Supplementary Material

Trade-off between object selectivity and tolerance in monkey inferotemporal cortex

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

(1.A) *Selection of the spike count window*

For each neuron, responses to each stimulus condition were computed over the same spike count time window. That window was optimally chosen for each neuron using the following algorithm (see also Supp. Fig. 2). Given a neuron, we first computed its average firing rate profiles $FR_{obj}(t)$ across multiple presentations of each object condition in overlapping time bins of 25 ms shifted in time steps of 1 ms. Then, we averaged the $FR_{obj}(t)$ evoked by all the object pairs and by the single objects whose average response (computed between 75 and 225 ms after stimulus onset) was at least 70% of the response produced by the most effective object of the set. By averaging across these same object conditions, we also computed the background rate FR_{bk} of the neuron (spikes were counted between 0 and 50 ms after stimulus onset). Then, by subtracting this stimulus-averaged background from the stimulus-averaged firing rate profile, we obtained a stimulus-averaged driven rate profile $FR_{driven}(t) = \langle FR_{obj}(t)\rangle_{obj}$ - FR_{bk}. Finally, we identified the samples for which $FR_{driven}(t)$ was at least 20% of its peak value. The largest continuous interval of samples fulfilling this requirement was always centered on the peak of the neuronal response. If no other samples, outside this main "peak" interval, fulfilled the requirement, the extremes of the interval were chosen as the extremes of the optimal spike count window for that neuron (Supp. Fig. 2A). If the firing profile exceeded 20% of its peak in other regions of the time axis, these were merged with the main interval only if they were within 25 ms from it (Supp. Fig. 2A). In this case, the extremes of the merged interval were chosen as the extremes of the optimal spike count window. For each neuron, the extremes of the spike count window were never allowed to be lower than 50 ms and higher than 300 ms from stimulus onset.

All analysis presented in the main text were carried out by counting spikes in these neuron-specific optimized time windows. The mean window start time (± s.d.) was

101 ± 18 ms, the mean window end time was 236 ± 48 ms, and the median duration was 135 ± 48 ms. These time windows are consistent with previous work[52] and with animal reaction times in recognition tasks[62]. All analyses were also repeated using fixed spike count windows for all the recorded neurons, yielding very similar results (see Supp. Tables 1- 2).

(1.B) *Independence of data sets used to compute selectivity and tolerance metrics*

Five to eight presentation repetitions were collected for each shape in the fixed object set. Ten to eighteen repetitions were collected for the object conditions used to measure the tolerance properties and selectivity within the morphed sets. The reference object was presented 20 to 30 times.

Selectivity and tolerance properties of neuronal responses were measured using the metrics defined in Methods. Since we sought to understand the relationship between these properties, we took care to measure them using independent data sets. In general, independent sets of object conditions were used to assess selectivity and tolerance properties. However, since a measure of the response to the very effective reference object (see Methods) was needed for both the sparseness metric and most of the tolerance metrics (and, occasionally, the morph tuning metric), we split the stimulus presentations collected for the reference object in such a way that: 1) a subset of presentations was used to compute selectivity; and 2) an independent subset with the remaining presentations was used to compute the tolerance metrics. As assessed by bootstrap, tolerance metrics were much more sensitive to spike train variability than the sparseness metric. In fact, the sparseness metric is computed using a huge number of stimuli (>200). As a consequence, this metric is pretty much insensitive to variations in the average firing rate produced by each stimulus. On the opposite, tolerance metrics are computed using a much more limited number of object conditions and, therefore, are more sensitive to spike train variability. Therefore, 90% of the presentations collected for the reference object were used in the tolerance metrics, while the remaining, non-overlapping 10% entered the sparseness metric. Similarly, the presentations collected for the reference object were split in two non-overlapping halves, one contributing to the morph tuning metric, the

other to the tolerance metrics. Therefore, all scatter plots and correlation coefficients (Fig. 4, Supp. Tables 1-3) presented in this study were obtained using completely independent data sets.

(1.C) *Inclusion of neurons in correlation analyses*

As explained in Methods, 94 neurons were selected for recording in two monkeys using a very inclusive criterion. Each analysis presented in the manuscript was carried out using a subset of neurons that was selected according to specific criteria.

Since the responses to the reference object (R_{ref}) and to identity-preserving transformations of this object entered all the tolerance metrics, we required that each neuron fired at least 10 spikes/s to the reference object in order to be included in any correlation analysis. To focus only on neurons that were strongly activated by the reference object, all analyses were also repeated using stricter thresholds on R_{ref}, such as: $R_{ref} \geq$ Th, with Th = 15, 20, 30 spikes/s (see Figs 4 and 6A and Supp. Tables 1-3).

As explained in Methods, clutter tolerance (*CT*) was assessed using responses to six isolated flanking objects ($R_{flanker}$) and to six object pairs composed of the reference object and each flanker ($R_{ref \& flanker}$). Clutter tolerance was meant to assess the amount of interference in the response to a preferred reference object caused by the presence of poorly-effective flanker objects. Although we used preliminary screening in an attempt to choose only flanker objects (see Methods) that were poorly effective, this was not always born out by more thorough testing. Thus, only object conditions such that $R_{flanker} < 0.5$ R_{ref} were allowed to contribute to the clutter tolerance metrics (but stricter criteria has been applied in Fig. 6A). Therefore, neurons for which no flankers satisfied such criterion were excluded from any correlation analysis involving clutter tolerance.

Size tolerance was measured for all neurons recorded from monkey 2 (34 cells) but only for a subset of neurons recorded from monkey 1 (18 cells), for a total of 52 neurons.

The tuning within a morphed set (morph tuning, *MT*) was measured only for a subpopulation of 49 neurons that responded to some of the morphed objects. In addition,

we required that each neuron fired at least 10 spikes/s to the most effective morphed object in order to be included in any correlation analysis (all 49 neurons met such requirement).

Since sensitivity to contrast changes of the reference object was measured 2.5º above the RF center, as a part of a more general assessment of clutter tolerance (see Methods), we required that the response to the reference object at the reference contrast in this RF position was at least 25% of the response to the same object in the RF center. Neurons that did not fulfill this requirement were excluded from any analyses involving contrast tolerance. Since the contrast tolerance metric measures how much of the response to the reference object at the reference (high) contrast is preserved when the contrast is reduced to much lower levels, we also excluded 4 neurons with contrast tolerance $CrT > 1$. In fact, these neurons had "inverted" contrast sensitivity, i.e. their average response across the low contrast stimuli was higher than the response to the reference (higher) contrast. These 4 neurons were excluded from our analysis, since our goal was to measure contrast tolerance for a population of neurons with "regular" contrast sensitivity function (i.e., that all preferred the object at the highest contrast).

