2006). A detailed analysis of the information about object identity conveyed by different temporal epochs of the neuronal response was not the primary goal of this study. However, by measuring the time course of object selectivity as a function of time (using spike count windows of 50 ms that overlapped of 25 ms) for the two subpopulations of weakly and highly selective neurons defined in Figure 5*B*, we found that sparseness was remarkably stable for the duration of the IT responses (Fig. 6*A*). That is, IT neurons that are very object selective in their initial response tend to remain very object selective, and vice versa. A similar analysis of the time course of the tolerance metrics showed that, despite some modulations as a function of time, highly selective neurons were consistently less tolerant than weakly selective cells for the whole duration of the neuronal response (Fig. 6*B*). Overall, these analyses show that selectivity and tolerance metrics are largely independent of the spike count win-

dow (see also supplemental Tables 1, 2, available at www.jneurosci.org as supplemental material) and that the trade-off between selectivity and tolerance holds for the duration of the response.

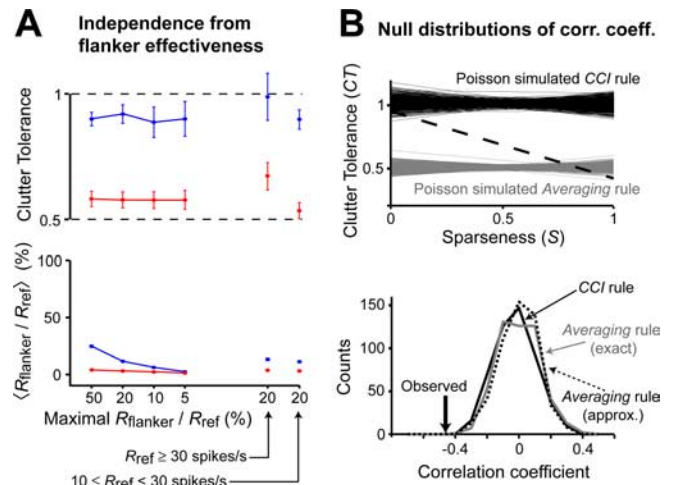
### Relative tolerance of object selectivity

Above, we measured tolerance in absolute terms. Although not the goal of our study, we also collected data that partially address the issue of how well preserved the rank order of object selectivity across different transformations (positions, sizes, and contrasts) is, independent of absolute response rate (i.e., relative tolerance). In particular, for each neuron, we asked how many times the response to the poorly effective object became higher than the response to the effective object, i.e., how many times the object preference of any given neuron reversed over our range of tested identity-preserving transformations (position, size, and contrast). This was quantified by computing a  $d'$  index (see Materials and Methods) that measures, for each transformation (e.g., a given position or size), how far apart the responses to the two objects are. A negative  $d'$  indicates that a given transformation produced a reversal in the object preference of the neuronal response. Such reversal happened rarely: (1) for position changes, it happened 15% (7%) of the times for the subpopulation of highly (weakly) selective neurons; (2) for size changes, it happened 0% (4%) of the times for the subpopulation of highly (weakly) selective neurons; (3) for contrast changes, the reversal happened 23% (14%) of the times for the subpopulation of highly (weakly) selective neurons. Interestingly, even when reversals occurred, they were typically small: the average of the reversed  $d'$  for highly selective cells was (1)  $-0.60 \pm 0.33$  (SD) for position changes; and (2)  $-0.64 \pm 0.50$  for contrast changes. The average of the reversed  $d'$  for weakly selective cells was (1)  $-0.31 \pm 0.28$  for position changes; (2)  $-0.24 \pm 0.17$  for size changes; and (3)  $-0.40 \pm 0.27$  for contrast changes (for comparison, the average  $d'$  in the reference position was 3.97). This suggests that the reversal of object preference typically happened when the response to the effective object became as small as the response to the poorly effective object (e.g., at the edges of the receptive fields or for the lowest contrasts), suggesting that the reversal was driven by the variability of two nearly identical neuronal responses, rather than by a true change of object preference. As expected based on our measurements of absolute tolerance (above), such reversals happened more often for the subset of very selective neurons, given the smaller size of their receptive fields and their lower contrast tolerance, compared with the poorly selective cells.

### A deeper look into clutter tolerance

Clutter tolerance differs from the other three tolerances tested here because the image changes that one might expect IT neurons to be tolerant to are less well defined. For example, size change has only one degree of freedom, but there is an unlimited number of distractor (clutter) objects that one might add to the visual scene. Considering the overarching goal of the present study (examining the relationship of shape selectivity and tolerance), the most important consideration was to be sure that our choices of flanker (clutter) objects did not produce artifactual dependency between shape selectivity and clutter tolerance.

