Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells

Giulio Matteucci[1] and Davide Zoccolan[1,*]

[1]Visual Neuroscience Lab, International School for Advanced Studies (SISSA), Trieste, Italy

*Correspondence should be addressed to: Davide Zoccolan, International School for Advanced Studies (SISSA), Via Bonomea, 265, 34136 Trieste (TS), ITALY, Phone: +39 040 3787 701, E-mail: zoccolan@sissa.it.

ABSTRACT

Unsupervised adaptation to the spatiotemporal statistics of visual experience is a key computational principle that has long been assumed to govern postnatal development of visual cortical tuning, including orientation selectivity of *simple cells* and position tolerance of *complex cells* in primary visual cortex (V1). Yet, causal empirical evidence supporting this hypothesis is scant. Here we show that degrading the temporal continuity of visual experience during early postnatal life leads to a sizable reduction of the number of complex cells and to an impairment of

their functional properties, while fully sparing the development of simple cells. This causally implicates adaptation to the temporal structure of the visual input in the development of transformation tolerance but not of shape tuning, thus tightly constraining computational models of unsupervised cortical learning.

INTRODUCTION

It has long been proposed that the tuning of sensory neurons is determined by adaptation to the statistics of the signals they need to encode (*1*, *2*). In the visual domain, this notion has given rise to two broad families of unsupervised learning algorithms: those relying on the spatial structure of natural images, referred to as *unsupervised spatial learning* (USL) models (*1–6*); and those leveraging the spatiotemporal structure of natural image sequences, referred to as *unsupervised temporal learning* (UTL) models (*7–15*). Both kinds of learning have been applied to explain the ability of visual cortical representations to selectively code for the identity of visual objects, a property known as shape tuning, while tolerating variations in their appearance (e.g., because of position changes), a property known as transformation tolerance (or invariance) (*16*). These properties are built incrementally along the *ventral stream* (the cortical hierarchy devoted to shape processing), but the earliest evidence of shape tuning and invariance in the visual system can be traced back to primary visual cortex, where simple cells first exhibit tuning for non-trivial geometrical patterns (oriented edges) and complex cells first display some degree of position tolerance (*17*).

In sparse coding theories (arguably the most popular incarnation of USL), maximizing the sparsity of the representation of natural images produces Gabor-like edge-detectors that

closely resemble the receptive fields (RFs) of V1 simple cells (*5*, *6*). A similar outcome is achieved by maximizing the independence of the representation, e.g., through Independent Component Analysis (*18*). Other USL models, by optimizing objective functions that depend on the combination of several linear spatial filters, also account for the emergence of position-tolerant edge detectors, such as V1 complex cells (*3*, *4*). The latter, however, have been more commonly modeled as the result of UTL, where the natural tendency of different object views to occur nearby in time is used to factor out object identity from other faster-varying, lower-level visual attributes. Interestingly, while some UTL models presuppose the existence of a bank of simple cells, upon which the complex cells' representation is learned (*7*, *11–15*), other models, such as *slow feature analysis* (SFA), directly evolve complex cells from the pixel (i.e., retinal) representation, thus simultaneously learning shape selectivity and invariance (*8*, *9*).

To date, it remains unclear what role these hypothesized learning mechanisms play in the developing visual cortex, despite the influence that early visual experience is known to exert on cortical tuning. This is demonstrated (e.g.) by the impact of monocular deprivation on the development of ocular dominance (*19*, *20*), by the bias in orientation tuning produced by restricting early visual experience to a single orientation (*21–25*), and by the need, for ferret visual cortex, to experience visual motion in order to develop direction selectivity (*26–28*). However, none of these manipulations was designed to specifically test the role of USL and/or UTL in mediating the development of simple and complex cells. As a result, empirical support for the role of sparse coding in determining orientation selectivity is still indirect (*29*, *6*), as no study has succeeded in abolishing (or at least interfering with) the development of simple cells with Gabor-like tuning through manipulations of the visual environment (*30*). Similarly, no clean causal evidence has been gathered yet to demonstrate the involvement of UTL in postnatal

development of invariance and/or selectivity in visual cortex. The only experiments suggesting the involvement of UTL in fostering invariant visual object representations during development comes from behavioral studies of chicks' object vision (*31*, *32*). In mammals, a few studies based on strobe rearing did investigate the effect of degrading the temporal continuity of the visual input on the developing cortex (*33*–*39*), but they did not quantitatively probe whether such manipulation led to a reduction of invariance (see Discussion). More critically, strobe rearing does not allow effectively and selectively altering the temporal statistics of the visual input, while sparing the spatial statistics (or vice versa). In fact, short light flashes (≤10 μs) severely limit also the experience with the spatial content of the visual input, as well as the overall amount of light exposure during development, especially when combined with low strobe rates (0.5-2 Hz). Conversely, higher strobe rates (8 Hz) allow still experiencing a strongly correlated visual input over time, given the dense, ordered sampling of the visual space performed by the visual system across consecutive flashes. This makes it impossible to disentangle the contribution of USL, UTL or simpler light-dependent plasticity processes to the changes of orientation and/or direction tuning reported in some of these studies.

In summary, the lack of conclusive evidence about the involvement of spatial and temporal learning processes in cortical development of selectivity and invariance calls for new studies based on tighter, better controlled manipulations of visual experience during postnatal development.

RESULTS

Our study was designed to causally test the involvement of unsupervised temporal learning in the development of shape selectivity and transformation tolerance (i.e., simple and complex cells) in primary visual cortex. To this aim, we took 18 newborn rats (housed in light-proof cabinets from birth) and, from postnatal day (P)14 (i.e., at eye opening) through P60 [i.e., well beyond the end of the critical period (*40*)], we subjected them to daily, 4-hours-long exposures inside an immersive visual environment. This consisted of a rectangular, transparent basin, surrounded on each side by a computer-controlled LCD monitor and placed inside a light-proof cabinet (Fig. S1). Eight animals (the *control* group) were exposed to a battery of 16 natural movies (lasting from a few minutes to half an hour), while the remaining ten rats (the *experimental* group) were exposed to their frame-scrambled versions (Fig. 1A). As a result of the scrambling, the correlation between the frames of a movie as a function of their temporal separation was close to zero at all tested time lags, while the image frames of the original movies remained strongly correlated over several seconds (compare the orange vs. blue curves in Fig. 1B-C; the average time constants of the exponential fits to the correlation functions were 6.9±1.3 ms and 1.47±0.10 s, respectively, for the frame-scrambled and original movies; see Fig. 1C, right). All movies were played at 15 Hz, which is approximately half of the critical flicker fusion frequency (~30-40 Hz) of the rat (*41*). This ensured that, while the temporal correlation of the input was substantially broken, no fusion occurred between consecutive frames of a movie, thus allowing the rats of the experimental group to fully experience the spatial content of the individual image frames. This likely enabled the experimental rats to also experience some amount of continuous transformation (e.g., translation) of the image frames, as the result of spontaneous head or eye movements during the 66.7 ms presentation time of each frame. This, along with the presence of

some stable visual features in the physical environment (e.g., the dark edges of the monitors) and the possibility for the rats to see parts of their own body, allowed for some residual amount of temporal continuity in the visual experience of the experimental group. Such incomplete disruption of temporal continuity was unavoidable, given the constraints of: 1) granting the animals full access to the spatial content of natural visual scenes; and 2) trying to foster visual cortical development and plasticity by leaving the rats free to actively explore the environment (*42*, *43*), thus avoiding body restraint and head fixation. Crucially, despite these constraints, the temporal statistics of the visual stream experienced by the two groups of animals at time scales larger than 66.7 ms was radically different (Fig. 1B-C), while the spatial statistics and overall amount of light exposure were very well matched. This allowed isolating the contribution of temporal contiguity to the postnatal development of V1 simple and complex cells.

