



Published in final edited form as:

Science. 2008 September 12; 321(5895): 1502–1507. doi:10.1126/science.1160028.

Unsupervised natural experience rapidly alters invariant object representation in visual cortex

Nuo Li and James J. DiCarlo

McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

Object recognition is computationally challenging because each object produces myriad retinal images. Neurons at the highest cortical stage of the primate ventral visual stream (inferior temporal cortex; IT) likely underlie the ability of the visual system to tolerate that image variation -- their responses are selective to different objects, yet tolerant (“invariant”) to changes in object position, scale, pose, etc. (1-6). Understanding object recognition will require solving the mystery of how the brain constructs this neuronal tolerance. Here we report a novel instance of neuronal learning that suggests the underlying solution. Specifically, we show that targeted alteration of the natural temporal contiguity of unsupervised visual experience causes specific changes in the position tolerance (“invariance”) of IT neuronal selectivity. This unsupervised temporal tolerance learning (UTL) is substantial, increases with experience, and is significant even in single IT neurons after just one hour. Coupled with previous theoretical work (7-9) and the finding that this same experience manipulation changes the position tolerance of human object perception (10), we speculate that UTL may reflect the mechanism by which the ventral visual stream builds and maintains tolerant object representations. The relatively fast time-scale and unsupervised nature of UTL open the door to advances in systematically characterizing the spatiotemporal image statistics that drive it, understanding if it plays a role in other types of tolerance, and perhaps connecting a central cognitive ability -- tolerant object recognition -- to cellular and molecular plasticity mechanisms.

When presented with a visual image, primates can rapidly (<200 ms) recognize objects even in the face of large variation in object position, scale, pose, etc., even without attentional pre-cueing (11,12). This ability likely derives from the responses of neurons at high levels of the primate ventral visual stream (1,3,13). But how are these powerful “invariant” neuronal object representations built by the visual system? Based on theoretical (e.g. (7,8)) and behavioral (9,10) work, one possibility is that tolerance (“invariance”) is learned from the temporal contiguity of object features during natural visual experience, potentially in an unsupervised manner. To look for evidence of that hypothesis, we focused on arguably the simplest tolerance that the visual system achieves -- position tolerance. Our overarching logic was as follows (also see (10)): During natural visual experience, objects tend to remain present for seconds or more, while object motion or viewer motion (e.g. eye movements) tend to cause rapid changes in the retinal image cast by each object over shorter time intervals (hundreds of ms). In theory, the ventral visual stream could construct a position-tolerant object representation by taking advantage of this natural tendency for temporally contiguous retinal images to belong to the same object. If so, it might be possible to uncover the signature of this learning by using targeted alteration of those spatiotemporal statistics. In particular, if two objects consistently swapped identity across temporal-contiguous

changes in retinal position then, following sufficient experience in this “altered” visual world, the visual system might incorrectly associate the neural representations of those objects viewed at different positions into a single object representation. To look for this signature, we focused on the top level of the primate visual object recognition pathway (IT). In the adult brain, many individual IT neurons have somehow achieved neuronal position tolerance -- they respond preferentially to different objects, their object selectivity is largely maintained across changes in object retinal position, even when images are simply presented to a fixating animal (5,6). Thus, our key prediction was that targeted alteration of natural temporal contiguity of real-world visual experience would produce a specific change in IT position tolerance (see Fig. 1C).

In this study, we tested a strong, “online” form of the temporal contiguity hypothesis -- we allowed two monkeys to visually explore an altered visual world for up to two hours (described below), while we intermittently tested the position tolerance of their IT neurons to look for any ongoing change in tolerance produced by that altered experience. Specifically, once we isolated a neuron, we determined its object selectivity among a set of 100 objects (Supplemental Fig. S2), and we selected two objects that elicited strong (object “P”, preferred) and moderate (object “N”, non-preferred) spiking responses from the neuron. We then tested the position-tolerance of that object selectivity (*Test Phase* in Fig. 1A) by briefly presenting each object at 3° above, below or at the center of gaze (see (14) and Fig. S1). (All neuronal data reported in this study were obtained in the *Test Phase*: task unrelated to the test stimuli, no attentional cuing, and completely randomized presentations of test stimuli, see (14)). Following the first *Test Phase*, we allowed the animal to explore the statistically altered visual world for ~15 minutes (*Exposure Phase*), and we then re-tested the position tolerance (*Test Phase*). We continued to alternate between these two phases (*Test Phase* ~5 min; *Exposure Phase* ~15 minutes) until neuronal isolation was lost. In formulating our study, we reasoned that, even if the temporal contiguity hypothesis were true, the ~1 hr period we could isolate each IT neuron might not allow us to provide enough experience to produce a strong change in its position tolerance. Thus, we designed our experiments so that, the more neurons we studied, the more power we gained to test our hypothesis (i.e. a pooled neuronal “subject” design).