First, we asked whether the observed strong tendency for weakly shape-selective IT neurons to be less suppressed by flanker objects could be explained by a tendency of those neurons to be more driven by the tested flanker objects when presented in isolation. Indeed, this tendency was found in our data (Fig. 7A,

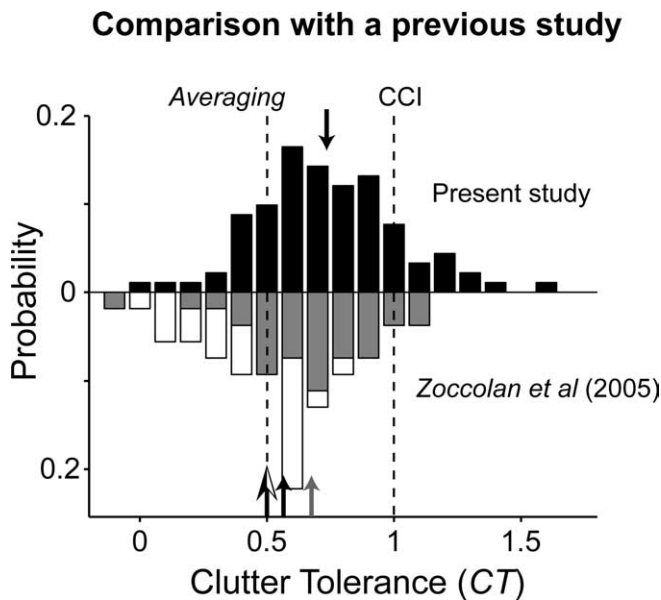


**Figure 7.** Additional controls on the relationship between clutter tolerance and object selectivity. **A**, Average CT (top plot) and average effectiveness of the flanker objects relative to the reference objects (bottom plot) for the two subpopulations of weakly (in blue) and highly (in red) shape-selective neurons (as defined in Fig. 5B) as a function of the maximal response evoked by flanker objects ( $R_{flanker}$ ) used to compute CT ( $R_{flanker}$  ranges from 50 to 5% of the response evoked by the reference object,  $R_{ref}$ ).  $R_{ref} \geq 10$  spikes/s for the four leftmost points in the plot. The two rightmost points refer to subsets of neurons with lower ( $10 \leq R_{ref} < 30$  spikes/s) and higher ( $R_{ref} \geq 30$  spikes/s) effectiveness of the reference object and  $R_{flanker} < 20\% R_{ref}$ . **B**, Relationship between sparseness and clutter tolerance obtained from Monte Carlo simulations of two populations of neurons following different clutter rules: CCI and averaging. Each simulation produced a regression line and the regression lines from 500 such simulations are shown in black (CCI) and gray (averaging) in the top panel, together with the regression line through the observed data (dashed). The distributions of correlation coefficients obtained from those simulations are shown at the bottom, together with the correlation coefficient observed in the neuronal data (thick arrow). Two different versions of the averaging rule were implemented (see Materials and Methods).

leftmost data points, described further below). However, our clutter tolerance metric (CT; see Materials and Methods) was designed to account for these differences in flanker effectiveness, because the response to the flanker (chosen to be poorly effective in isolation) is subtracted in both the CT numerator and denominator (see Materials and Methods). That is, CT is 1.0 when  $R_{ref} \& \text{flanker} = R_{ref}$  (CCI), and CT is 0.5 when  $R_{ref} \& \text{flanker} = 0.5 (R_{ref} + R_{flanker})$  (averaging rule), independent of the effectiveness of the isolated flanker ( $R_{flanker}$ ). To show this lack of bias directly, we performed Monte Carlo simulations with populations of the same size and response rate distribution as our neuronal data (see Materials and Methods for details). These simulations (Fig. 7B) showed that the observed inverse relationship between sparseness and clutter tolerance could not arise ( $p < 0.002$ , one-tailed test) from an implicit correlation between these metrics computed over a population of noisy neuronal responses that were all equally tolerant to clutter [i.e., either all completely clutter invariant (CCI rule) or all equally suppressed by clutter (averaging rule)]; see Materials and Methods for details].

Thus, despite possible variations in flanker effectiveness, our CT metric remains unbiased with respect to either a CCI or an averaging rule in clutter. This can be directly appreciated by comparing the amount of clutter tolerance in the two subpopulations of weakly and highly selective IT neurons (Fig. 5B) as we gradually restricted our analysis to flanker objects with increasingly lower effectiveness (Fig. 7A). As already indicated for our standard analysis (above), flanker objects were, on average, slightly more effective for weakly shape-selective than for highly shape-selective neurons (Fig. 7A, bottom, compare blue and red





**Figure 8.** Distributions of clutter tolerance values obtained in this and a previous study (Zoccolan et al., 2005). Distributions of CT values obtained for the neuronal populations recorded in the present study (black bars) and from Experiments 1 (gray bars) and 2 (white bars) of Zoccolan et al. (2005). The arrows show the mean CT for the neuronal populations recorded in (1) the present study (top black); (2) the previous study (bottom black); and (3) Experiment 1 of the previous study (bottom gray). The big black and white arrow (bottom) shows the mean CT for the very selective neurons that were analyzed in Figure 4 of Zoccolan et al. (2005).