Postnatal rearing in a temporal discontinuous visual environment leads to a reduction of V1 complex cells, but leaves spatial tuning of V1 neurons unaltered.

Shortly after the end of the controlled-rearing period, we performed multi-channel extracellular recordings from V1 of each rat under fentanyl/medetomidin anesthesia (*44*, *45*) (see Materials and Methods for details). Our recordings mainly targeted layer 5, where complex cells are known to be more abundant (*46*), as well as layer 4, with the distributions of recorded units across the cortical depth and the cortical laminae being statistically the same for the control and experimental groups (Fig. S2). During a recording session, each animal was presented with drifting gratings spanning 12 directions (from 0° to 330° in steps of 30°), and with contrast-modulated movies of spatially and temporally correlated noise (*44*, *46*). Responses to the noise movies allowed inferring the linear RF structure of the recorded units using the spike-triggered

average (STA) analysis, as well as the temporal scale over which the stimulus representation unfolded (see Materials and Methods). Responses to the drifting gratings were used to estimate the tuning of the neurons with standard orientation and direction selectivity indexes (OSI and DSI; defined in Materials and Methods), as well as to probe their sensitivity to phase-shifts of their preferred gratings, thus measuring their position tolerance (*44*, *46*) (see Discussion).

This is illustrated in Fig. 2A, which shows a representative complex cell from the control group (left: blue lines) and a representative simple cell from the experimental group (right: orange lines). Both units displayed sharp orientation tuning (polar plots), but the STA method successfully recovered a sharp, Gabor-like RF only for the simple cell – as expected, given the nonlinear stimulus-response relationship of complex cells (*44*). Consistently, the response of the complex cell was only weakly modulated at the temporal frequency (4 Hz) of its preferred grating (middle plots), with the highest power spectral density concentrated at frequencies < 4 Hz (bottom plot). By contrast, the response of the simple cell was strongly phase modulated, with a power spectrum narrowly peaked at the grating frequency. Thus, by z-scoring the power spectral density of the response at the preferred grating frequency, it was possible to define a modulation index (MI) that distinguished between complex (MI < 3) and simple (MI > 3) cells (*44*, *47*) (see Materials and Methods).

We applied this criterion to the neuronal populations of 105 and 158 well-isolated single units recorded from, respectively, the control and experimental group, and we found a significantly lower fraction of complex cells in the latter (39%, 61/158) with respect to the former (55%, 58/105; $p < 0.01$, Fisher exact-test). Consistently, the median MI for the control population (2.69 ± 0.29) was significantly smaller than for the experimental one (3.52 ± 0.25; $p < 0.05$, Wilcoxon test). Such a difference became very sharp after restricting the comparison to the

neurons that, in both populations, were at least moderately orientation tuned (i.e., 50 control and 75 experimental units with OSI > 0.4). The resulting MI distribution for the control group had a typical double-peak shape (*44*), featuring two maxima, at MI~2 and MI~5, corresponding to the two classes of the complex and simple cells (Fig. 2B, blue curve). Instead, for the experimental group, the peak at low MI was flattened out, leaving a single, prominent peak at MI~5 (orange curve). This resulted in a large, significant difference between the two distributions and their medians (dashed lines), with the fraction of complex cells being almost half in the experimental (35%, orange bar) than in the control group (60%, blue bar).

The lower incidence of complex cells in the experimental group was confirmed when a different metric (the $F1/F0$ ratio; see Material and Methods) was applied to quantify the modulation of neuronal responses at the temporal frequency of the gratings (Fig. S3; see the Discussion for a thorough comparison among the MI and $F1/F0$ indexes, and an explanation of why our main analyses have been carried out using the MI). We also verified that the difference in the fraction of complex cells found between the two groups was not driven by a few outlier recording sessions. To this aim, we performed a bootstrap analysis in which: 1) we obtained 100 surrogate MI distributions for the populations of control and experimental units by sampling with replacement the available sessions for the two groups; and 2) we computed the fraction of complex cells found in each surrogate distribution. This allowed estimating the spread of the fraction of complex cells measured in each group, as a result of the variable sampling of the recorded sessions. The overlap between the spreads obtained for the two groups was minimal (Fig. S4A) and not significant (Fig. S4B, $p < 0.05$), thus showing that the lower incidence of complex cells in the experimental group was robust against the sampling of V1 units performed across different recordings/animals.

Conversely, no difference was observed between the two groups in terms of orientation tuning (Fig. 2C), with the OSI distributions (blue and orange curves) and their medians (dashed lines) being statistically undistinguishable, as well as the fraction of sharply orientation-tuned units (i.e., neurons with OSI > 0.6; blue vs. orange bar). A similar result was found for direction tuning (Fig. S5; see Discussion for an interpretation of this finding). Taken together, these results suggest that our experimental manipulation substantially impaired the development of complex cells, but not the emergence of orientation and motion sensitivity.

This conclusion was confirmed by comparing the quality of the RFs inferred through STA for the experimental and control units. To ease the comparison, the pixel intensity values in a STA image were z-scored, based on the null distributions of STA values obtained for each pixel after randomly permuting the association between frames of the movie and spike times 50 times. This allowed reporting the intensity values of the resulting z-scored, STA images in terms of their difference (in units of standard deviation σ) from what expected in the case of no frame-related information carried by the spikes. As illustrated by the examples shown in Fig. 3A, we found that STA was as successful at yielding sharp, linear RFs (often similar to Gabor filters) for the experimental units as for the control ones. The sharpness of the STA images, as assessed through an expressly devised contrast index (*44*) (CI; see Materials and Methods), was similar for the two groups, with the CI distributions and their medians being statistically undistinguishable (Fig. 3B, blue vs. orange curve/line). As expected, for both groups, the mean CI was significantly larger for the simple than for the complex cells (dark vs. light bars), reflecting the better success of STA at inferring the linear RFs of the former, but no difference was found between the mean CIs of the simple cells of the two groups (dark blue vs. brown bar) and the mean CIs of the complex cells (light blue vs. yellow bar).

To further explore the extent to which the spatial structure of the STA-based RFs was similar for the experimental and control units, we measured the size of the RFs and we counted how many distinct lobes they contained (this analysis was applied only to the units with well-defined linear RFs, i.e., to STA images within the top quartiles of the CI distributions shown in Fig. 3B, left). To count the lobes, we binarized each STA image, by applying a threshold to the modulus of its intensity values. This allowed identifying the lobes as distinct connected regions that crossed the binarization threshold [a more detailed description of this procedure is provided in Materials and Methods and a graphical illustration can be found in Figure 5B of (*44*)]. Since these regions became progressively smaller and fewer as a function of the magnitude of the binarization threshold, we compared the distributions of lobe counts obtained for the experimental and control units across different thresholds – from 3.5 to 6.5 σ. At every tested threshold, the distributions of lobe counts for the two populations were statistically indistinguishable (Fisher's exact test, $p > 0.05$; compare matching rows in Fig. 3C). The same was true for the distributions of RF sizes (compare matching rows in Fig. 3D), with the RF size of a unit being defined as the mean of the major and minor axes of the ellipse that best fitted the area covered by the detected lobes. These results confirmed that our experimental manipulation did not alter the spatial tuning properties of V1 units.