To create the altered visual world (*Exposure Phase* in Fig. 1A), we focused on just three retinal positions (the center of gaze, 3° below, and 3° above) and two objects for each neuron (P and N). We took advantage of the fact that changes in the position of an object’s retinal image occur naturally as rapid eye movements sample the visual scene (saccades) -- each monkey freely viewed the video monitor on which isolated objects appeared intermittently, and its only task was to freely look at each object. This exposure “task” is a natural, automatic primate behavior in that it requires no training. Importantly however, by using real-time eye-tracking ((15)), the images that played out on the monkey’s retina during exploration of this world were under precise experimental control (see (14)). Specifically, even though the monkey was free-viewing, the objects were placed on the video monitor so as to (initially) cast their image at one of two possible *retinal* positions (+3° or -3°). One of these retinal positions was pre-chosen for targeted alteration in visual experience (the “swap” position; counterbalanced across neurons, see Fig. 1B and (14)); the other position acted as a control (the “non-swap” position). Of course, the monkey quickly saccaded to each object (mean: 108 ms after object appearance), which rapidly brought the object image to the center of its retina (mean saccade duration 23 ms). For cases in which the object had appeared at the “non-swap” position, its identity remained stable even as the monkey saccaded to it, typical of real-world visual experience (“Normal exposure”, Fig. 1A, see (14)). Critically however, for cases in which the object had appeared at the “swap” position, it was always replaced by the other object (e.g. P->N) precisely as the monkey saccaded to it (Fig. 1A, “Swap exposure”). This manipulation of experience took advantage of the fact that

primates are effectively blind during the brief time it takes to complete a saccade (16). Most importantly, it consistently made the image of one object at a peripheral retinal position (“swap” position) temporally contiguous with the retinal image of the other object at the center of the retina (see Fig. 1A,B).

The main prediction of the temporal contiguity hypothesis is that, confronted with this altered world, each IT neuron that prefers object P over object N should tend to change its initial position-tolerance (i.e. its ability to maintain that object selectivity across retinal position) in a specific way: it should tend to alter its object selectivity primarily at the “swap” position so that it begins to respond less to object P and more to object N (given enough experience, the neuron might even begin to reverse its object selectivity; see Fig. 1C). Indeed, we found that, the longer the monkey was exposed to the altered visual world, the more its IT neurons changed their position tolerance in the predicted manner (n=50 IT neurons in Monkey 1; 51 in Monkey 2). Specifically, for each neuron, we measured its object selectivity at each position as the difference in response to the two objects (P-N; all key effects reported below were also found with a contrast index of selectivity, see Supplemental Fig. S6). At the “swap” position, IT neurons (on average) decreased their initial object selectivity for P over N, and this change in object selectivity grew monotonically stronger with increasing numbers of “swap” exposure trials (Fig. 2). However, while that change was occurring, the same IT neurons showed no average change in their object selectivity at the control (“non-swap”) position, and little change in their object selectivity among two other control objects that were never shown to the animals during *Exposure Phase* (Fig. 2A; also see below).

Because each IT neuron was tested for different amounts of exposure time, to quantify the IT object selectivity change across the entire single-unit population (n=101), we computed object selectivity change, $\Delta(P-N)$, between the first and last available *Test Phase* for each IT neuron. Following the logic outlined above, the prediction is that $\Delta(P-N)$ should be negative (i.e. in the direction of object preference reversal), and greatest at the “swap” position. Again, we found that, on average across the IT population, this prediction was born out by the data (Fig. 3A). The position-specificity of the experience-induced changes in object selectivity was confirmed by two different statistical approaches: 1) direct comparison of $\Delta(P-N)$ between the swap and non-swap position (n=101; p=0.005, one tailed paired t-test); 2) a significant interaction between position and exposure – that is, object selectivity decreased (i.e. moved in the direction of object preference reversal) at the swap position with increasing amounts of exposure (p=0.009 by one-tailed bootstrap; p=0.007 by one-tailed permutation test; tests were done on (P-N), see details in Supplementary).