curves). Of course, by construction this difference in flanker effectiveness disappeared as the analysis was restricted to flanker objects that were less and less effective (Fig. 7A, bottom, rightmost points). Nevertheless, the difference in the amount of clutter tolerance in the two neuronal subpopulations was largely unaffected by this manipulation (Fig. 7A, top, compare blue and red curves). That is, regardless of flanker effectiveness, the weakly selective IT neurons tended to show high flanker (clutter) tolerance ( $CT \sim 1$ ), whereas the highly shape-selective IT neurons showed stronger clutter suppression ( $CT \sim 0.5$ ). Furthermore, we also checked that this result did not depend on the effectiveness of the reference object (Fig. 7A, rightmost points).

Finally, we also verified that the inverse relationship between object selectivity and clutter tolerance did not arise from any bias in the low-level visual properties of flanker objects (such as contrast) (see supplemental Fig. 3, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) or flanker object identity (see supplemental Fig. 4, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) used to test weakly versus highly shape selective neurons. In summary, IT neurons that are intrinsically the most selective for object shape are typically the least tolerant to the addition of other poorly effective (clutter) objects in the visual field, and this result is not simply attributable to the choice of clutter objects.

The observed trade-off between selectivity and clutter tolerance implies that the clutter tolerance observed in a recorded population of IT neurons depend on the underlying distribution of object selectivity in that population. For example, most CT values in this study (Fig. 8, black bars) ranged between 0.5 (averaging) and 1 (CCI), with a mean  $CT = 0.74$  (top black arrow). In a previous study (Zoccolan et al., 2005), we reported that most IT neurons ( $\sim 67\%$ ) had clutter tolerance that was consistent with an averaging rule (whereas only  $\sim 12\%$  were closer to CCI, and another  $\sim 20\%$  in between CCI and averaging rule). Although the

distributions of clutter tolerance are broad in both of these studies (Fig. 8A, compare black to gray and white bars) [but see also Zoccolan et al. (2005), Fig. 4B], they are significantly different (Kolmogorov–Smirnov test,  $p = 0.01$ ), with the previous study distribution skewed toward an average rule (mean  $CT = 0.57$ ) (Fig. 8, bottom black arrow). This difference in the distribution of clutter tolerance is likely explained by a difference in the distribution of object selectivity resulting from a difference in neuronal selection criteria. Specifically, in the present study, neurons were collected with very inclusive criteria (see Materials and Methods). In our previous study, approximately one-half of the analyzed neurons were recorded with similarly inclusive criteria [Zoccolan et al. (2005), Experiment 1] (Fig. 8, gray bars), but the other half were recorded only if very shape selective [Zoccolan et al. (2005), Experiment 2] (Fig. 8, white bars), resulting in an overall neuronal population that was enriched in shape-selective neurons. In fact, when only neurons of Experiment 1 in Zoccolan et al. (2005) were considered (gray bars), no significant difference was observed between the clutter tolerance distributions in the two studies (Fig. 8, compare gray to black bars) (Kolmogorov–Smirnov tests,  $p = 0.7$ ; mean  $CT = 0.67$  shown by the bottom gray arrow), showing the consistency between the studies. It should also be noted that, although both studies required focusing on neurons in which noneffective “clutter” objects could be found, the fact that we had many more objects available in the current study allowed us to examine clutter tolerance in weakly shape-selective neurons that would have been overlooked in our previous study. In sum, our current study confirms a tendency of ventral visual stream neurons to be suppressed by clutter (Sato, 1989; Miller et al., 1993; Rolls and Tovee, 1995b; Chelazzi et al., 1998; Missal et al., 1999; Reynolds et al., 1999; Sheinberg and Logothetis, 2001; Reynolds and Chelazzi, 2004; Zoccolan et al., 2005; De Baene et al., 2007), but it points to variation in object selectivity as a key factor that can explain a range of clutter tolerance found within and across studies.