Postnatal rearing in a temporal discontinuous visual environment reduces the ability of complex cells to represent stimulus orientation in a translation-invariant manner

Next, we tested the extent to which the experimental units that had been classified as complex cells fully retained the functional properties of this class of neurons. As already shown in the previous section, the key property of complex cells is their ability to fire more persistently than

simple cells, in response to a continuous, spatiotemporally correlated visual input. This can be understood based on intuitive considerations, i.e., the local invariance of complex cells to (e.g.) translations of their preferred oriented edges. In the original work of Hubel and Wiesel (*17*), this property emerged when static oriented bars matching the preferred orientation of a complex cell were shown in different RF positions and, despite such translations, were found to elicit strong responses in the recorded unit. More recent investigations of V1 have relied instead on moving stimuli, such as the full-field drifting gratings used in our study, that allow probing at once the invariance properties of all the units recorded with a multi-electrode array. In these experiments, the translation invariance of complex cells manifests itself as the phase invariance of the response – despite the phasic alternation of light and dark oriented stripes, produced by the drifting of the preferred grating across its RF, a complex cell is able to respond to the stimulus with a more sustained, temporally persistent firing, as compared to a simple cell (compare the blue and orange rasters/PSTHs in Fig. 2A). More in general, such persistent, slowly changing responses should be expected every time a complex cell is probed with a spatiotemporally correlated stimulus, such as the noise movies used in our study to map the RFs through STA. From a theoretical point of view, this is consistent with the predictions of UTL models, such as SFA (*8*, *9*), that are based exactly on maximizing the slowness (or persistence) of neuronal responses in order to learn invariance. Critically, the different persistency of the responses of complex and simple cells to spatiotemporally correlated stimuli is not expected to result from intrinsic differences in terms of membrane excitability, temporal integration of the synaptic inputs or firing dynamics. In other words, complex cells are not expected to fire more persistently than simple cells when probed with brief, static stimuli (e.g., a complex will not continue to fire persistently in the absence of the stimulus). It is the invariance of the stimulus

representation afforded by complex cells that is at the origin of their slower responses. As such, the more persistent firing of complex cells can only emerge when V1 neurons are tested with spatiotemporally continuous stimuli.

To measure the persistence of neuronal responses in our recorded populations, we computed the time-constants of the exponential fits to the autocorrelograms of the spike trains evoked by the noise movies. This analysis was restricted to those units whose firing was strongly modulated at the frequency of variation of the contrast in the noise movies (i.e., 0.1 Hz; see examples in Fig. 4A, top and see Material and Methods for details). This ensured that our analysis measured the stimulus-dependent amount of slowness in the neuronal responses, as determined by the interplay between the temporal continuity of the visual stimulus and the transformation invariance afforded by the recorded neurons. As expected, the average time constant was larger for the control than for the experimental units (Fig. 4B). Such difference, however, was not merely driven by the larger fraction of complex cells in the control group (Fig. 2B). In fact, while the average time constants did not significantly differ between the simple cells of the two groups (Fig. 4C, dark blue vs. brown bar), the responses of complex cells unfolded over a shorter time scale for the experimental than for the control units (yellow vs. light blue bar).

To understand the functional implication of such abnormally fast-changing stimulus representations, we assessed the ability of the four distinct populations of simple and complex cells of the two groups to support stable decoding of stimulus orientation over time. To this aim, we randomly sampled 300 neurons from each population (after having first matched the populations in terms of OSI and orientation preference distributions; see Materials and Materials and Methods), so as to obtain four, equally sized and similarly tuned pseudo-populations, whose

units homogenously covered the orientation axis. We then trained binary logistic classifiers to discriminate between 0°- and 90°-oriented gratings (drifting at 4 Hz), based on the activity of each pseudo-population. Each classifier was trained using neuronal responses (spike counts) in a 33 ms-wide time bin that was randomly chosen within the presentation epoch of the gratings. We then tested the ability of each classifier to generalize the discrimination to test bins at increasingly larger time lags (TLs) from the training bin (see Fig. 5A and Materials and Methods for details). As expected, given the strong phase dependence of their responses (see cartoon in Fig. 5A, top), the simple cells from both groups yielded generalization curves that were strongly modulated over time and virtually identical (Fig. 5B, dark blue and brown curves). The performance was high (≥ 80% correct) at test bins where the phase of the grating was close to that of the training bin (i.e., at TLs that were multiple of the 250 ms grating period), but it dropped to less than 30% correct (i.e., well below chance; dashed line) at test bins where the grating was in opposition of phase with respect to the training bin (e.g., at TL ~ 125 ms). By comparison, the complex cells of the control group, by virtue of their weaker phase dependence (see cartoon in Fig. 5A, bottom), afforded a decoding of grating orientation that was substantially more phase-tolerant, with the performance curve never dropping below chance level at any TL (Fig. 5B, light blue curve). However, for the complex cells of the experimental group, the performance curve (in yellow) was not as stable – at most TLs, it was 5-10 percentage points smaller than the performance yielded by the control complex cells, dropping significantly below chance at test bins where the grating was in opposition of phase with respect to the training bin. That is, the ability of the experimental complex cells to support phase-tolerant orientation decoding was somewhat in between that of properly developed complex cells and that of simple cells. This shows that, even if some complex cells survived our experimental manipulation (i.e.,

the rearing in temporally broken visual environments), their functional properties were nevertheless impaired by the controlled rearing, as demonstrated by their reduced ability to support phase-invariant decoding of stimulus orientation.

DISCUSSION

The findings reported in our study show that breaking the temporal continuity of early visual experience severely interferes with the typical development of complex cells in V1, leading to a sizable reduction of their number (Fig. 2B) and an impairment of their functional properties (Fig. 4C and 5B). This implies that experience with the temporal contiguity of natural image sequences over time scales longer than 66.7 ms (i.e., the frame duration used during our controlled rearing) plays a critical role in postnatal development of the earliest form of invariance found along the ventral stream. Such an instructive role of temporal continuity of visual stimuli, so far, has been empirically demonstrated only in adult monkeys, at the very last stage of this pathway, the inferotemporal cortex (*48*, *49*), likely supported by the same associative plasticity mechanisms underlying visual long-term memory (*50*). At the same time, our experiments show that degrading the amount of the temporal continuity experienced during development does not affect the emergence of orientation tuning (Fig. 2D), with simple cells exhibiting unaltered spatial (Fig. 3), temporal (Fig. 4C) and functional (Fig. 5B) properties. Interpreting these findings requires a careful discussion of our procedure to classify simple and complex cells, as well as of the strengths and limits of our protocol for controlled rearing, along with a thorough review of the previous studies in which early visual experience was altered during postnatal development.