The changes in object selectivity at the swap position were not only position-specific, but were largely shape-specific. For 88 of the 101 neurons, we monitored the neuron’s selectivity among two control objects not shown to the animal during the *Exposure Phase* (chosen in similar way as the P and N objects, fully-interleaved testing in each *Test Phase*; see (14)). Across the IT population, control object selectivity at the swap position did not significantly change (Fig. 2A), and swap object selectivity changed significantly more than control object selectivity (Fig. 3B; n=88 neurons, p = 0.009 one tailed paired t-test of swap vs. control objects at swap position).

These changes in object selectivity were substantial in magnitude and grew ever larger with increasing amounts of exposure (average change of ~5 spikes/s per 400 exposures at swap position; Figs. 2C, 3C), and, at the population level were visibly obvious and highly significant (above). In the face of well-known Poisson spiking variability (17,18), these effects were only weakly visible in most single IT neurons recorded for short durations, but were much more apparent over the maximal one hour exposure time that we could hold

these isolated neurons (Fig. 2C, lower panels). So we asked, would the object selectivity change at the swap position continue to grow even larger with longer periods of exposure? To answer this question, we recorded multi-unit activity (MUA) during ten new experiment sessions in one animal (Monkey 2), which allowed us to record from a number of (non-isolated) neurons around the electrode tip (which all tend to have similar selectivity(19,20), see (14)) while the monkey was exposed to the altered visual world for the entire experimental session (~2+ hrs). Just like the single-unit data, the MUA data showed a change in object selectivity only at the swap position (Fig. 2C; “position \times exposure” interaction: $p=0.03$ one-tailed bootstrap; $p=0.014$ one-tailed permutation test, $n=10$ sites). This shows that just ten MUA recording sessions are sufficient to replicate the main effect we first observed in the single-unit population. Furthermore, the MUA data revealed that the object selectivity change at the swap position continued to increase as the animal received even more exposure to the altered visual world, followed a very similar time course in the *rate* of object selectivity change (~5 spikes/s per 400 exposures; see Fig. 3C), and even showed a slight reversal in object selectivity ($N>P$ in Fig. 4D).

Our main results were similar in magnitude (Fig. 3A, B) and statistically significant in each of the two monkeys (Monkey 1: $p=0.019$; Monkey 2: $p=0.0192$; one-tailed t-test). Because each monkey performed a different task during the *Test Phase* (unrelated to the rapidly-presented test stimuli; see (14)), this suggests that these neuronal changes are not task dependent. Moreover, the response changes were present in the earliest IT spikes (~100ms, see Supplemental Fig. S4). Thus the observed learning most likely reflects changes in the feed-forward response properties of the ventral stream.

Because we selected the objects P and N so that they both tended to drive the neuron (see (14)), the population distribution of selectivity for P and N at each position was very broad (95% range: [-5.7 spikes/s to 31.0 spikes/s] pooled across position; $n=101$). However, our prediction (and result, above) assumes that the IT neurons we recorded from were initially object-selective (i.e. response to object P greater than object N). Consistent with this, we found that neurons in our population with no initial object selectivity at the center of gaze showed little average change in object selectivity with exposure (see Supplemental Fig. S5). To test the learning effect in the most selective IT neurons, we selected the neurons with significant object selectivity ($n=52$ of 101; two-way ANOVA test (2 objects \times 3 positions), $p<0.05$, significant main object effect or interaction). Even among just these neurons, the learning effect remained highly significant and still specific to the swap position ($p=0.002$ by t-test; $p=0.009$ by bootstrap; $p=0.004$ by permutation test; see above).

To further characterize the changes in response to individual objects, we closely examined the selective neurons held for at least 300 exposures ($n=28/52$) and the multi-unit sites ($n=10$). For each neuron/site, we used linear regression to measure any trend in the neuron’s responses to each object as a function of exposure time (Fig 4A). Changes in response to P and N at the swap position were visibly obvious in a fraction of single neurons/sites (Fig. 4A), and statistically significant object selectivity change was encountered on 32% of instances (12/38; Fig. 4C; see (14)). Across our neuronal population, the change in object selectivity at the swap position was due to both a decreased response to object P and an increased response to object N (approximately equal change; Fig. 4B). These response changes are highly visibly in the single-units and multi-units held for the longest exposure times (Fig. 4D).