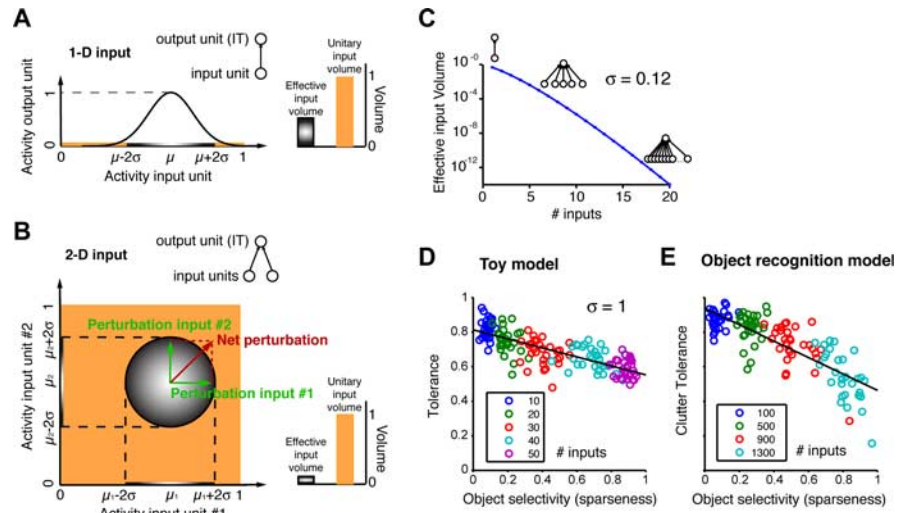
#### Possible neuronal mechanisms underlying the trade-off

The experimental results presented above show that object selectivity and tolerance are related in a potentially nonintuitive way: high object selectivity does not guarantee tolerance, but comes at the expense of tolerance. What family of neuronal mechanisms and architectures are consistent with the observed trade-off? Although a resolution of this issue is beyond the scope of a single study, as a first step toward understanding the neuronal computations underlying this trade-off, we considered basic computational mechanisms that might be at work. Our goal was not to obtain a quantitative comparison or a fit with neuronal data, but rather testing the consistency of class of tuning mechanisms and hierarchical architectures with the main trend observed in the data (i.e., the trade-off between selectivity and tolerance).

To do this, we used two computational models that employ a key tuning mechanism assumed by some object recognition models (Fukushima, 1980; Perrett and Oram, 1993; Wallis and Rolls, 1997; Riesenhuber and Poggio, 1999, 2000). The basic computational idea is that neurons are thought to attain selectivity for increasingly complex visual patterns by combining feature-conveying afferents in an AND-like manner. A convenient implementation of this is provided by a multidimensional Gaussian tuning function (Riesenhuber and Poggio, 1999) over a space of continuous-value inputs (i.e., activation values of afferent neurons). Each neuron is thus tuned to give its maximal response only when all of its inputs approximately match the neuron’s preferred input pattern (i.e., the center of the Gaussian).

We simulated responses of such toy model neurons over abstract input spaces of different dimensionality (see examples in Fig. 9A,B). When the sensitivity to each input dimension (i.e., the standard deviation  $\sigma$  of the Gaussian function) is held constant, the volume of total input space covered by the Gaussian (Fig. 9A,B, dashed gray areas) shrinks approximately exponentially with increasing number of inputs (i.e., input dimensionality) (Fig. 9C). Because this volume represents the chance that a random input pattern will drive the simulated neuron, the number of input patterns (i.e., arbitrary shapes) that drive the neuron decreases with the number of inputs, and thus object selectivity (sparseness) increases (Fig. 9D). To model tolerance, we assumed that any image transformation causes each input to have an equal, independent chance of being perturbed from its preferred value. Consistent with our experimental findings, this simple toy model results in the poorest tolerance for neurons with the highest object selectivity (Fig. 9D) because, on average, there is greater net perturbation of the input pattern from its optimum (Fig. 9B). This trade-off is also obtained when the number of inputs is held constant, but the SD of the Gaussian is allowed to vary.

The toy model likely captures the essence of the mechanisms underlying the observed trade-off of selectivity and tolerance. However, the ventral stream leading up to IT is clearly more complex than this toy model, and it is far from clear that mechanisms that try to build both selectivity and tolerance would still lead to the same predictions in IT. Thus, to go beyond the toy model, we also applied an existing object recognition model (Serre et al., 2005, 2007b), which reflects the basic architecture and physiology of the ventral visual stream, supports tolerant identification and categorization of visual objects and can make predictions over the exact same visual images tested experimentally. This model uses layers of neural units in a feedforward hierarchy, tuned to produce a gradual buildup of selectivity (AND-like operations) and tolerance (OR-like operations) (Fukushima, 1980; Perrett and Oram, 1993; Wallis and Rolls, 1997; Riesenhuber and Poggio, 1999, 2000). To compare any relationship between object selectivity and tolerance in this model with our experimental results, we used the same experimental stimuli as model input, recorded from model IT neurons, and analyzed responses using the same metrics (see supplemental Fig. 6, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). We found that variation in the number of afferents to the model IT neurons resulted in a clear trade-off between selectivity and tolerance within the model IT neuronal population, similar to both the toy model (Fig. 9D) and the experimental results (Fig. 4). This is shown in Figure 9E for the relationship between selectivity (sparseness) and clutter tolerance. The trade-off between selectivity and the other tolerance properties observed for the model units is shown in supplemental Figure 7 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). This trade-off was also obtained when the number of afferents was kept constant within the population of simulated IT units, but the sensitivity to each input dimension (i.e., the width of the Gaussian functions)