These different screenings were combined yielding the neuronal subpopulations used in the correlation analyses shown in Fig. 4 and Supp. Tables 1-3.

(2.A) *Overview of the computational model of object recognition*

We used a model of object recognition to explore the mechanisms underlying the trade-off between selectivity and tolerance that was observed in IT. The model is composed of hierarchical layers of neuronal units performing feedforward operations that gradually build up selectivity and tolerance for robust object recognition. A short overview of the model is provided below, and more details and other aspects of the model can be found in previous reports (Riesenhuber and Poggio, 1999a; Serre, 2005; Serre et al., 2007a; Serre et al., 2007b).

There are two principle operations in the model: one for Gaussian-like tuning and the other for max-like selection or tolerance (Riesenhuber and Poggio, 1999a; Serre, 2005; Serre et al., 2007a; Serre et al., 2007b). Each unit ("neuron") in the model is connected to n afferent or input units (i.e., the "presynaptic neurons"). The tuning operation is performed by a Gaussian-like function over a space of continuous-value inputs (i.e., the activation values of the afferent units). A high response is produced when an input pattern across the afferent units matches a stored preferred pattern (e.g., the center of the Gaussian-like tuning function; see Figs. 8A-B). The other principle operation in the model is the *maximum* (MAX) and its approximations, to implement translation and scale tolerances, based on some physiological evidence and biologically plausible neural circuits[1,2].

Along the hierarchy of the model, these two operations are repeated in an interleaved manner by the units performing the Gaussian-like operation and the units performing the maximum-like operation. As a result, progressively more tolerant (to changes in position and scale) and more complex neural selectivities are generated, as observed along the ventral pathway.

The model is constructed to reflect the anatomy and physiology of the ventral pathway, so that the hierarchical layers in the model correspond to the visual areas V1,

---

[1] Lampl, I., Ferster, D., Poggio, T. & Riesenhuber, M. Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J Neurophysiol* 92, 2704-13 (2004).
[2] Yu, A. J., Giese, M. A. & Poggio, T. A. Biophysiologically plausible implementations of the maximum operation. *Neural Comput* 14, 2857-81 (2002).

V2, V4, and IT, and the receptive field sizes and tuning properties of the model units closely match those of the neurons in the corresponding areas (see Serre, 2005; Serre et al., 2007a; Serre et al., 2007b). The model units in the top layer are the most relevant in this study, as they correspond to the neurons in the anterior inferotemporal (IT) cortex. Each model IT unit is connected to a subset of input units from the previous layer (PIT) through a Gaussian-like tuning operation. The tuning function of each model IT unit was centered on the activation pattern produced, across its set of input units, by one of the objects of the fixed stimulus set. This means that a model IT neuron had, as a preferred stimulus, one of objects of the fixed stimulus set. Simulation results did not change when the preferred stimuli of the model IT units did not exactly match the objects of the fixed stimulus set used to probe their selectivity and clutter tolerance.
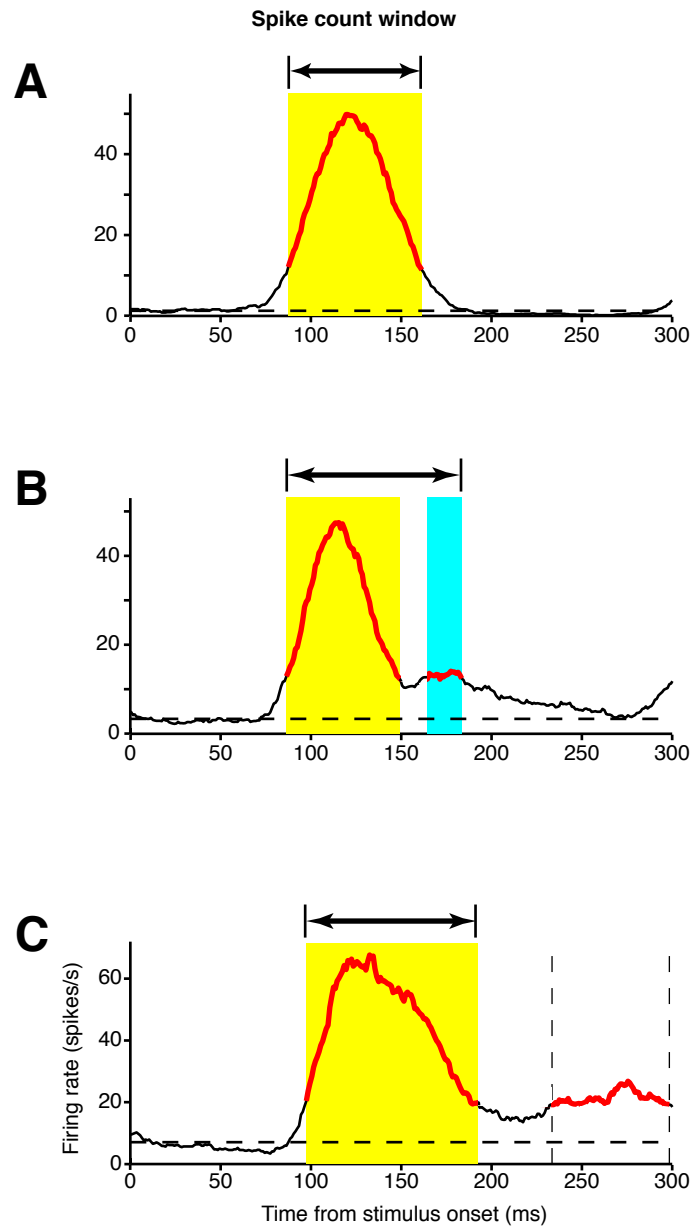
(2.B) *Trade-off between selectivity and tolerance in the model of object recognition*

Our simulations tried to follow as closely as possible the experimental protocols used during recordings. The same object stimuli (Supp. Fig. 1) were "presented" as inputs to the model and the same metrics (with the exception of position tolerance, see Legend of Supp. Fig. 7) were computed to measure selectivity and tolerance of the responses of the model units. We fixed all the parameters in the lower layers of the model and asked whether a range of selectivity and tolerance similar to that found in the recorded IT neuronal population (as well as the trade-off between these properties) could be observed in the layer of the model corresponding to anterior IT.