These neuronal changes in the position profile of IT object selectivity (i.e. position tolerance) cannot be explained by changes in attention or retinotopic adaptation. First, a simple fatigue-adaptation model cannot explain the position-specificity of the changes because, during the recording of each neuron, each object was experienced equally often at

the “swap” and “non-swap” positions. Second, we measured these object selectivity changes with briefly presented, fully randomized stimuli while the monkeys performed tasks unrelated to the stimuli (see (14)), arguing against an attentional account. Third, both of these explanations predict a decrease in response to *all* objects at the swap position, yet we found that the change in object selectivity at the swap position was due to an *increase* in response to object N (+2.3 spikes/s per 400 swap exposures) as well as a decrease in response to object P (−3.0 spikes/s per 400 swap exposures; see Fig. 3D, S4). Fourth, neither possibility can explain the shape-specificity of the changes. In sum, even when the animal is doing a task unrelated to the test objects, we used briefly presented stimuli to reveal position-specific, shape-specific changes in IT responses to those objects consisting of increased responses to some objects (N) and decreased responses to others (P). This cannot be explained by any known form of “adaptation” or “attention”.

We term this effect “unsupervised temporal tolerance learning” (UTL), because the tolerance changes depend on the temporal contiguity of object images on the retina. Although, such temporal contiguity of images on the retina may be all that is required to drive UTL, our current data cannot rule out the possibility that the brain’s saccade generation mechanisms (or the associated attentional mechanisms (21,22)) may also be needed (e.g. to internally signal “when” to learn and the saccade target/direction, which could make tolerance learning more efficient). Interestingly, eye-movement signals are present in the ventral stream (23,24), and our previous results showing alteration in human perception with a similar paradigm (see below) suggested that temporal contiguity alone was not sufficient (10). The relatively fast time-scale and unsupervised nature of UTL may allow rapid advances in answering these questions, systematically characterizing the spatiotemporal sensory statistics that drive it, and understanding if and how it extends to other types of image tolerance (e.g. changes in object scale, pose (25,26)).

It has been shown that IT neurons “learn” to give similar responses to different visual shapes (“paired associates”) when reward is used to *explicitly* teach monkeys to associate those shapes over long time scales (1-5 sec between images, e.g.(27,28)), in some cases without explicit instruction (29,30). UTL might be an instance of the same underlying plasticity mechanisms: here the “associations” are between object images at different retinal positions (which, in the real world, are typically images of the same object). However, UTL is qualitatively different – the changes here are position-specific, do not require external supervision (the “reward” was only used to motivate visual exploration), and operate over the much shorter time scales of natural visual exploration (~200 ms). These distinctions are important because we naturally receive orders of magnitude more such experience (e.g. $\sim 10^8$ unsupervised temporal-contiguity saccadic “experiences” in just one year of life).

In summary, our results show that targeted alterations in natural, unsupervised visual experience change the position tolerance of IT neurons in the manner predicted by the hypothesis that the brain employs a temporal contiguity learning strategy to build that tolerance in the first place. That is, visual features that co-occur across short time intervals are, on average, likely to correspond to different images of the same object. We do not yet know if UTL reflects mechanisms that are fundamental to building tolerant representation. However, we have previously found these same experience manipulations change the position tolerance of human object perception -- producing a tendency to (e.g.) perceive one object to be the same identity as another object across a “swap” position (10). Moreover, given that the animal had a lifetime of visual experience to potentially build its IT position tolerance, the strength of UTL is substantial (~5 spikes/s change per hour). For comparison, the change in IT selectivity over just one hour is comparable to attentional effect sizes (31), and is more than double that observed in previous IT learning studies over much longer training intervals (32-34). We do not yet know how far we can push this learning, but we se

that just two hours of (highly targeted) unsupervised experience begins to reverse the object preferences of IT neurons (Fig. 4D). This discovery re-emphasizes the importance of plasticity in vision (1,30,32,33,35-38) by showing that it extends to a bedrock property of the adult ventral visual stream -- position tolerant object selectivity (also see (39-41)), and studies along the post-natal developmental time line are now needed.