Distinguishing simple from complex cells

The original definition of simple cells provided by Hubel and Wiesel was based on the subjective assessment of distinct, elongated ON and OFF flanking regions in the RF of this class of neurons, which endowed them with the property of being both orientation selective and very sensitive to the position of their preferred oriented edges (*17*). By contrast, no clearly defined ON and OFF regions could be found for complex cells, which retained the ability to selectively respond to specific orientations, but in a locally position-invariant way – a complex cell would still respond vigorously, despite displacements of the preferred oriented edge within its RF. Later studies proposed more objective measures to distinguish simple from complex cells (*51*, *52*), by relying instead on the level of modulation of the neuronal response during the presentation of a drifting grating. This approach has gained increased popularity with the advent of multielectrode arrays. In fact, recording many tens of neurons in parallel does not allow probing each individual unit with cell-specific stimuli (such as the oriented bars originally used by Hubel and Wiesel) – full-field stimuli (such as drifting gratins) are necessary to simultaneously test the recorded population (*44*, *46*, *53*). However, assessing the level of modulation of neuronal firing to distinguish simple from complex cells raises two important issues. The first is methodological and concerns the definition of the most suitable metric to measure response modulation (*47*, *54*). The second, deeper issue concerns the validity itself of the classification of V1 neurons into distinct functional cell types, with some authors proposing that a continuum of cell properties, rather than a segregation into discrete cell classes, better describes the organization of visual cortex (*55*, *56*).

With regard to the first issue, the traditional metric that has been proposed, and is still often used, to characterize response modulation is the so-called $F1/F0$ ratio – i.e., the ratio

between the amplitude of the Fourier spectrum at the temporal frequency of the drifting grating and the mean spike rate of the neuron (see Materials and Methods for details). This metric however has been criticized in a recent study (*47*), which quantitatively demonstrated the already-known drawbacks of the $F1/F0$ ratio in terms of consistency and reliability. This ratio, in fact, is very sensitive to the relative magnitude of the evoked and background firing rate of a neuron. Specifically, it tends to yield low values not only in the absence of modulation, but also when the amplitude of the modulation is weak, relative to the background rate. In this scenario, the $F1/F0$ ratio tends to underestimate the level of modulation, thus misclassifying as complex cells units that exhibit clearly modulated activity in their PSTHs. In addition, the $F1/F0$ ratio is not a standardized metric and the threshold traditionally used to distinguish complex from simple cells (i.e., $F1/F0 = 1$) is arbitrary and not based on statistical considerations. This led the authors of (*47*) to define a new modulation index (which they named standardized $F1$, or $zF1$), in which the spectral intensity at the temporal frequency of the drifting grating (i.e., $F1$) is referred to the mean spectral intensity and divided by its standard deviation. As shown in (*47*), this metric is more reliable in capturing the level of modulation of neuronal firing that is apparent from the PSTHs. Also, being a standardized metric, a criterion to distinguish highly modulated (i.e., simple) from poorly modulated (i.e., complex) cells can be defined on statistical grounds – i.e., by measuring how distant $F1$ is from the mean spectral intensity in units of standard deviation.

In our study, we also used a standardized $F1$ metric to quantify the level of modulation of neuronal responses to drifting gratings (simply referred to as the MI index; see Material and Methods). This choice was motivated by the considerations explained in the previous paragraph and by having verified, in an earlier study, the effectiveness and robustness of this index at quantifying the level of response modulation not only in rat V1 and higher-level visual cortical

regions, but also across the layers of deep, artificial neural networks for image classification, such as HMAX and VGG16 (*44*). Notably, following our adoption of this metric, the key advantages of the standardized $F1$ index were recently acknowledged by the Allen Institute, which used it for its large scale surveys of mouse visual cortex (*57*, *58*).

In our current study, for completeness, we have also assessed the modulation of neuronal firing using two different instances of the $F1/F0$ ratio – the most commonly applied definition (*54*) and a modified version that has the advantage of being bounded between 0 and 2 (*59*) (see Materials and Methods). As expected, both $F1/F0$ ratios tended to inflate the proportion of units falling below the $F1/F0 = 1$ threshold that is typically used to classify a cell as complex (Fig. S3). Despite such reduced sensitivity to capture variations in the level of modulation of the firing rate, the experimental units still displayed a significantly larger response modulation than the control units (orange vs. blue curves; $p < 0.05$, Wilcoxon test). As a result, a significantly lower proportion of experimental cells was classified as complex (orange vs. blue bars; $p < 0.05$, Fisher exact-test), thus confirming the impact of rearing newborn rats in visually discontinuous environments on the development of complex cells.

As mentioned above, the debate about the best choice of the modulation index relates to the deeper issue of whether it is appropriate in the first place to segregate visual cortical neurons into discrete functional classes. Critically, the decoding analysis presented in our study (see Fig. 5) addresses both questions. In fact, from a computational perspective, the key functional property distinguishing simple from complex cells is the larger translation invariance that the latter are supposed to afford in the representation of stimulus orientation (*60*). Modulation indexes or other metrics measure this ability only indirectly and with a variable degree of reliability. On the other hand, reading-out stimulus orientation using a linear classifier directly

quantifies the amount of translation-invariant information that can be easily (i.e., linearly) extracted from the underlying neuronal representation (*16*). As such, our decoding analysis (Fig. 5) validates at once the existence of two functionally distinct subpopulations of visual cortical neurons and the metric (i.e., the MI index) we used to distinguish them. In fact, the radically different degree of phase invariance in the representation of stimulus orientation afforded by the two populations of units classified as simple and complex in the control group (dark vs. light blue curves) demonstrates that: 1) such populations are indeed functionally distinct, with respect to their ability to code invariantly stimulus orientation; 2) the MI provides a measure of response modulation that is highly consistent with the degree of translation invariance of the recorded units; and 3) the 3 σ threshold used to distinguish simple from complex cells effectively partitions the range of measured MI values into distinct functional classes.

Breaking temporal continuity of early visual experience: a comparison with strobe rearing studies

The development of complex cells in the animals reared with the temporally discontinuous movies (i.e., the experimental group) was strongly impaired, with the experimental animals showing a median modulation index that was almost twice as large as that of the control rats, and a fraction of complex cells that was almost half (Fig. 2B). However, it was not fully abolished – a small amount of complex cells survived the experimental manipulation, although with a diminished capability of supporting translation-invariant decoding of stimulus orientation (Fig. 5B). At first glance, this may seem at odd with the hypothesis that temporal continuity is strictly necessary for the development of transformation tolerance in V1. However, it should be considered that, as explained in the Results, the disruption of temporal continuity achieved with

our controlled rearing was not complete. Even if the frame-scrambled rearing videos lacked temporal structure at time scales longer than 66.7 ms (Fig. 1B-C), the experimental rats could still experience some residual amount of temporal continuity in the visual experience because of head and/or eye movements (*61–63*). Specifically, the visual features that the animals may have experienced as continuously transforming (e.g., translating) include: 1) structural parts of the physical environment (e.g., the edges of the monitors; see Fig. S1); 2) parts of their own bodies; and 3) the content of individual movie frames, although over very short temporal spans (≤ 66.7 ms). As already explained, such residual temporal continuity was not accidental but intentional. It was dictated by the need of allowing the rats full access to the spatial content of the individual image frames, which prevented using frame rates higher than rat flicker fusion frequency (~30-40 Hz) (*41*). In addition, although experience with the motion of physical features and/or body parts may have been strongly limited by the use of head fixation, we preferred to avoid such procedure. In fact, head fixation would have prevented a natural and active exploration of the visual environment, which, in rodents, has been shown to strongly affect the plasticity and development of visual cortex (*42*, *43*) – a phenomenon that is consistent with the tight relationship between the encoding of visual and locomotory/positional signals recently reported in rodent V1 (*63–67*). The fear that head fixation could limit the impact of controlled visual rearing on the developing visual cortex was reinforced by the failure of a previous study (performed on head-fixed ferrets) to causally demonstrate that experience with oriented visual patterns is necessary for the development of orientation tuning in V1 (*30*). On the basis of these considerations, we reasoned that the rearing would have been more effective if the newborn rats were left unrestrained inside the immersive visually environments, even at the cost of allowing some residual temporal continuity in their visual experience. The fact that, despite such residual

continuity, the development of complex cells was strongly impaired in the experimental rats testifies to the paramount importance of experiencing a fully continuous visual environment for the development of translation tolerance. At the same time, the residual temporal continuity experienced during rearing can easily explain why the development of complex cells was not fully abolished.