As shown in Supp. Fig. 7, having a population of model IT units with a variable number of afferents produced wide ranges of selectivity and tolerance and a trade-off between these two properties in a very close agreement with the experimental results. The reason is that, as the number of afferents increases, it becomes harder for an arbitrary visual stimulus to produce a pattern of activation across the afferents that matches the stored, preferred pattern. As a consequence, the selectivity of the unit increases. At the same time, the tolerance of the unit decreases, since it becomes easier, for some stimulus transformations (such as adding arbitrary flanker (clutter) stimuli and changing the position, size and contrast of the preferred stimulus), to produce large deviations from the

optimal input pattern. Increasing the dimensionality of the input space (i.e., the number of afferents) as done in Fig. 8 and Supp. Fig. 7 is just one way to shrink such a volume. Another obvious way is to increase the sensitivity of the tuning along each input dimension (i.e., to reduce σ of the Gaussian-like tuning function), while keeping the number of afferents constant. This again produces a similarly robust trade-off between selectivity and tolerance in the model IT units. Finally, it should be noticed that any plasticity mechanism able to modify the number of inputs to a given neuron, by strengthening or pruning certain synapses, can be used to control the amount of selectivity and tolerance of the neuron.

Object set: 213 grayscale images of natural objects

Supplemental Figure 1. The fixed stimulusset. The full set of 213 stimuli used to probe the selectivity of each recorded neuron. The set consists of: i) 188 images of real-world objects belonging to 94 different categories (e.g., two hats, two accordions, two monkey faces, etc.); ii) 5 objects belonging to each of three different sets of morphed shapes (cars, faces and 2D silhouettes; see below); iii) 5 patches of texture (e.g., random dots and oriented bars); iv) a blank frame (used to compute background rate); v) 4 low contrast (10%, 3%, 2% and 1.5%) images of one of the objects (a camera).

**Optimal spike count window computed for three example neurons**

Supplemental Figure 2. Spikes count windows computed for three example neurons. In each plot, the black continuous line and the horizontal dashed line are, respectively, the stimulus-averaged firing rate profile $\langle FR_{obj}(t)\rangle_{obj}$ and background rate FR_{bk} (see Supp. Material 1.A). The red line shows the portion of the firing profile that exceeds 20% of its peak value (relative to the background rate). **A**, For this example neuron, the 20% threshold was crossed in a single continuous interval centered around the peak of the firing profile (yellow patch). This interval was chosen as the spike count window for the neuron. **B**, For this example neuron, the 20% threshold was crossed in two distinct regions (yellow and cyan patches) that were merged to obtain the final spike count window, since they were separated by less than 25 ms. **C**, For this example neuron, the 20% threshold was crossed in two distinct regions (yellow patch and interval between the vertical dashed lines) that were not merged because they were separated by more than 25 ms. Therefore, the interval centered around the peak of the firing (yellow patch) was taken as the spike count window for the neuron.

Independence of clutter tolerance from contrast of flankers

A

B
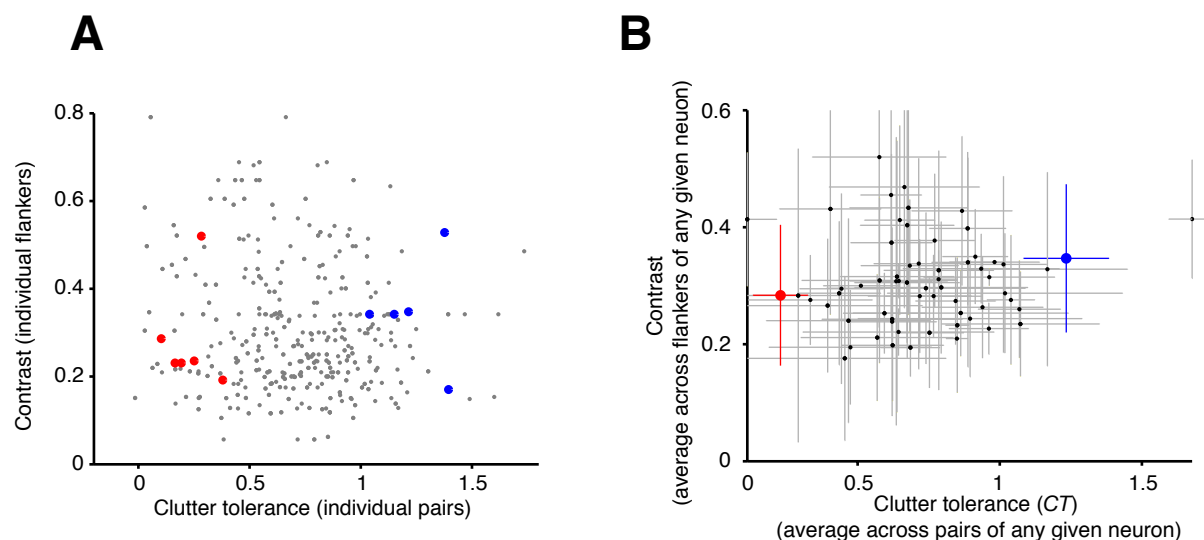
Supplemental Figure 3. The trade-off between selectivity and clutter tolerance does not depend on the contrast of the flankers in the object pairs used to measure clutter tolerance. We asked if covariation of flanker contrast with selectivity could explain the inverse relationship. This could, in principle, happen, since: i) flankers chosen for less selective neurons might be expected to have, on average, a lower contrast than flankers used for very selective cells; and ii) flankers with low contrast may have a lower suppressive power when paired to a very effective object. If both i) and ii) were true, then a strong inverse relationship may be expected between contrast of individual flankers and the amount of response suppression they produced (i.e., the clutter tolerance for individual flankers). Panel **A** shows that this was not the case. In fact, contrast and clutter tolerance of individual flankers were only weakly and not significantly negatively correlated (left panel; $r = -0.092$, $p = 0.094$, two-tailed t-test, $n = 330$). In addition, multiple flankers (up to 6) were used to compute the clutter tolerance (*CT*) of each neuron (see Methods), and they typically spanned a broad range of contrasts, regardless of the neuron's selectivity (compare the sets of red and blue dots, which refer to the flankers used to assess the clutter tolerance of the example neurons shown, with corresponding colors, in Figs. 2 and 3). Interestingly, the average contrast of the flankers used for a given neuron was positively, although weakly and not significantly, correlated with the clutter tolerance across the neuronal population (panel **B**; $r = 0.065$, $p = 0.61$, two-tailed t-test, $n = 63$; vertical and horizontal bars are the SD of the average contrast and clutter tolerance across the flankers used for each neuron; same color code as in the left panel). This shows that the inverse relationship between selectivity and CT cannot simply be explained by differences in the contrast of the flanker objects (see also Supp. Fig. 4).