Several computational models have shown how temporal contiguity strategies can build tolerance (7-9), and these models can be implemented using Hebbian-like learning rules (9) that are consistent with spike-timing-dependent plasticity (42). For example, one can imagine IT neurons using almost-temporally-coincident activity to learn which sets of its afferents correspond to features of the same object at different positions. The time-course and task-independence of UTL is consistent with synaptic plasticity (36,43), but our data do not constrain the locus of plasticity. Based on previous work, changes at multiple levels of the ventral visual hierarchy are likely (37,44). In sum, we speculate that UTL may be a neuronal learning signature of underlying brain algorithms for building position tolerant object representation, and it may provide a new bridge to connect a central cognitive ability -- tolerant object recognition -- to cellular and molecular plasticity mechanisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

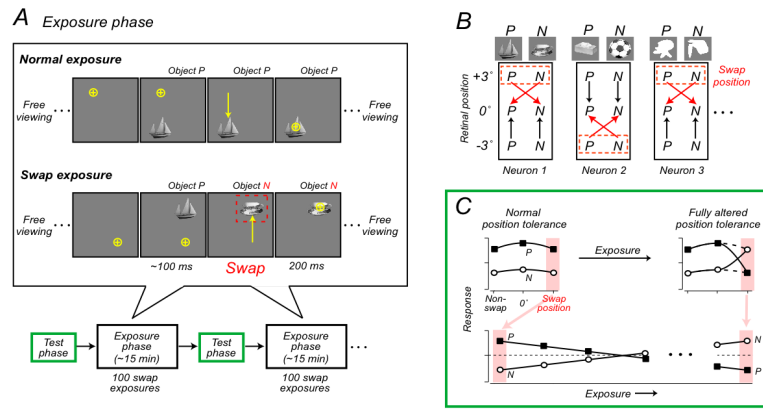
Acknowledgments

We thank D. Cox, R. Desimone, N. Kanwisher, J. Maunsell and N. Rust for helpful comments and discussion, and J. Deutsch, B. Kennedy, M. Maloof and R. Marini for technical support. This work was supported by the National Institutes of Health (NIH-R01-EY014970) and The McKnight Endowment Fund for Neuroscience.

References

1. Logothetis NK, Sheinberg DL. *Ann. Rev. Neurosci.* 1996; 19:577. [PubMed: 8833455]
2. Tanaka K. *Annual Review of Neuroscience.* 1996; 19:109.
3. Hung CP, Kreiman G, Poggio T, DiCarlo JJ. *Science.* Nov 4.2005 310:863. [PubMed: 16272124]
4. Brincat SL, Connor CE. *Nat Neurosci.* Aug.2004 7:880. [PubMed: 15235606]
5. Ito M, Tamura H, Fujita I, Tanaka K. *Journal of Neurophysiology.* 1995; 73:218. [PubMed: 7714567]
6. Op de Beeck H, Vogels R. *J Comp Neurol.* 2000; 426:505. [PubMed: 11027395]
7. Wiskott L, Sejnowski TJ. *Neural Comput.* Apr.2002 14:715. [PubMed: 11936959]
8. Foldiak P. *Neural Computation.* 1991; 3:194.
9. Wallis G, Rolls ET. *Progress in Neurobiology.* 1997; 51:167. [PubMed: 9247963]
10. Cox DD, Meier P, Oertelt N, DiCarlo JJ. *Nat Neurosci.* Sep.2005 8:1145. [PubMed: 16116453]
11. Thorpe S, Fize D, Marlot C. *Nature.* 1996; 381:520. [PubMed: 8632824]
12. Potter MC. *J Exp Psychol [Hum Learn].* Sep.1976 2:509.
13. Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. *Nature.* Jun 23.2005 435:1102. [PubMed: 15973409]
14. Supplementary Methods.
15. DiCarlo JJ, Maunsell JHR. *Nat Neurosci.* 2000; 3:814. [PubMed: 10903575]
16. Ross J, Morrone MC, Goldberg ME, Burr DC. *Trends Neurosci.* 2001; 24:113. [PubMed: 11164942]
17. Tolhurst DJ, Movshon JA, Dean AF. *Vision Res.* 1983; 23:775. [PubMed: 6623937]
18. Shadlen MN, Newsome WT. *J. Neuroscience.* 1998; 18:3870.
19. Tanaka K. *Cereb Cortex.* Jan.2003 13:90. [PubMed: 12466220]