**Figure 9.** Modeling of the trade-off between selectivity and tolerance. **A, B**, Examples of toy model neurons with AND-like Gaussian tuning over 1-D (**A**) and 2-D (**B**) input spaces. Within the input space (in orange), the volume (a length in the 1-D case) occupied by the region within  $2\sigma$  from the Gaussian center (dashed gray; effective input volume) is much smaller for a simulated neuron with two afferents than for a neuron with only one afferent (compare dashed gray bars in **A** and **B**, right). **B** also shows how independent and identical perturbations of each inputs (green arrows) sum to produce a resulting net perturbation (red arrow) that reaches beyond the  $2\sigma$  radius around the center of the Gaussian. **C**, The effective volume shrinks as function of the number of inputs. **D**, For a toy model neuron, sparseness increases and tolerance decreases with the number of inputs, resulting in an inverse relationship between the two metrics. **E**, Model units at the last stage of a hierarchical model of object recognition (Serre et al., 2005) show a trade-off between sparseness and clutter tolerance, as the number of inputs is varied.

was allowed to vary, in such a way to obtain a range of broadness of tuning functions.

In summary, the observed trade-off between object selectivity and tolerance across the IT population can be naturally explained by hierarchical neuronal architectures in which (1) AND-like tuning operations are used to build object selectivity; and (2) the amount of object selectivity varies across the population, as a consequence of either a variable number of afferents (inputs) from earlier levels and/or variable sensitivity to each input dimension.

## Discussion

Some of the most remarkable findings about anterior IT concern the highly selective neuronal tuning for complex visual patterns reported by some authors (Desimone et al., 1984; Logothetis et al., 1995; Op de Beeck et al., 2001; Brincat and Connor, 2004; Freedman et al., 2005; Brincat and Connor, 2006) and the very high tolerance to stimulus transformations reported by others (Gross et al., 1972; Desimone and Gross, 1979; Perrett et al., 1982; Rolls and Baylis, 1986; Tovee et al., 1994; Rolls, 2000). Remarkably however, to our knowledge, no previous study has investigated the possible connection between these properties in IT. In this study, by systematically measuring both object selectivity and tolerance to different identity-preserving image transformations in the same IT neuronal population, we showed that IT neurons that attained the highest object selectivity were, on average, the least tolerant to identity-preserving image transformations and vice versa. That is, selectivity and tolerance trade off within IT.

The observed trade-off is not a trivial consequence of the definitions of object selectivity and tolerance; these are two very distinct types of image changes, and they have very different perceptual consequences. The result is also not an artifact of analysis methods; independent data were always used for each measure and extensive control analyses were conducted (Figs. 4–8; see supplemental material 1 and supplemental Figs. 3–5, available at

www.jneurosci.org as supplemental material). The trade-off cannot be accounted for by a differential sensitivity of highly versus weakly selective neurons to variations in low-level properties of the tested objects (see Results). We also showed that the result is robust to the way in which object selectivity and tolerance are measured, and it holds for all types of tolerance we tested (position, size, contrast, and clutter) (Figs. 4–6; supplemental Tables 1–4, available at www.jneurosci.org as supplemental material). In addition, we found that object selectivity, tolerance, and their inverse relationship are largely independent of the temporal epoch of the neuronal response (Fig. 6), suggesting that they are largely established as feedforward properties of the processing of visual objects (at least for the behavioral task and stimulus presentation rate tested here).

### Comparison with previous findings

This trade-off between selectivity and tolerance is consistent with previous work showing that selectivity and tolerance both grow along the ventral visual stream, to reach peak values in anterior IT (Kobatake and Tanaka, 1994; Logothetis and Sheinberg, 1996; Tanaka, 1996; Edelman, 1999; Rolls, 2000; Rousselet et al., 2004). Indeed, we suspect that the mean level of object selectivity and the mean level of tolerance are higher in IT than in lower ventral visual areas. However, our experiments indicate that, although single IT neurons are capable of achieving either very high selectivity or very high tolerance, they do not manage to attain both at once. This empirical discovery has several important implications and connections with previous work.