Trade-off between sparseness and clutter tolerance (only flankers with fixed identity)

A

B
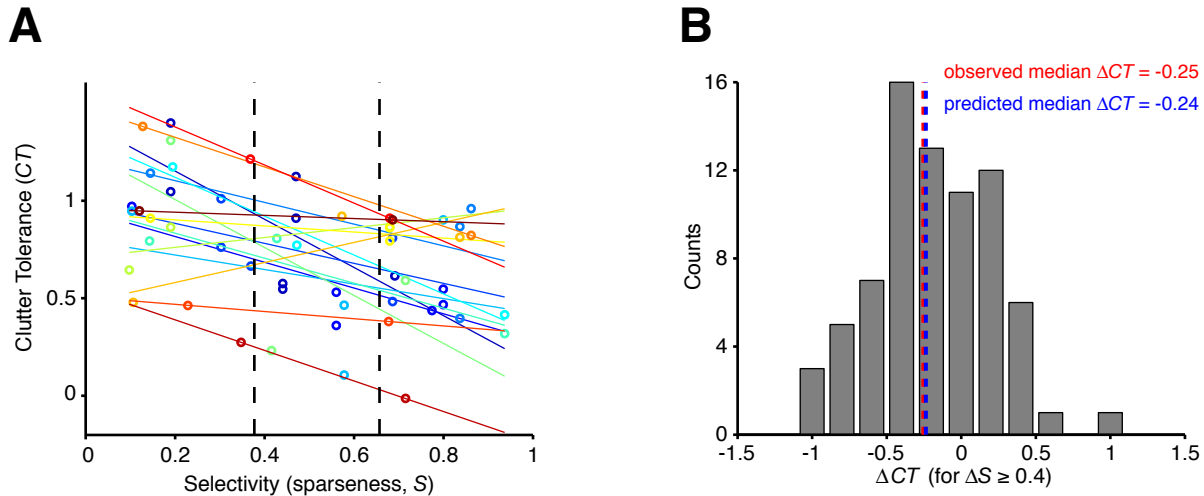
Supplemental Figure 4. The trade-off between selectivity and clutter tolerance does not depend on the identity of the flankers in the object pairs used to measure clutter tolerance. **A**, To fully remove all possible properties of the chosen flanker objects that might happen to covary with shape selectivity, we specifically examined cases where neurons were tested with the exact same flanker objects. The figure shows 16 sets of neurons (each set is identified by a color, and ranges from 2 to 7 neurons in size). For neurons in a given set, clutter tolerance (ordinate) was measured using exactly the same flanker object. Each set contained at least one neuron with selectivity (sparseness, S) within the first third of its possible range of values (i.e., to the left of the leftmost dashed line) and one neuron with S within the last third (i.e., to the right of the rightmost dashed line). For each set, we computed a linear regression between selectivity and clutter tolerance (regression lines for each set are shown with color matching the color of the corresponding data points). The mean slope of the regression lines was significantly negative (-0.52, p = 0.00039, one-tailed t-test, n = 16) and very close to the slope of the regression line obtained from the whole data set (slope = -0.54, with $R_{ref} \geq 30$ spikes/s). Also, the mean value of the clutter tolerance for neurons within the first third of S was significantly higher than the mean clutter tolerance of neurons within the last third of S (p = 0.0015, one-tailed t-test). **B**, We directly compared the strength of the inverse relationship (shape selectivity vs. clutter tolerance) when measured with the exact same flanker objects with that observed in the entire data set (Fig 4), in which the flanker identity typically varied from neuron to neuron. To do this, we again focused on those sets of neurons that were tested with the same flankers (as in **A**). For each pair of neurons within a set, we computed the difference in sparseness ($\Delta S > 0$, since we always subtracted the lower sparseness value from the higher) and the corresponding change in clutter tolerance (ΔCT). The histogram shows the distribution of the ΔCTs, for $\Delta S > 0.4$. As expected, there was a significantly higher number of negative, rather than positive, ΔCTs (p = 0.00034, χ^2 test). More importantly, the observed median ΔCT (red dashed line) was almost identical to the median of the predicted ΔCTs (blue dashed line) based on applying the slope of the linear regression obtained from the whole data set (as in Fig 4; slope = -0.46 with $R_{ref} \geq 10$ spikes/s) to the observed ΔSs. Together, the analyses presented in **B** and **C**, show that the trade-off between S and CT: i) was present in neuronal populations in which clutter tolerance was measured using the exact same flanker objects; and ii) was quantitatively very close to that obtained from the whole data set.
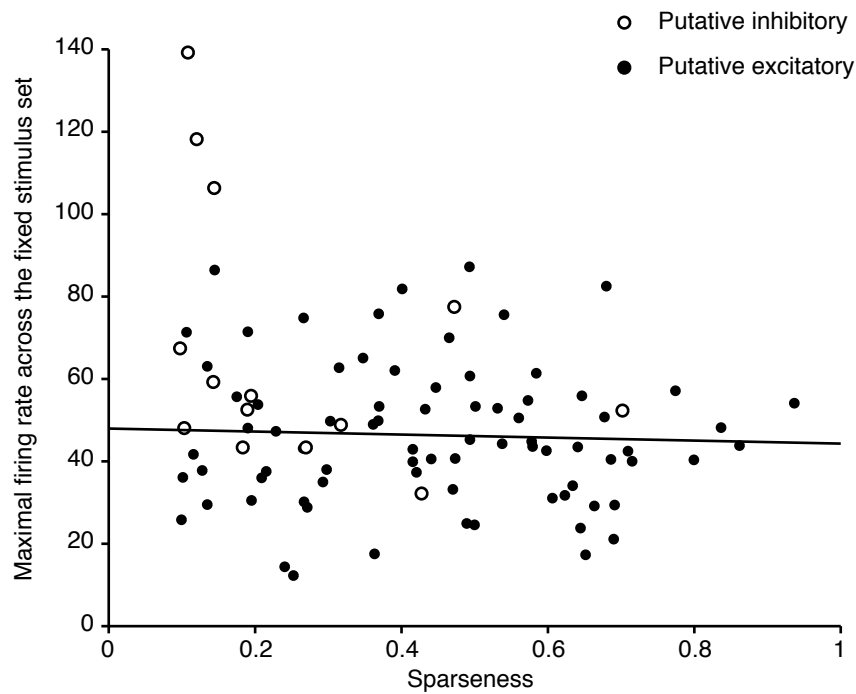
Supplemental Figure 5. Correlation between sparseness and neuronal maximal firing rate over the fixed stimulus set
When the whole neuronal population ($n = 94$) was taken into account, a significant, negative correlation was found between the maximal firing of neurons across the stimulus set and their sparseness ($r = -0.21^*$, $p = 0.043$, $n = 94$; two-tailed t-test). Such correlation, however, was mainly driven by a few "outlier" putative inhibitory neurons (open circles; see Fig. 5A) with very high firing rate and minimal selectivity. Indeed, the correlation coefficient computed over the population of excitatory neurons only (filled circles) was not significantly different from zero ($r = -0.045$, $p = 0.7$, $n = 80$; two-tailed t-test).