20. Kreiman G, et al. *Neuron*. Feb 2.2006 49:433. [PubMed: 16446146]
21. Moore T, Fallah M. *Proc Natl Acad Sci U S A*. Jan 30.2001 98:1273. [PubMed: 11158629]
22. Kowler E, Anderson E, Doshier B, Blaser E. *Vision Res*. 1995; 35:1897. [PubMed: 7660596]
23. Ringo JL, Sobotka S, Diltz MD, Bunce CM. *Journal of Neurophysiology*. 1994; 71:1285. [PubMed: 8201422]
24. Moore T, Tolias AS, Schiller PH. *Proc Natl Acad Sci U S A*. Jul 21.1998 95:8981. [PubMed: 9671790]
25. Edelman S, Duvdevani-Bar S. *Neural Comput*. 1997; 9:701. [PubMed: 9161020]
26. Wallis G, Bulthoff H. *Trends Cogn Sci*. Jan.1999 3:22. [PubMed: 10234223]
27. Sakai K, Miyashita Y. *Nature*. 1991; 354:152. [PubMed: 1944594]
28. Messinger A, Squire LR, Zola SM, Albright TD. *Proc Natl Acad Sci U S A*. Oct 9.2001 98:12239. [PubMed: 11572946]
29. Miyashita Y. *Nature*. 1988; 335:817. [PubMed: 3185711]
30. Erickson CA, Desimone R. *J Neurosci*. 1999; 19:10404. [PubMed: 10575038]
31. Maunsell JHR, Cook EP. *Philos Trans R Soc Lond B Biol Sci*. Aug 29.2002 357:1063. [PubMed: 12217174]
32. Baker CI, Behrmann M, Olson CR. *Nat Neurosci*. Nov.2002 5:1210. [PubMed: 12379864]
33. Kobatake E, Wang G, Tanaka K. *Journal of Neurophysiology*. 1998; 80:324. [PubMed: 9658053]
34. Sigala N, Gabbiani F, Logothetis NK. *J Cogn Neurosci*. Feb 15.2002 14:187. [PubMed: 11970785]
35. Rolls ET, Baylis GC, Hasselmo ME, Nalwa V. *Exp Brain Res*. 1989; 76:153. [PubMed: 2753096]
36. Meliza CD, Dan Y. *Neuron*. Jan 19.2006 49:183. [PubMed: 16423693]
37. Yang T, Maunsell JH. *J Neurosci*. Feb 18.2004 24:1617. [PubMed: 14973244]
38. Seitz A, Watanabe T. *Trends Cogn Sci*. Jul.2005 9:329. [PubMed: 15955722]
39. Dill M, Fahle M. *Perception & Psychophysics*. 1998; 60:65. [PubMed: 9503912]
40. Dill M, Edelman S. *Perception*. 2001; 30:707. [PubMed: 11464559]
41. Nazir TA, O'Regan JK. *Spat Vis*. 1990; 5:81. [PubMed: 2090197]
42. Sprekeler H, Michaelis C, Wiskott L. *PLoS Comput Biol*. Jun 29.2007 3:e112. [PubMed: 17604445]
43. Markram H, Lubke J, Frotscher M, Sakmann B. *Science*. Jan 10.1997 275:213. [PubMed: 8985014]
44. Kourtzi Z, DiCarlo JJ. *Curr Opin Neurobiol*. Apr.2006 16:152. [PubMed: 16563736]

**Fig. 1.**

Experimental design and predictions. **(A)** For each visually responsive neuron encountered, a preferred object (P) and less-preferred object (N) were chosen from a set of 100 objects. We then alternated between a *Test Phase* and *Exposure Phase* for as long as we could record from the neuron. In the *Test Phase*, the neuron's response to P and N was measured with fully randomized images while each monkey performed a task unrelated to the object images (see (14)). In the *Exposure Phase*, free viewing monkeys spontaneously saccaded to objects that initially appeared either 3° above or below the center of gaze. For half of these appearances (“Normal exposure”), the peripheral object remained unchanged as the monkey's natural eye movement brought it to the center of its retina (until it was removed 200 ms later). For the other half (randomly interleaved; “Swap exposure”), the peripheral object (e.g. P) was always replaced by the other object (e.g. N) as the monkey brought it to the center of its retina (see (14)). All aspects of the stimulus presentation were otherwise identical for the two exposure types. Stimulus size was 1.5°, and is only schematic in this figure (see Supp. Fig. S2). Each *Exposure Phase* consisted of 100 normal exposures (50 each of P->P and N->N) and 100 “swap” exposures (50 each of P->N and N->P). **(B)** Each box shows the *Exposure Phase* design for a single neuron. Arrows show the saccade-induced temporal contiguity of retinal images (arrow head points to the retinal image occurring later in time, i.e. at the end of the saccade). Note the crossed red arrows from the “swap” position (red) indicating the key manipulation -- experience with spatiotemporal image statistics that are consistently different than the real-world (non-crossed) black arrows. The “swap” position was strictly alternated (neuron-by-neuron) so that it was perfectly counterbalanced across neurons. **(C)** Prediction: if the visual system builds tolerance using temporal contiguity (here driven by saccades), the “swap” exposure should cause incorrect grouping of two different object images (here P and N). Thus, the predicted effect is a decrease in object selectivity at the “swap” position (red dot) that increases with increasing exposure (in the limit, reversing object preference), and little or no change in object selectivity at the “non-swap” position.