The incomplete disruption or temporal continuity during postnatal rearing can also explain why the development of direction selectivity was unaffected by our experimental manipulation (Fig. S5). This finding was somewhat unexpected, given that, in agreement with the temporal extension of the sparse coding principle (*68*), postnatal rearing under stroboscopic illumination has been found to produce a substantial loss of direction selectivity in V1 (*33–39*). Such discrepancy with our result can be understood by considering that strobe light flashes in these earlier studies had a much shorter duration (typically ~10 μs) than the frame duration in our movies. Thus, in strobe rearing studies, the animals were fully deprived of experience with smooth motion signals, while our controlled rearing allowed the content of individual image frames to be experienced as smoothly moving (e.g., translating) over time scales ≤ 66.7 ms. On the other hand, our rearing ensured that the temporal correlation of the visual stream delivered through the displays was close to zero over time scales > 66.7 ms (see Fig. 1B-C). By contrast, strobe rearing, especially at high rates (8 Hz), allowed for such a high-frequency sampling of the visual environment to resemble a "normal patterned input" (*38*), leading to "human subjective experience […] of a series of jerky images, reminiscent of the early motion picture" (*33*). This implies that, despite the disruption of smooth motion signals at the microsecond time scale, the animals subjected to strobe rearing likely experienced a strongly correlated visual input at timescales as large as several hundred of milliseconds or a few seconds (i.e., of the order of what

experienced by our control rats; see blue curves in Fig. 1B-C). This likely explains why several studies based on strobe rearing at 4-8 Hz mention the existence of complex cells in the strobe-reared animals, without explicitly reporting any loss of such neurons (*33*, *36*, *39*), with one study in particular reporting no qualitative differences in the sampling of simple and complex cells between the strobe-reared and control subjects (*37*).

In summary, when our results are considered together with those of earlier strobe rearing studies, an intriguing double dissociation emerges with regard to the instructive role of temporal continuity during cortical development. The temporal learning mechanisms leading to the development of invariance appear to be distinct and independent from those supporting the development of direction tuning, with the former operating over time scales that are several orders of magnitude longer than the latter. As a result, successful disruption of temporal continuity at the microsecond time scale, but preservation of temporal correlations at time scales of the order of tens/hundreds of milliseconds (as in most strobe rearing studies), interferes with the development of direction tuning but spares the development of complex cells. Vice versa, preserving time contiguity at the microsecond/millisecond level, but destroying correlations at longer time scales (as in our study), impairs the development of complex cells, without preventing the emergence of direction selectivity.

Another finding of our study that is worth discussing in the context of the limitations of our rearing procedure and previous strobe rearing studies is the typical development of orientation tuning (Fig. 2C) and spatial RF properties (Fig. 3) observed in the experimental rats. Given that the access to the image content of the individual movie frames was the same as for the control animals, this result strongly suggests that development of shape tuning depends on the exposure to the spatial statistics of natural images, rather than on the temporal continuity of the

visual stream. Thus, our results would add to the indirect evidence in favor of the role played by USL during development (*29*, *6*). However, given the residual amount of temporal continuity allowed by our rearing procedure, we cannot exclude that, as for the case of direction tuning, development of orientation tuning too may rely on UTL mechanisms working at smaller temporal scales than those required to support the development of invariance. The fact that strobe rearing at 4-8 Hz impairs the development of direction tuning but not of orientation selectivity makes this scenario unlikely (*33*, *36*, *37*, *39*). Nevertheless, this does not fully exclude the possibility that an intermediate time scale of temporal continuity exists that is necessary for the development of spatial selectivity, but is neither sufficiently long to support the development of invariance, neither sufficiently short to sustain the development of direction tuning. To settle this question, future studies will need to rear newborn animals with purely static images, possibly varying image duration from a few tens of milliseconds to a few tens of microseconds in different experimental groups. This will require combing head fixation with eye tracking in closed-loop experiments, where initiation of a saccade should abort stimulus presentation, so as to fully deprive the subjects of the experience of continuous transformations of the visual input at any time scale.

Nature vs. nurture

While our findings, as those of previous strobe rearing studies, points to a pivotal, instructive role of early visual experience in determining the tuning properties of visual cortical neurons, the residual amount of complex cells in our experimental animals, as well as the unimpaired tuning for orientation and direction, could also be explained as the result of genetically-encoded, experience-independent developmental programs (*69*). Support for such “hardwiring” hypothesis

comes from studies in which orientation and direction selectivity in various species was found to be already highly developed at the onset of visual experience, i.e., right after eye opening (EO) (*20*, *30*, *70–74*). However, this does not seem to apply to rat V1, whose functional properties have been reported to remain immature after postnatal rearing in complete darkness (*40*). This may point to differences among species, but also among experimental manipulations, since in many studies the animals were kept in a normal dark-light cycle before EO (*70–72*). Differently from dark rearing (DR), such procedure allows for a very blurred and dimmed stimulation of the retina through the closed eyelids, which could drive the development of cortical tuning in an experience-dependent way, either by directly evoking neuronal responses or by fostering the generation of waves of spontaneous activity (*75*) (see next paragraph). In addition, even a few hours of visual experience after eye opening may be enough to drive fast development of cortical tuning properties, as demonstrated in juvenile ferrets (*27*, *28*). To date, the most convincing demonstration of experience- and activity-independent formation of orientation and direction tuning comes from a mouse study in which DR was paired with genetic silencing of spontaneous cortical activity during development (*74*) (unfortunately, the study did not test whether complex cells developed normally).

The possible role played by spontaneously generated activity in instructing the development of cortical tuning is yet another explanation for the residual fraction of complex cells and the unaltered orientation and direction selectivity found in our study. Key to this concept, often referred to as “innate learning” (*76*), is the idea that, during development, neural circuits, by virtue of their genetically-determined structure, could self-generate activity patterns that are able to act as “training examples” to sculpt and refine their own wiring or the wiring of other downstream circuits. Such activity-dependent structuring may be driven by the same

unsupervised plasticity rules (such as USL and UTL) that would later act on stimulus-evoked activity, after the onset of sensory experience. An example of innate learning is the role played by the spatiotemporally correlated patterns of cortical activity evoked by retinal waves (*77*) in driving the development of topographic visual maps (*78–80*). From a theoretical standpoint, computational studies have shown that such spontaneous activity patterns could also support the development of simple and complex cells via, respectively, sparse coding (*76*, *81*) and temporal learning mechanisms (*82*). This may explain the finding of a recent study, where the presence of complex cells in mouse V1 was reported at EO already (*53*). However, the animals included in that study were not subjected to DR and were also allowed normal visual experience for several hours before the neuronal recordings. This makes it difficult to infer what developmental mechanism was at the origin of the complex cells reported by (*53*) – whether experience-dependent or independent and, in the latter case, whether activity-driven (innate learning) or purely genetically encoded.