A **Object selectivity (sparseness) vs. tolerance**
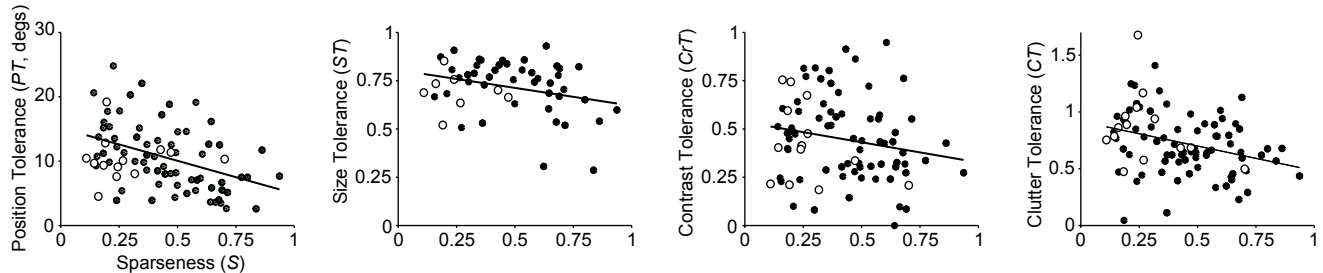
B **Object selectivity (morph tuning) vs.tolerance**
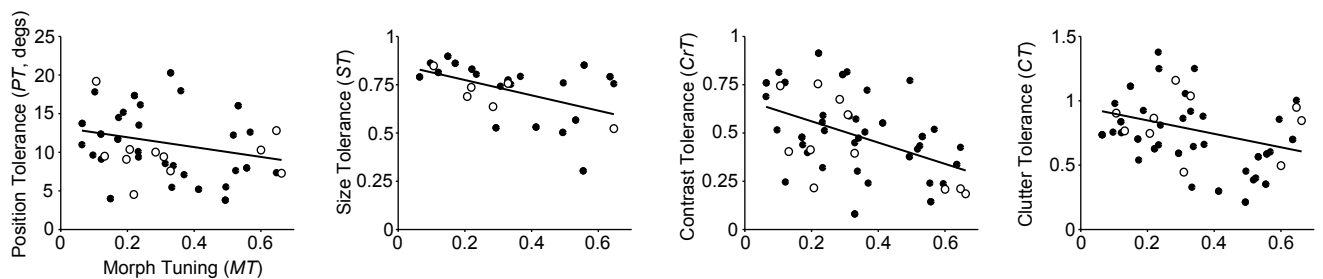
Supplemental Figure 6. Trade-off between object selectivity and tolerance to identity-preserving transformations in IT after subtraction of minimal rates. For each neuron, sparseness, morph tuning and each of the tolerance metrics were computed using firing rates that were corrected by subtracting the minimal neuronal response across the fixed object set of 213 stimuli (see Supp. Fig. 1). *A,* The scatter plots show the inverse relationship between sparseness 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 Fig. 5A. Regression lines through, respectively, 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 and each of the tested tolerance properties. Same symbols as in A. Neurons that fired at least 10 spikes/s to the reference object were included in the plots shown in A and B. Correlation coefficients between object selectivity (either sparseness or morph tuning) and each of the tolerance properties are reported in Supp. Table 4.
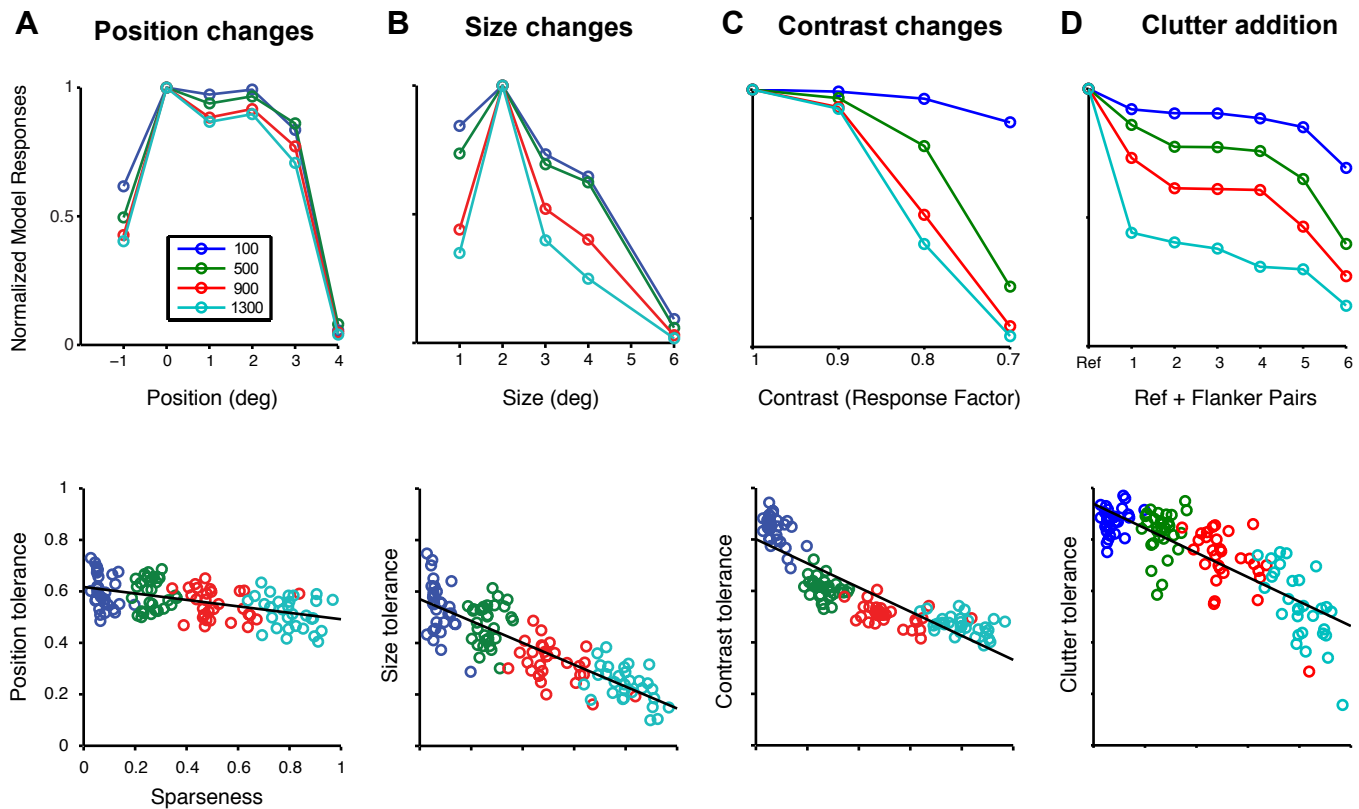
Supplemental Figure 7. Trade-off between selectivity and tolerance to stimulus transformations in the model of object recognition In the top panels, the tuning curves of 4 model IT units (with different number of afferent units as indicated in the legend of the first panel) to four different stimulus transformations (position, size, contrast, and clutter) are shown (cf. Fig. 3). In the bottom panels, the sparseness and the tolerance metrics for these four transformations are plotted for a population of model units (a total of 120 model units with 4 different numbers of afferent units) (cf. Fig. 4). As discussed in the text, the units with more afferents inputs are more selective and less tolerant resulting in an inverse relationship, similar to that observed in the recorded IT neuronal population (Fig. 4). **A, B,** Because of the maximum-like operations across different positions and scales in the lower layers of the model, the responses of the model IT units show some degree of tolerance to position and scale changes. However, it should be noted that in general the presence of OR-like (e.g., maximum) operations does not guarantee perfect invariance across position and size changes, because these tolerance operations are applied in a hierarchical manner, interleaved by AND-like (e.g., Gaussian) operations. In the model, a hierarchy of OR-like operations followed by AND-like operations is not equivalent to a global OR operation at the top level. Additionally, pooling over only a finite number of positions and sizes over which the model units perform the OR-like operations contribute to the imperfect invariance of the model. Since model IT units perform a Gaussian-like tuning operation over their afferents, those IT units that are more selective (e.g., they receive more inputs) will still be more sensitive to the (not completely invariant) activation of their afferents due to position and size changes. Note that, although these simulations show a qualitative agreement with the data, the ranges of positions and scale changes tested in the model do not match those probed during recordings, because of the limited span of visual field simulated in the model. Moreover, a more realistic and quantitative comparison with the data would require modeling the variations in size and location of the receptive fields of inferotemporal neurons and the drop of retinal sampling as a function of retinal eccentricity. For these reasons, since a quantitative comparison with neuronal position data is currently beyond the scope of our simulations, we decided to quantify position tolerance as the average response of the model IT unit to the reference object presented in the non-preferred RF locations rather than as the unit RF size (as done for the experimental data). **C,** For the simulations on the contrast sensitivity of the model units, we assumed that stimuli with lower contrast would produce smaller responses in the model units representing V1 neurons. Specifically, we assumed that, in the model, the responses of each V1 unit to stimuli at 1.5, 2 and 3% contrast (cf. Fig. 3C) were, respectively, 70, 80 and 90% (abscissa in the top panel) of the response to the same stimulus at its reference contrast. **D**, The simulation of the clutter condition followed the same protocols as in the physiological experiment.