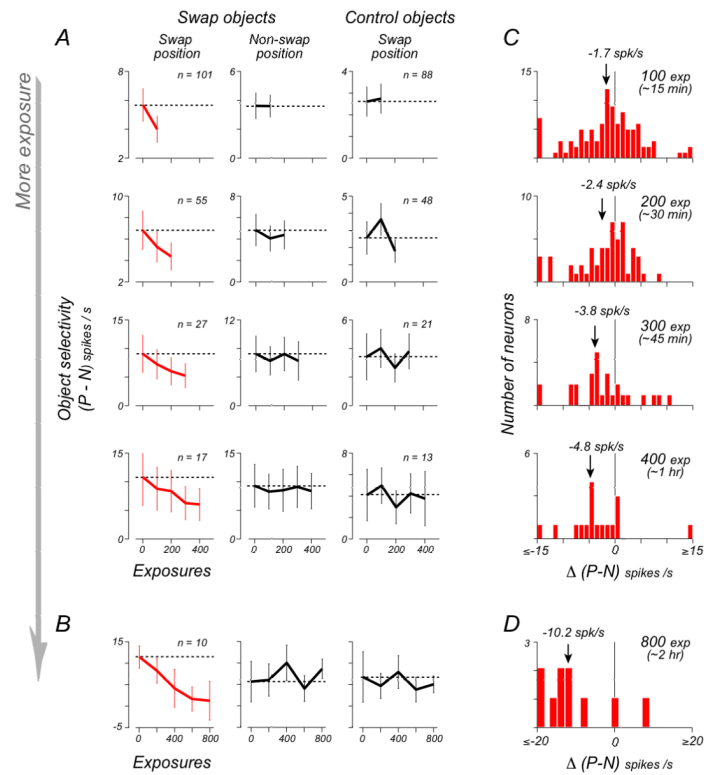


Fig. 2. Change in the population object selectivity. **(A)** Mean population object selectivity at the swap and non-swap position, and for the control objects at the swap position. The control objects were never shown to the animals during the *Exposure Phase*. Neuronal responses to all objects and positions were obtained during fully-randomized testing in the *Test Phase* (see Fig. 1). Each row of plots show neurons held for different amount of time, (e.g. the top row shows all neurons held for over 100 swap exposures -- including the neurons from the lower rows; the second row shows the neurons held for over 200 exposures; etc). The object selectivity for each neuron was the difference in its response to object P and N. To avoid any bias in this estimate, for each neuron we defined the labels “P” and “N” (P is the preferred object in summed response across position) by splitting the pre-exposure response data and used a portion of it (10 repetitions) to determine these labels. All remaining data were used to compute the displayed results in all analyses using these labels. **(B)** Mean population object selectivity of ten multi-unit sites plotting in the same format as **(A)**. Error bars **(A, B)** are standard error of the mean. **(C)** Histograms of the object selectivity change at the swap position, $\Delta(P-N) = (P - N)_{\text{post-exposure}} - (P - N)_{\text{pre-exposure}}$. The arrow and text indicate the mean of the distribution. The mean $\Delta(P-N)$ at the non-swap position was: -0.01 , -0.5 , -0.9 , -0.9 spikes/s respectively (not shown). **(D)** Object selectivity change at multi-unit sites. The mean $\Delta(P-N)$ at the non-swap position was 1.6 spikes/s (not shown).