Conclusions and implications

Summing up, it is difficult to fully reconcile the conclusions of the studies reviewed in the previous two sections, especially given the variability found across species and the variety of experimental approaches that have been devised to manipulate visual experience and/or retinal/cortical activity during early postnatal development. This makes it hard to know whether our altered rearing acted on visual cortical circuits in a “blank”, immature state, or, rather, reshaped the wiring of circuits that had already been structured by innate developmental programs, possibly combined with the effect of internally generated activity. Nevertheless, what our data causally demonstrate is that a form of plasticity based on unsupervised temporal

learning must be at work in the developing visual cortex to build up (or maintain) invariance, in a way that is highly susceptible to the degree of temporal correlation of visual experience.

From a theoretical standpoint, this result causally validates the family of UTL models (*7–15*) at the neural level, albeit strongly suggesting that their scope is limited to the development of invariance and not of shape selectivity. More in general, since slowness has been related to predictability (*83–85*), our results are also consistent with normative approaches to sensory processing that are based on temporal prediction (*86*). On the other hand, our findings, by showing that exposure to the spatial structure of natural images alone is not enough to enable proper development of complex cells, reject computational accounts of invariance based exclusively on USL (*3*, *4*), while leaving open the possibility that the latter may govern the development of shape tuning (*1*, *2*, *5*, *6*). As a result, our study tightly constrains unsupervised models of visual cortical development, supporting theoretical frameworks where the objectives of sparseness and slowness maximization coexist to yield, respectively, shape selectivity and transformation tolerance (*13*, *14*, *87*).

MATERIALS AND METHODS

All animal procedures were in agreement with international and institutional standards for the care and use of animals in research and were approved by SISSA Institutional Animal Care and Use Committee and by the Italian Ministry of Health (Project DGSAF 22791-A, submitted on September 7, 2015 and approved on December 10, 2015, approval 1254/2015-PR).

Animal subjects and controlled rearing protocol

Data were obtained from 18 Long–Evans male rats that were born and reared in our facility for visually controlled rearing. The facility consists of a small vestibule, where the investigators can wear the infrared goggles that are necessary to operate in total darkness, and a larger, lightproof room, containing a lightproof housing cabinet (Tecniplast) and four custom cabinets (Tecniplast) for exposure of the rats to controlled visual environments.

Pregnant mothers (Charles River Laboratories) where brought into the housing cabinet about one week before delivery. Pups were born inside the cabinet and spent the first two weeks of their life in total darkness with their mothers. Starting from P14 (i.e., at eye opening) through P60 (i.e., well beyond the end of the critical period), each rat, while still housed in full darkness (i.e., inside the housing cabinet) with his siblings, was also subjected to daily, 4-hours-long exposures inside an immersive visual environment (referred to as the *virtual cage*), consisting of a transparent basin (480x365x210 mm, Tecniplast 1500U), fully surrounded by 4 computer-controlled LCD monitors (one per wall; 20" HP P202va; see Fig. S1), placed on the shelf of one of the custom cabinets (each cabinet had 4 shelves, for a total of 16 rats that could be simultaneously placed in the visually controlled environments). These controlled-rearing environments, which are reminiscent of those used to study the development of object vision in chicks (*31, 32, 88*), were custom designed in collaboration with Videosystem, which took care of building and installing them inside the custom cabinets.

Different visual stimuli were played on the monitors, depending on whether an animal was assigned to the experimental or the control group. Rats in the control group (n = 8) were exposed to natural movies, including both indoor and outdoor scenes, camera self-motion and moving objects. Overall, the rearing playlist included 16 videos of different duration, lasting

from a few minutes to half an hour. The playlist was played in random order and looped for the whole duration of the exposure. Rats from the experimental group (n = 10) were exposed to a time-shuffled version of the same movies, where the order of the frames within each video was randomly permuted, so as to destroy the temporal continuity of the movie (see Fig. 1B-C), while leaving unaltered the natural spatial statistics of the individual image frames. All movies were played at 15 Hz, which is approximately half of the critical flicker fusion frequency (~30-40 Hz) that has been measured for the rat (*41*), to make sure that the animals could experience the image content of the individual frames of the movies.

Animal care, handling and transfer operations were always executed in absolute darkness, using night vision goggles (Armasight NXY7), in such a way to prevent any unwanted exposure of the animals to visual inputs different from those chosen for the rearing.

Quantification of the temporal correlations in the rearing videos

To assess the level of temporal structure in the videos that were administered to the control and experimental rats during the controlled rearing inside the virtual cages, we computed the average pixel-level temporal autocorrelation function for each movie. This function was then fitted with an exponential decay model, whose time constant provided a measure of the time scale of temporal continuity in the movie.

The first step to compute the temporal autocorrelation function was to chunk each frame in a movie into blocks of 6x6 pixels and then average the pixel intensity values inside each block, so as to lower the resolution of the movie frames. Such downsampling was necessary to ease the computational load of the analysis. Each movie frame was then unrolled into a vector and the correlation matrix of the ordered ensemble of frame vectors was computed. Finally, all

the elements of the correlation matrix that were located along the kth diagonal (where k denotes the distance from the main diagonal) were averaged to obtain the value of the mean temporal autocorrelation function at lag k (with k ranging from 1 to the maximal separation between two frames in a movie).

The following exponential model was used to fit the mean temporal autocorrelation function obtained for each movie:

$$f(\Delta t) = Ae^{-\frac{\Delta t}{\tau}} + C,$$

where Δt is the time lag (obtained by multiplying the frame lag k by the frame duration of 66.7 ms), and τ is the time constant of the exponential decay, whose value was taken as a measure of the amount of temporal structure in a movie. A and C are free parameters. Only the first 4.95 s of the mean temporal autocorrelation functions were taken into account for the fitting procedure (see Fig. 1B-C).

Surgery and recordings

Acute extracellular recordings were performed between P60 and P90 (last recording). During this 30-days period, the animals waiting to undergo the recording procedure were maintained on a reduced visual exposure regime (i.e., 2-hours-long visual exposure sessions every second day; see previous section).

The surgery and recording procedure was the same as described in (*44*). Briefly, the day of the experiment, the rat was taken from the rearing facility and immediately (within 5-10 minutes) anesthetized with an intraperitoneal injection of a solution of 0.3 mg/kg of fentanyl (Fentanest, Pfizer) and 0.3 mg/kg of medetomidin (Domitor, Orion Pharma). A constant level of

anesthesia was then maintained through continuous intraperitoneal infusion of the same aesthetic solution used for induction, but at a lower concentration (0.1 mg/kg/h fentanyl and 0.1 g/kg/h medetomidine), by means of a syringe pump (NE-1000; New Era Pump Systems). After induction, the rat was secured to a stereotaxic apparatus (Narishige, SR-5R) in flat-skull orientation (i.e., with the surface of the skull parallel to the base of the stereotax) and, following a scalp incision, a craniotomy was performed over the target area in the left hemisphere (typically, a 2x2 mm^2 window), and the dura was removed to allow the insertion of the electrode array. The coordinates of penetration used to target V1 were ~6.5 mm posterior from bregma and ~4.5 mm left to the sagittal suture (i.e., anteroposterior 6.5, mediolateral 4.5). Once the surgical procedure was completed, and before probe insertion, the stereotax was placed on a rotating platform and the rat's left eye was covered with black, opaque tape, while the right eye (placed at 30 cm distance from the monitor) was immobilized using a metal eye-ring anchored to the stereotax. The platform was then rotated in such a way to bring the binocular visual field of the right eye to cover the left side of the display.