Supp. Table 1

Correlation between selectivity and tolerance properties
Selectivity metric = _sparseness (S) within the fixed object set_

Th (spikes/s)	Position tolerance	(n)	Size tolerance	(n)	Contrast tolerance	(n)	Clutter tolerance	(n)
Spike count window = neuron specific (same as first two rows of Fig. 4C)								
10	-0.39***± 0.08	(77)	-0.33**± 0.13	(52)	-0.32**± 0.1	(85)	-0.35***± 0.09	(88)
30	-0.36**± 0.11	(56)	-0.36**± 0.17	(42)	-0.3*± 0.11	(60)	-0.46***± 0.08	(63)
Spike count window = neuron specific (DRIVEN rates – background rate computed at trial onset)								
10	(above)		-0.21 ± 0.14	(52)	-0.18*± 0.1	(85)	(above)	
30	(above)		-0.26*± 0.17	(42)	-0.15 ± 0.11	(60)	(above)	
Spike count window = neuron specific (DRIVEN rates – background rate computed as response to a blank frame)								
10	-0.38***± 0.08	(79)	-0.28*± 0.14	(52)	-0.19*± 0.1	(85)	(above)	
30	-0.34**± 0.12	(57)	-0.29*± 0.18	(42)	-0.14 ± 0.13	(60)	(above)	
Spike count window = neuron specific (NO putative INHIBITORY)								
10	-0.43***± 0.09	(64)	-0.4**± 0.12	(44)	-0.3**± 0.1	(72)	-0.29**± 0.11	(74)
30	-0.36**± 0.12	(44)	-0.5***± 0.14	(34)	-0.3*± 0.11	(48)	-0.41***± 0.11	(50)
Spike count window = [75, 225] ms								
10	-0.38***± 0.08	(80)	-0.26*± 0.15	(52)	-0.38***± 0.1	(90)	-0.36***± 0.08	(92)
25	-0.4***± 0.1	(53)	-0.25 ± 0.19	(40)	-0.35**± 0.12	(57)	-0.43***± 0.09	(59)
Spike count window = [100, 200] ms								
10	-0.26**± 0.12	(76)	-0.21 ± 0.15	(52)	-0.38***± 0.08	(87)	-0.3**± 0.09	(91)
30	-0.23**± 0.13	(57)	-0.23 ± 0.19	(41)	-0.33**± 0.11	(61)	-0.36**± 0.09	(65)
Spike count window = [100, 250] ms								
10	-0.29**± 0.1	(78)	-0.2 ± 0.14	(52)	-0.3**± 0.1	(87)	-0.37***± 0.07	(90)
25	-0.31**± 0.13	(58)	-0.29* ± 0.15	(41)	-0.25*± 0.13	(59)	-0.45***± 0.08	(62)
Spike count window = [75, 250] ms								
10	-0.41***± 0.08	(81)	-0.26* ± 0.14	(52)	-0.39***± 0.1	(89)	-0.35***± 0.08	(90)
25	-0.38**± 0.11	(52)	-0.26 ± 0.18	(38)	-0.24*± 0.13	(53)	-0.46***± 0.09	(53)

Supplemental Table 1. Correlation between selectivity (_sparseness_) and tolerance properties for different choices of the parameters used to compute the sparseness and tolerance metrics. The first column (Th) shows the minimal spike rate each neuron had to fire to the reference object in order to be included in the analysis. The first two rows show the correlations obtained using our _standard_ analysis parameters (already reported in Fig. 4C), i.e. by computing raw rates in neuron-specific spike count windows (see Supp. Material 1.A) and including both putative excitatory and inhibitory neurons (see in Fig. 5A). The other rows show the correlations obtained using either: i) driven rates instead of raw rates; or ii) excluding putative inhibitory neurons from the analysis; or iii) using various fixed spike count windows for all the neurons included in the analysis. Driven rates were computed by subtracting from each response of a neuron its background rate. The latter was computed either as: i) the average firing rate in a 100 ms window at the beginning of each behavioral trial, after fixation was acquired but before any stimulus was presented; or ii) the average firing rate to a blank frame embedded in the stimulus sequence. Note that, since some parameter choices are the same (or equivalent) to the _standard_ parameter choice, correlation values are not repeated (cells filled with "above"). This applies to the clutter tolerance metric that, by definition, is the same using either raw or driven rates, and to the position tolerance metric that, by default, is computed using driven rates (with background rate computed at the beginning of each trial). Correlation coefficients are all negative and, in most cases, highly significant (* $p \le 0.05$; ** $p \le 0.01$; ***$p \le 0.001$; one-tailed permutation test; SE computed by bootstrap).