The present result helps unify a body of previous IT work by explaining why the amount of tolerance is highly varied within and across those studies (Gross et al., 1972; Desimone and Gross, 1979; Perrett et al., 1982; Desimone et al., 1984; Rolls and Baylis, 1986; Sato, 1989; Lueschow et al., 1994; Tovee et al., 1994; Ito et al., 1995; Logothetis et al., 1995; Logothetis and Sheinberg, 1996; Tanaka, 1996; Chelazzi et al., 1998; Edelman, 1999; Op De Beeck and Vogels, 2000; Rolls, 2000; Sheinberg and Logothetis, 2001; DiCarlo and Maunsell, 2003; Rousselet et al., 2004). Specifically, our results show that, because each IT neuron's tolerance depends on its object selectivity, the wide variation in object selectivity across the IT population (Fig. 5B) might explain the wide variation in tolerance. For instance, the amount of clutter tolerance of neuronal responses along the ventral visual stream is the subject of debate. Although many studies found that neuronal responses in IT and earlier visual areas (V4 and V2) are often strongly suppressed by clutter (Sato, 1989; Miller et al., 1993; Rolls and Tovee, 1995b; Chelazzi et al., 1998; Missal et al., 1999; Reynolds et al., 1999; Sheinberg and Logothetis, 2001; Reynolds and Chelazzi, 2004; Zoccolan et al., 2005; De Baene et al., 2007), most of these studies found that some fraction of neurons were much more tolerant to clutter (Chelazzi et al., 1998; Reynolds et al., 1999; Zoccolan et al., 2005), sometimes a large fraction (Sato, 1989; Sheinberg and Logothetis, 2001) or a majority (Gawne and Martin, 2002) of neurons. Our results suggest that, because the amount of clutter tolerance of any given IT neuron depends strongly on its object selectivity, all of these answers are correct; there is no single amount of clutter tolerance for all ventral stream neurons.

### Possible neuronal mechanisms underlying the trade-off

An exhaustive examination of the neuronal mechanisms underlying the observed trade-off is beyond the scope of the present work. However, using both a toy model and a hierarchical model of object recognition (Fig. 9), we showed the consistency between

the observed trade-off and neuronal tuning mechanisms (AND-like or Gaussian tuning) that are often assumed to be at the base of neuronal selectivity for visual objects (Fukushima, 1980; Perrett and Oram, 1993; Wallis and Rolls, 1997; Riesenhuber and Poggio, 1999, 2000). Because AND-like, approximately Gaussian tuning functions can be implemented through divisive normalization (Poggio and Bizzi, 2004), these simulations show the consistency between the observed trade-off and divisive normalization models (Heeger, 1992; Carandini et al., 1997; Recanzone et al., 1997; Britten and Heuer, 1999; Reynolds et al., 1999; Schwartz and Simoncelli, 2001; Reynolds and Chelazzi, 2004). Interestingly, although the hierarchical recognition model also employs OR-like (MAX) pooling operations to build tolerance to position and size changes, it still reveals an IT trade-off between object selectivity and tolerance to position and size variation. Future computational work may illuminate the family of mechanisms and architectures that lead to the observed trade-off.

### Generality and limitations of the present study

The main strength of our results is that a significant inverse relationship between object selectivity and tolerance was consistently observed between each tolerance property and each selectivity metric (Fig. 4). Importantly, we did not attempt to uncover what combination of visual features each neuron responded to. Rather, we aimed to estimate the sensitivity of each IT neuron to object identity changes (object selectivity) by measuring what fraction of objects, in two complementary sets, evoked a response in any neuron. Admittedly, however, object selectivity as measured here is not a perfect predictor of tolerance at the level of individual neurons (Fig. 4). This might reflect a failure to perfectly measure object selectivity and tolerance for each neuron, rather than a weak link between these properties. Another, nonexclusive possibility is that the trade-off produced by AND-like tuning mechanisms is partially opposed by OR-like pooling mechanisms used by ventral stream neurons to build tolerance to specific transformations (e.g., position and size). The interaction between such mechanisms would likely affect the strength of the correlation between selectivity and some tolerance properties.

When considering the possible mechanisms underlying the observed trade-off, we note that the lower retinal sampling in the periphery could contribute to limit position and size tolerance (Op De Beeck and Vogels, 2000). This might be especially true for highly selective neurons. Nevertheless, because we found that selectivity trades off with clutter and contrast tolerance, our experiments already provide evidence that retinal sampling cannot be the only underlying mechanism. In this regard, the AND-like tuning operation simulated in the computational models provides a more general mechanism that can explain the inverse relationship between object selectivity and every type of tolerance. The extent to which this relationship is affected by the balance between AND-like and OR-like operations (see above), as well as other properties/mechanisms (such as retinal sampling), may be revealed by further simulations of biologically grounded object recognition models.

The interaction between different tuning and tolerance mechanisms and retinal sampling could also explain why other trends, beside the trade-off between selectivity and each tolerance, seem to be present in some of the scatter plots of Figure 4. For instance, the range of variation of position tolerance seems to be wider for weakly selective than highly selective neurons. We do not yet know whether these other trends only reflect a failure to perfectly



measure object selectivity and tolerance for each neuron (see discussion above). Future studies could explore this further by measuring object selectivity and just one tolerance property using many more stimuli.