Extracellular recordings were performed using either single- (or double-) shank 32- (or 64-) channel silicon probes (NeuroNexus Technologies) with site recording area of 775 μm^2 and 25 µm of intersite spacing. After grounding (by wiring the probe to the animal’s head skin), the electrode was manually lowered into the cortical tissue using an oil hydraulic micromanipulator (Narishige, MO-10; typical insertion speed: 5 µm/s), up to the chosen insertion depth (800 –1000 m from the cortical surface), either perpendicularly or with a variable tilt, between 10° and 30°, relative to the vertical to the surface of the skull. Extracellular signals were acquired using a system 3 workstation (Tucker Davis Technologies) with a sampling rate of 25 kHz.

Since, in rodents, the largest fraction of complex cells in found in layer 5 of V1 (*46*), our recordings aimed at sampling more densely that layer. This was verified *a posteriori* (Fig. S2), by estimating the cortical depth and laminar location of the recorded units, based on the patterns of *visually evoked potentials* (VEPs) recorded across the silicon probes used in our recording sessions. More specifically, we used a template-matching algorithm for laminar identification of cortical recording sites that we recently developed and validated in an appositely dedicated methodological study (*89*). Briefly, the method finds the optimal match between the pattern of VEPs recorded, in a given experiment, across a silicon probe and a template VEP profile, spanning the whole cortical thickness, that had been computed by merging an independent pool of 18 recording sessions, in which the ground-true depth and laminar location of the recording sites had been recovered through histology. The method achieves a cross-validated accuracy of 79 µm in recovering the cortical depth of the recording sites and a 72% accuracy in returning their laminar position, with the latter increasing to 83% for a coarser grouping of the layers into supagranular (L1-3), granular (L4) and infragranular (L5-6).

Visual stimuli

During a recording session, each animal was presented with: i) 20 repetitions (trials) of 1.5-s-long drifting gratings, made of all possible combinations of two spatial frequencies (0.02 and 0.04 cpd), two temporal frequencies (2 and 4 Hz), and 12 directions (from 0° to 330°, in 30° increments); and ii) 20 different 60-s-long spatially and temporally correlated, contrast modulated, noise movies (*44*, *46*). All stimuli were randomly interleaved, with a 1-s-long inter-stimulus interval (ISI), during which the display was set to a uniform, middle-gray luminance level. To generate the movies, random white noise movies were spatially correlated by

convolving them with a Gaussian kernel having full width at half maximum (FWHM) corresponding to a spatial frequency of 0.04 cpd SF. Temporal correlation was achieved by convolving the movies with a causal exponential kernel with a 33 ms decay time-constant. To prevent adaptation, each movie was also contrast modulated using a rectified sine wave with a 10 s period from full contrast to full contrast (*46*).

Stimuli were generated and controlled in MATLAB (MathWorks) using the Psychophysics Toolbox package and displayed with gamma correction on a 47-inch LCD monitor (SHARP PNE471R) with 1920x1080-pixel resolution, 220 cd/m^2 maximum brightness, and spanning a visual angle of 110° azimuth and 60° elevation. Grating stimuli were presented at 60 Hz refresh rate, whereas noise movies were played at 30 Hz.

Single unit isolation

Single units were isolated offline using the spike sorting package KlustaKwik-Phy (*90*). Automated spike detection, feature extraction, and expectation maximization clustering were followed by manual refinement of the sorting using a customized version of the Phy interface. Specifically, we took into consideration many features of the candidate clusters: a) the distance between their centroids and their compactness in the space of the principal components of the waveforms (a key measure of goodness of spike isolation); b) the shape of the auto- and cross correlograms (important to decide whether to merge two clusters or not); c) the variation, over time, of the principal component coefficients of the waveform (important to detect and take into account possible electrode drifts); and d) the shape of the average waveform (to exclude, as artifacts, clearly non physiological signals). Clusters suspected to contain a mixture of one or more single units were separated using the “reclustering” feature of the GUI. After the manual

refinement step, we included in our analyses only units that were: i) well-isolated, i.e. with less than 0.5% of "rogue" spikes within 2 ms in their autocorrelogram; and ii) grating-responsive, i.e., with the response to the most effective grating condition being larger than 2 spikes per second (baseline-subtracted), and being larger than 6 z-scored points relative to baseline activity. The average baseline (spontaneous) firing-rate of each well-isolated unit was computed by averaging its spiking activity over every ISI condition. These criteria led to the selection of 105 units for the control group and 158 units for experimental group.

Quantification of selectivity

The response of a neuron to a given drifting grating was computed by counting the number of spikes during the whole duration of the stimulus, averaging across trials and then subtracting the spontaneous firing rate (see previous section). To quantify the tuning of a neuron for the orientation and direction of drifting gratings, we computed two standard metrics, the orientation and direction selectivity indexes (OSI and DSI), which are defined as: $\mathrm{OSI} = (R_{\mathrm{pref}} - R_{\mathrm{ortho}})/(R_{\mathrm{pref}})$, and $\mathrm{DSI} = (R_{\mathrm{pref}} - R_{\mathrm{opposite}})/(R_{\mathrm{pref}})$, where R_{pref} is the response of the neuron to the preferred direction, R_{ortho} is the response to the orthogonal direction, relative to the preferred one (i.e., $R_{\mathrm{ortho}} = R_{\mathrm{pref}} + \pi/2$), and R_{opposite} is the response to the opposite direction, relative to the preferred one (i.e., $R_{\mathrm{opposite}} = R_{\mathrm{pref}} + \pi$). Values close to one indicate very sharp tuning, whereas values close to zero are typical of untuned units.

Quantification of phase modulation (i.e. position tolerance)

Since phase shifts of a grating are equivalent to positional shifts of the whole, 2D sinusoidal pattern, a classical way to assess position tolerance of V1 neurons (thus discriminating between

simple and complex cells) is to probe the phase sensitivity of their responses to optimally oriented gratings. Quantitatively, the phase-dependent modulation of the spiking response at the temporal frequency f_1 of a drifting grating was quantified by a modulation index (MI) adapted from (*47*) and used in (*44*), defined as:

$$\mathrm{MI} = \left| \frac{PS(f_1) - \langle PS \rangle_f}{\sqrt{\langle PS^2 \rangle_f - \langle PS \rangle_f^2}} \right|,$$

where PS indicates the power spectral density of the stimulus-evoked response, i.e., of the peri-stimulus time histogram (PSTH), and $\langle\ \rangle_f$ denotes the average over frequencies. This metric measures the difference between the power of the response at the stimulus frequency and the average value of the power spectrum in units of its standard deviation. The power spectrum was computed by applying the Blackman-Tukey estimation method to the baseline-subtracted, 10ms-binned PSTH. Since MI is a standardized measure, values greater than 3 can be interpreted as signaling a strong modulation of the firing rate at the stimulus frequency (typical of simple cells), whereas values smaller than 3 indicate poor modulation (typical of complex cells). On this ground, we adopted MI = 3 as a threshold for classifying neurons as simple or complex. The choice of this classification criterion, as well as the use of the MI index itself, were determined, prior to seeing the data collected for the current study, exclusively on the basis of our experience with the same metric and criterion in a previous study (*44*).