Supp. Table 2

Correlation between selectivity and tolerance properties
Selectivity metric = _tuning within a morphed set (morph tuning, MT)_

Th (spikes/s)	Position tolerance	(n)	Size tolerance	(n)	Contrast tolerance	(n)	Clutter tolerance	(n)
	Spike count window = neuron specific (same as last two rows of Fig. 4C)							
10	-0.26*± 0.12	(41)	-0.47**± 0.16	(26)	-0.52***± 0.09	(48)	-0.38**± 0.09	(48)
20	-0.29*± 0.1	(38)	-0.43*± 0.16	(25)	-0.52***± 0.1	(45)	-0.32*± 0.1	(45)
	Spike count window = neuron specific (DRIVEN rates – background rate computed at trial onset)							
10	(above)		-0.45* ± 0.16	(26)	-0.45***± 0.11	(48)	(above)	
20	(above)		-0.43*± 0.18	(25)	-0.44**± 0.11	(45)	(above)	
	Spike count window = neuron specific (DRIVEN rates – background rate computed as response to a blank frame)							
10	-0.2 ± 0.12	(42)	-0.47**± 0.16	(26)	-0.41**± 0.14	(48)	(above)	
20	-0.25 ± 0.11	(39)	-0.46*± 0.18	(25)	-0.43** ± 0.14	(45)	(above)	
	Spike count window = neuron specific (NO putative INHIBITORY)							
10	-0.24 ± 0.12	(30)	-0.48*± 0.17	(20)	-0.48**± 0.11	(37)	-0.37*± 0.11	(37)
20	-0.26 ± 0.14	(27)	-0.45*± 0.17	(19)	-0.47**± 0.13	(34)	-0.3*± 0.13	(34)
	Spike count window = [75, 225] ms							
10	-0.22 ± 0.13	(35)	-0.48**± 0.2	(25)	-0.57***± 0.09	(44)	-0.42**± 0.09	(44)
20	-0.22 ± 0.14	(33)	-0.48** ± 0.2	(25)	-0.54***± 0.1	(41)	-0.39**± 0.11	(41)
	Spike count window = [100, 200] ms							
10	-0.2 ± 0.15	(37)	-0.6*** ± 0.16	(24)	-0.57***± 0.1	(42)	-0.39**± 0.1	(45)
25	-0.19 ± 0.15	(32)	-0.6** ± 0.16	(23)	-0.52***± 0.12	(35)	-0.36*± 0.12	(38)
	Spike count window = [100, 250] ms							
10	-0.18 ± 0.13	(38)	-0.56** ± 0.18	(24)	-0.61***± 0.09	(43)	-0.41**± 0.1	(44)
20	-0.3*± 0.11	(34)	-0.54** ± 0.21	(22)	-0.61***± 0.11	(38)	-0.37**± 0.11	(39)
	Spike count window = [75, 250] ms							
10	-0.21 ± 0.12	(39)	-0.53** ± 0.17	(25)	-0.64***± 0.08	(44)	-0.45***± 0.1	(43)
15	-0.26 ± 0.11	(32)	-0.52* ± 0.19	(22)	-0.63***± 0.1	(37)	-0.43**± 0.1	(36)

Supplemental Table 2. Correlation between selectivity (_morph tuning_) and tolerance properties for different choices of the parameters used to compute the sparseness and tolerance metrics. Same conventions as in Supp. Table 1. All correlation coefficients are negative and, in most cases, highly significant.

Supp. Table 3

Correlation between _different selectivity measures over the fixed stimulus set_ and tolerance properties

Th (spikes/s)	Position tolerance	(n)	Size tolerance	(n)	Contrast tolerance	(n)	Clutter tolerance	(n)
Sparseness (raw rates; same as first two rows of Fig. 4C)								
10	-0.39***± 0.08	(77)	-0.33**± 0.13	(52)	-0.32**± 0.1	(85)	-0.35***± 0.09	(88)
30	-0.36**± 0.11	(56)	-0.36**± 0.17	(42)	-0.3*± 0.11	(60)	-0.46***± 0.08	(63)
Sparseness (after subtraction of minimal rates)								
10	-0.37***± 0.09	(77)	-0.33**± 0.13	(52)	-0.29**± 0.1	(85)	-0.32**± 0.09	(88)
30	-0.33**± 0.11	(56)	-0.37** ± 0.17	(42)	-0.27*± 0.11	(60)	-0.42***± 0.08	(63)
Sparseness (absolute driven rates)								
10	-0.37**± 0.1	(77)	-0.21 ± 0.12	(52)	-0.22*± 0.1	(85)	-0.32**± 0.1	(88)
30	-0.28*± 0.13	(56)	-0.19 ± 0.15	(42)	-0.2 ± 0.12	(60)	-0.46***± 0.08	(63)
Fraction of responses NOT significantly different from background rate ($p = 0.1$, t-test)								
10	-0.4***± 0.09	(77)	-0.24* ± 0.12	(52)	-0.17± 0.11	(85)	-0.27**± 0.12	(88)
30	-0.34**± 0.11	(56)	-0.21 ± 0.12	(42)	-0.24 ± 0.16	(60)	-0.4**± 0.11	(63)
Average absolute difference between responses to twin objects								
10	-0.36***± 0.09	(77)	-0.22 ± 0.12	(52)	-0.25*± 0.09	(85)	-0.26**± 0.09	(88)
25	-0.34**± 0.11	(56)	-0.25 ± 0.14	(42)	-0.3*± 0.09	(60)	-0.38**± 0.09	(63)