### Possible functional consequences of the trade-off

In discussing the functional implications of the observed trade-off, it should be noticed that tolerance to identity-preserving transformations can be conceptualized and measured in different ways. In this study, we focused on absolute tolerance. That is, we asked each IT neuron the same question one can ask of behavioral performance: How well preserved is its response to a very effective object when this object undergoes different identity-preserving transformations? However, this is not the only potentially relevant measure of neuronal tolerance. A related question is how well IT neurons maintain the rank order of their object preference across positions, sizes, etc. (relative tolerance). These measures are related. In particular, the results presented here on absolute tolerance provide a bound on relative tolerance: when the response to an effective object falls to near zero (the response to ineffective objects), no relative tolerance can remain. Similarly, when the absolute response is reduced, signal-to-noise in relative tolerance will also be reduced.

Which tolerance measure is most appropriate remains an important open question and ultimately requires an understanding of how IT neuronal activity is “read out” to guide behavior (Hung et al., 2005). IT neurons with absolutely tolerant selectivity could directly support tolerant object recognition behavior. If IT neurons have only relative tolerance, this would require further processing beyond IT to support that behavior. For example, appropriate combinations of IT neurons with only relative position and size tolerance can likely support position and size tolerant object recognition (Dicarlo and Cox, 2007). However, given the virtually infinite number of clutter conditions that can be encountered in natural vision, it is less obvious that populations of neurons with limited clutter tolerance can support recognition in clutter. Indeed, there is some empirical evidence that the brain aims to achieve absolute clutter tolerance at the level of individual ventral stream neurons. Specifically, when an animal is precued with the identity or position of a target object, existing studies (Moran and Desimone, 1985; Chelazzi et al., 1998; Reynolds et al., 1999; Reynolds and Chelazzi, 2004) suggest that ventral stream neurons move their absolute responses toward levels seen without clutter (i.e., as if no clutter stimuli were present). This suggests that, when possible, the brain employs resources to actively compensate for the interference produced by clutter and to achieve complete clutter tolerance (in absolute terms) at the level of individual neurons.

Our finding of broad ranges of selectivity and tolerance properties within IT and the trade-off of those properties suggests that the ventral visual stream is not striving to create a representation containing only cells with very sharp object selectivity (Kobatake and Tanaka, 1994). Rather, its goal may be to maximize the amount of tolerance given the constraint of achieving a spectrum of shape selectivity. Such a spectrum might provide a flexible basis in IT to support recognition tasks ranging from fine-grained identification to coarse categorization (Hung et al., 2005; Serre et al., 2005). In this regard, highly selective IT neurons, although capable of supporting very fine shape discriminations, likely require further feedforward stages of processing or top-down modulation (such as attentional feedback to filter out clutter interference). On the contrary, weakly selective IT neurons may quickly and reliably report the presence of some coarse classes of objects,

without clutter interference and over a large span of visual field, even without attentional control and the need of further processing stages.

### References

- Andermann ML, Ritt J, Neimark MA, Moore CI (2004) Neural correlates of vibrissa resonance; band-pass and somatotopic representation of high-frequency stimuli. *Neuron* 42:451–463.
- Bartho P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsaki G (2004) Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J Neurophysiol* 92:600–608.
- Brincat SL, Connor CE (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7:880–886.
- Brincat SL, Connor CE (2006) Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* 49:17–24.
- Britten KH, Heuer HW (1999) Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19:5074–5084.
- Bruno RM, Simons DJ (2002) Feedforward mechanisms of excitatory and inhibitory cortical receptive fields. *J Neurosci* 22:10966–10975.
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17:8621–8644.
- Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol* 80:2918–2940.
- Connors BW, Gutnick MJ (1990) Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci* 13:99–104.
- Constantinidis C, Goldman-Rakic PS (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J Neurophysiol* 88:3487–3497.
- De Baene W, Premereur E, Vogels R (2007) Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *J Neurophysiol* 97:2900–2916.
- Desimone R, Gross CG (1979) Visual areas in the temporal cortex of the macaque. *Brain Res* 178:363–380.
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051–2062.
- DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. *Trends Cogn Sci* 11:333–341.
- DiCarlo JJ, Maunsell JH (2003) Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89:3264–3278.
- Edelman S (1999) Representation and recognition in vision. Cambridge, MA: MIT.
- Fei-Fei L, Fergus R, Perona P (2004) Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. Paper presented at IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Generative-Model Based Vision, Washington DC, June 27–July 2, 2004.
- Frank LM, Brown EN, Wilson MA (2001) A comparison of the firing properties of putative excitatory and inhibitory neurons from CA1 and the entorhinal cortex. *J Neurophysiol* 86:2029–2040.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2005) Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex* 16:1631–1644.
- Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36:193–202.
- Gawne TJ, Martin JM (2002) Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J Neurophysiol* 88:1128–1135.
- Gross CG (2002) Genealogy of the “grandmother cell”. *Neuroscientist* 8:512–518.
- Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol* 35:96–111.
- Gur M, Beylin A, Snodderly DM (1999) Physiological properties of macaque V1 neurons are correlated with extracellular spike amplitude, duration, and polarity. *J Neurophysiol* 82:1451–1464.
- Hasenstaub A, Shu Y, Haider B, Kraushaar U, Duque A, McCormick DA (2005) Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron* 47:423–435.

- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9:181–197.
- Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310:863–866.
- Intraub H (1980) Presentation rate and the representation of briefly glimpsed pictures in memory. *J Exp Psychol [Hum Learn]* 6:1–12.
- Ito M, Tamura H, Fujita I, Tanaka K (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73:218–226.
- Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856–867.
- Logothetis NK, Sheinberg DL (1996) Visual object recognition. *Annu Rev Neurosci* 19:577–621.
- Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5:552–563.
- Lueschow A, Miller EK, Desimone R (1994) Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex* 4:523–531.
- Matsumoto N, Okada M, Sugase-Miyamoto Y, Yamane S, Kawano K (2005) Population dynamics of face-responsive neurons in the inferior temporal cortex. *Cereb Cortex* 15:1103–1112.
- McCormick DA, Connors BW, Lighthall JW, Prince DA (1985) Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J Neurophysiol* 54:782–806.
- Miller EK, Gochin PM, Gross CG (1993) Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res* 616:25–29.
- Missal M, Vogels R, Li CY, Orban GA (1999) Shape interactions in macaque inferior temporal neurons. *J Neurophysiol* 82:131–142.
- Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55:131–141.
- Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784.
- Mountcastle VB, Talbot WH, Sakata H, Hyvarinen J (1969) Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys. Neuronal periodicity and frequency discrimination. *J Neurophysiol* 32:452–484.
- Nowak LG, Azouz R, Sanchez-Vives MV, Gray CM, McCormick DA (2003) Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *J Neurophysiol* 89:1541–1566.
- Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. *Curr Opin Neurobiol* 14:481–487.
- Op De Beeck H, Vogels R (2000) Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426:505–518.
- Op de Beeck H, Wagemans J, Vogels R (2001) Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat Neurosci* 4:1244–1252.
- Perrett DI, Oram MW (1993) Neurophysiology of shape processing. *Image Vision Comput* 11:317–333.
- Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47:329–342.
- Poggio T, Bizzi E (2004) Generalization in vision and motor control. *Nature* 431:768–774.
- Potter MC (1976) Short-term conceptual memory for pictures. *J Exp Psychol [Hum Learn]* 2:509–522.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102–1107.
- Rainer G, Asaad WF, Miller EK (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393:577–579.
- Recanzone GH, Wurtz RH, Schwarz U (1997) Responses of MT and MST neurons to one and two moving objects in the receptive field. *J Neurophysiol* 78:2904–2915.
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647.
- Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19:1736–1753.
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.
- Riesenhuber M, Poggio T (2000) Models of object recognition. *Nat Neurosci* 3 [Suppl]:1199–1204.
- Rolls ET (2000) Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27:205–218.
- Rolls ET, Baylis GC (1986) Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp Brain Res* 65:38–48.
- Rolls ET, Tovee MJ (1995a) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* 73:713–726.
- Rolls ET, Tovee MJ (1995b) The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Exp Brain Res* 103:409–420.
- Rousset GA, Thorpe SJ, Fabre-Thorpe M (2004) How parallel is visual processing in the ventral pathway? *Trends Cogn Sci* 8:363–370.
- Rubin GS, Turano K (1992) Reading without saccadic eye movements. *Vision Res* 32:895–902.
- Sato T (1989) Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp Brain Res* 77:23–30.
- Schwartz O, Simoncelli EP (2001) Natural signal statistics and sensory gain control. *Nat Neurosci* 4:819–825.
- Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T (2007a) Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* 29:411–426.
- Serre T, Oliva A, Poggio T (2007b) A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci USA* 104:6424–6429.
- Serre T, Kouh M, Cadieu C, Knoblich U, Kreiman G, Poggio T (2005) A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. *CBCL Paper #259/AI Memo #2005-036*. Cambridge, MA: MIT.
- Sheinberg DL, Logothetis NK (2001) Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J Neurosci* 21:1340–1350.
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–873.
- Swadlow HA (2003) Fast-spike interneurons and feedforward inhibition in awake sensory neocortex. *Cereb Cortex* 13:25–32.
- Tamura H, Kaneko H, Kawasaki K, Fujita I (2004) Presumed inhibitory neurons in the macaque inferior temporal cortex: visual response properties and functional interactions with adjacent neurons. *J Neurophysiol* 91:2782–2796.
- Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109–139.
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522.
- Tovee MJ, Rolls ET, Azzopardi P (1994) Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *J Neurophysiol* 72:1049–1060.
- Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287:1273–1276.
- Wallis G, Rolls ET (1997) Invariant face and object recognition in the visual system. *Prog Neurobiol* 51:167–194.
- Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci* 25:8150–8164.