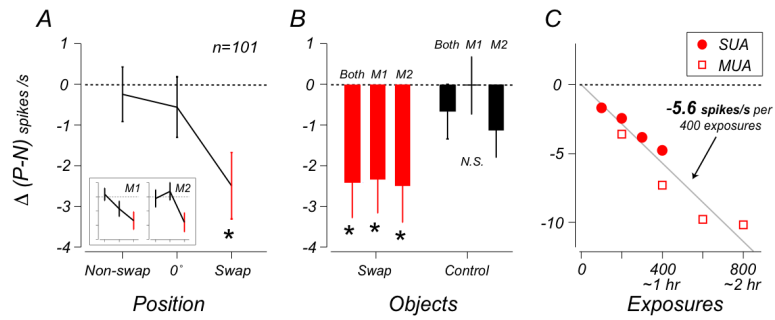
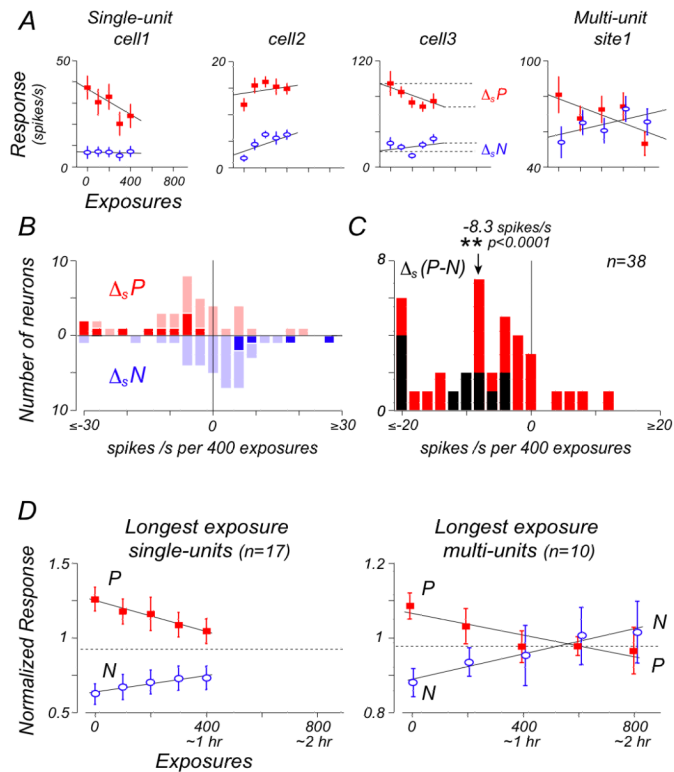


Fig. 3.

Position-, object-specificity and time course. **(A)** Mean object selectivity change, $\Delta(P-N)$, at the swap, non-swap, and central (0°) retinal position; data from all neurons are included (by using each neuron's last available *Test Phase*, mean ~ 200 swap exposures). $\Delta(P-N)$ was computed as in Fig. 2c. The insets show the same analysis performed separately for each monkey. **(B)** Mean object selectivity change for the (exposed) swap objects and (non-exposed) control objects at the swap position. Error bars (**A**, **B**) are standard error of the mean. The swap object selectivity change at the swap position is statistically significant (*) in the pooled data as well as in individual animals ($p < 0.05$, one-tailed t-test against 0). **(C)** Mean object selectivity change as a function of the number of swap exposures for all single-units ($n=101$) and multi-units ($n=10$). Each data point shows the average across all the neurons/sites held for a particular amount of time. Gray line is best linear fit with a zero intercept; slope is mean effect size: -5.6 spikes/s per 400 exposures. The slope at the non-swap position using the same analysis was 0.6 spikes/s (not shown).

**Fig. 4.**

Responses to objects P and N. **(A)** Response data to object P and N at the swap position for three example neurons and one multi-unit sites as a function of exposure time. The solid line is the best-fit linear regression. The slope of each line (ΔP and ΔN) provided a measure of the change in response to object P and N for each neuron. Some neurons showed a response decrease to P, some showed a response enhancement to N, while others showed both (see examples). **(B)** Histograms of the slopes obtained for the object selective neurons/sites tested for at least 300 exposures (see main text). The slope values indicate the change in response per 400 exposures (~ 1 hr). The dark-colored bars indicate neurons with significant change in response by a permutation test ($p < 0.05$; see (14)). **(C)** Histograms of the slopes from linear regression fits to object selectivity (P-N) as a function of exposure time, same units as in **(B)**. Arrow indicates the mean of the distribution, (for comparison, the mean $\Delta(P-N)$ at the non-swap position was -1.7 spikes/s, $p=0.38$). The black bars indicate instances (32%; 12/38) that showed a significant change in object selectivity by permutation test ($p < 0.05$). **(D)** Data from all the neurons/sites tested for the longest exposure time. The plot shows the mean normalized response to object P and N as a function of exposure time. (see Supplemental Fig. S3 for data at the non-swap position and for the control objects). Error bars **(A, D)** are standard error of the mean.