Supplemental Table 3. Correlation between different selectivity metrics and tolerance properties. In this study, we quantified neuronal selectivity across the fixed object set (see Supp. Fig. 1) using a well-established metric: the sparseness of the raw neuronal responses (first two rows). Here we show that the tradeoff between selectivity and tolerance properties is also observed when other selectivity metrics are computed (using the responses across the fixed object set). i) Sparseness after subtracting from each response the minimal response across the stimulus set. ii) Sparseness after subtracting the background rate and taking the absolute value of the resulting responses (note: that the sparseness is ill defined if some of the responses are negative). iii) The fraction of stimuli that did not produce a response significantly different from background rate according to a t-test ($p = 0.1$); iv) The average absolute difference between the responses to two objects belonging to the same object category ("twin" objects). The last metric was based on the fact that the fixed stimulus set consisted of 94 object categories, each containing two exemplars/twins (e.g., two hats, two monkey faces, etc...). To compute this metric, we selected all the objects that evoked a response that was at least 50% of the response to most effective object. For each of these objects, we found its twin and we computed the absolute difference between the firing rates evoked by the two objects. Finally, we averaged the firing differences obtained for all these pairs of twin objects. Overall, a negative correlation was always found between each of these selectivity metrics and the tolerance metrics, and most of these correlations were significant, showing that the trade-off between selectivity and tolerance does not fundamentally depend on the choice of metric.

Supp. Table 4

Correlation between selectivity and tolerance properties after subtraction of minimal rates

Th (spikes/s)	Position tolerance	(n)	Size tolerance	(n)	Contrast tolerance	(n)	Clutter tolerance	(n)
Correlation between _sparseness_ and tolerance properties (after subtraction of minimal rates)								
10	-0.39***± 0.08	(81)	-0.29*± 0.14	(52)	-0.19*± 0.1	(85)	-0.32**± 0.09	(88)
30	-0.35**± 0.11	(58)	-0.32*± 0.16	(42)	-0.16 ± 0.12	(60)	-0.42***± 0.08	(63)
Correlation between _morph tuning_ and tolerance properties (after subtraction of minimal rates)								
10	-0.22 ± 0.11	(42)	-0.5**± 0.15	(26)	-0.47***± 0.11	(48)	-0.31*± 0.09	(48)
20	-0.25 ± 0.11	(39)	-0.47**± 0.15	(25)	-0.48***± 0.12	(45)	-0.26*± 0.11	(45)

Supplemental Table 4. Correlation between different selectivity metrics and tolerance properties after subtraction of minimal rates. For each neuron, sparseness, morph tuning and each of the tolerance metrics were computed using firing rates that were corrected by subtracting the minimal neuronal response across the fixed set of 213 stimuli (see Supp. Fig. 1). The first column (Th) shows the minimal spike rate each neuron had to fire to the reference object in order to be included in the analysis. Overall, a negative correlation was always found between each of the selectivity and tolerance metrics, and most of these correlations were significant (* $p \le 0.05$; ** $p \le 0.01$; ***$p \le 0.001$; one-tailed permutation test; SE computed by bootstrap) and similar to what obtained using raw rates (Fig. 4C).

Supp. Table 5

Correlation between different tolerance properties (Th ≥ 10 spikes/s)

A	Size tolerance	(n)	Contrast tolerance	(n)	Clutter tolerance	(n)
Position tolerance	0.15 ± 0.16	(48)	0.4***± 0.12	(72)	0.06± 0.1	(75)
Size tolerance			0.15 ± 0.13	(47)	0.29*± 0.11	(52)
Contrast tolerance					0.13 ± 0.09	(83)

Correlation between different tolerance properties (Th ≥ 30 spikes/s)

B	Size tolerance	(n)	Contrast tolerance	(n)	Clutter tolerance	(n)
Position tolerance	0.29* ± 0.14	(39)	0.37**± 0.14	(52)	0.13 ± 0.12	(55)
Size tolerance			0.14 ± 0.13	(38)	0.22 ± 0.14	(42)
Contrast tolerance					0.1 ± 0.11	(59)

Supplemental Table 5. Correlation between different tolerance properties. The minimal spike rate each neuron had to fire to the reference object in order to be included in the analysis was selected to be either 10 spikes/s (**A**) or 30 spikes/s (**B**). The asterisks indicate the significance level (* $p ≤ 0.05$; ** $p ≤ 0.01$; one-tailed permutation test; SE computed by bootstrap).

Supp. Table 6

A. Correlation between sparseness and different temporal properties of the response

Th (spikes/s)	(n)	Latency of response onset	Latency of response peak	Response duration
		Background corrected response is at least 10% of its peak value		
10	(91)	0.19*± 0.11	0.09 ± 0.11	0.23*± 0.09
30	(64)	0.27*± 0.12	0.15 ± 0.12	0.13± 0.1
		Background corrected response is at least 20% of its peak value		
10	(91)	0.28**± 0.1	above	0.21*± 0.1
30	(64)	0.32**± 0.11	above	0.15 ± 0.11

B. Average values of the temporal properties of the response for the populations of weakly and highly selective cells

Th (spikes/s)	(n)	Latency of response onset (ms)	Latency of response peak (ms)	Response duration (ms)
	weak sel. / high sel.	weak sel. / high sel.	weak sel. / high sel	weak sel. / high sel
		Background corrected response is at least 10% of its peak value		
10	(38 / 17)	95±3 / 104±5	137±4 / 141±6	124±7 / 155±10 (**)
30	(25 / 11)	93±4 / 108±6 (*)	135±4 / 147±6	122±8 / 140±12
		Background corrected response is at least 20% of its peak value		
10	(38 / 17)	97±3 / 108±5 (*)	above	122±7 / 151±11 (*)
30	(25 / 11)	98±3.5 / 112±5 (*)	above	118±8 / 136±12

Supplementary Table 6. Relationship between object selectivity and time properties of the neuronal responses. ***A,*** Correlation coefficients (± SE) between sparseness and: i) latency of the onset of the response; ii) latency of the peak of the response; and iii) duration of the response (* $p \leq 0.05$; ** $p \leq 0.01$; one-tailed permutation test; SE computed by bootstrap). ***B,*** Average values (± SE) of response onset latency, response peak latency and response duration for the two populations of weakly and highly selective neurons (as defined in Fig. 5). Onset and offset of neuronal responses were computed using the algorithm described in Methods and Supp. Material 1.A., i.e., by considering epochs in which the background corrected response was at least either 10% or 20% of its peak value. Asterisks indicate that a significant difference was observed between weakly and highly selective neurons (* $p \leq 0.05$; ** $p \leq 0.01$; one-tailed t-test). The first column (Th) in ***A*** and ***B*** shows the minimal spike rate each neuron had to fire to the reference object in order to be included in the analysis.