We also quantified the phase sensitivity of the recoded neurons using two other popular metrics of response modulation: the standard $F1/F0$ ratio, as well as a modified version of this metric that has the advantage of being bounded between 0 and 2 (we will refer to this metric as $F1/F0^*$). The $F1/F0$ ratio (*51*, *52*, *54*) is typically defined as:

$$F1/F0 = \frac{F1}{F0}$$

$F1$ is the value of the amplitude of the Fourier spectrum at the stimulus frequency f_1, whereas F0 is its value at the zero frequency f_0 (i.e., the "DC" or constant component of the response):

$$F1 = AS(f_1) \qquad F0 = AS(f_0 = 0)$$

On the other hand, the $F1/F0^*$ ratio (*59*) has been defined as:

$$F1/F0^* = \frac{2 \cdot F1}{(F0 + F1)}$$

This allows obtaining an index that is bounded to have a maximum value of 2 rather than infinity (as in the case of the $F1/F0$ ratio). The amplitude spectra used to compute the $F1/F0$ and $F1/F0^*$ ratios were obtained by subjecting each trial of the preferred grating orientation of a neuron to Fourier analysis. Trials with firing rate < 2Hz were excluded from the analysis. Specifically, Fourier amplitude spectra were obtained by applying the fast Fourier transform algorithm to the baseline-subtracted, 10 ms-binned PSTH of the steady-state grating response (i.e., from 250 ms to 1500 ms after stimulus onset). As done in previous studies (*52*, *54*, *59*), the threshold we adopted to classify neurons as simple or complex via these ratios was 1 for both indexes.

Estimation of linear receptive fields through STA and characterization of their properties

We used the spike-triggered average (STA) method (*91*, *92*) to estimate the linear RF structure of each recorded neuron. The method was applied to the spike trains fired by neurons in response to the spatiotemporally correlated and contrast modulated noise movies described above. To account for the correlation structure of our stimulus ensemble and prevent artifactual blurring of the reconstructed filters, we "decorrelated" the raw STA images by dividing them by the

covariance matrix of the whole stimulus ensemble (*91*, *92*). We used Tikhonov regularization to handle covariance matrix inversion. Statistical significance of the STA images was then assessed pixel-wise, by applying the following permutation test. After randomly reshuffling the spike times, the STA analysis was repeated multiple times (n = 50) to derive a null distribution of intensity values for the case of no linear stimulus-spike relationship. This allowed z-scoring the actual STA intensity values using the mean and standard deviation of this null distribution. The temporal span of the spatiotemporal linear kernel we reconstructed via STA extended till 330 ms before spike generation (corresponding to 10 frames of noise at 30 Hz frame rate). The STA analysis was performed on downsampled noise frames (16x32 pixels) and the resulting filters were later spline-interpolated at higher resolution for better visualization.

To estimate the amount of signal contained in a given STA image, we used the contrast index (CI) metric that we have introduced in a previous study (*44*) (see the method section and Figure 5A of that study). The CI is a robust measure of maximal local contrast in a z-scored STA image. Since the intensity values of the original STA images were expressed as z-scores (see above), a given CI value can be interpreted in terms of peak-to-peak (i.e., white-to-black) distance in sigma units of the z-scored STA values. For the analysis shown in Fig. 3B, the STA image with the highest CI value was selected for each neuron.

We also characterized the structural complexity of the RFs yielded by STA by counting the number of excitatory/inhibitory lobes that were present in a STA image and measuring the overall size of the resulting RF. The procedure is the same described in our previous study (*44*) (see the method section and Figure 5B of that study). Briefly, we applied a binarization threshold over the modulus of the z-score values of the image (ranging from 3 to 6 units of standard deviations). We then computed the centroid positions of the simply connected regions within the

resulting binarized image (i.e., the candidate lobes), as well as their center of mass (i.e., the candidate RF center). Finally, we applied a refinement procedure, which is detailed in (*44*), to prune spurious candidate lobes (often very small) that were far away from the RF center. Obviously, the number of lobes and the size of the RF (computed as the mean of the major and minor axes of the ellipse that best fitted the region covered by the detected lobes) depended on the binarization threshold. For this reason, in Fig. 3C-D, we have compared the lobe number and the RF size of the recorded populations of experimental (orange) and control (blue) units over a range of possible choices of this threshold.

Quantification of response slowness

For each neuron, we quantified the slowness of its response to the same noise movies used to estimate its RF, by computing the time constant of the autocorrelogram of the evoked spike trains (i.e., the probability density function of inter-spike intervals - ISI). Being the noise movies composed of richer visual patterns than drifting gratings (i.e., richer orientation and spatial frequency content), this was a way to assess the response properties of the recorded population in a slightly more naturalistic stimulation regime. The time constants τ were computed by fitting autocorrelograms with the following exponential function:

$$f(\Delta t) = Ae^{-\frac{\Delta t}{\tau}} + C$$

where Δt is the inter-spike time interval (see Fig. 3A, bottom) and τ is the time constant of the exponential decay, whose value was taken as a measure of the slowness of the response of each neuron to the noise movies. A and C are free parameters. Only the first 200 ms of the ISI distributions were taken into account for the fitting procedure (see Fig. 4A, bottom panels).

Only neurons that were strongly modulated at the frequency of variation of the contrast in the movies (i.e., 0.1 Hz) were included in the analysis. To select the neurons that met such criterion, the level of response modulation was quantified by a standardized contrast modulation index (MIc). The MIc was defined exactly as the MI that was used to assess the phase sensitivity of the responses to the gratings (see above), with the only difference that the target frequency to measure F1 (i.e., the power spectral density at the frequency of the modulated input) was now the frequency of the contrast modulation in the noise movies (i.e., 0.1 Hz). To this aim, we built PSTHs for the noise movies, by considering each of the 20 different movies we presented as a different trial of the same stimulus, so as to highlight the effect of contrast modulation (see examples of highly contrast modulated neurons in Fig. 3A, top). The MIc for each unit was computed over these PSTHs and only units with MIc >3 (i.e., units that were significantly contrast modulated) were included in the analysis. Furthermore, in order to ensure a robust estimation of the response time constants, we rejected units for which the R^2 of fit with the best exponential model was lower than 0.5.

Orientation decoding analysis

The goal of this analysis was to build four pseudo-populations of neurons – i.e., control simple (CS), control complex (CC), experimental simple (ES) and experimental complex (EC) cells – with similar distributions of orientation tuning and orientation preference, and then compare their ability to support stable decoding of the orientation of the gratings over time.

The pseudo-populations were built as follows. We first matched the control and experimental populations in terms of the sharpness of their orientation tuning. To this aim, we took the OSI distributions of the two populations (i.e., the blue and orange curves in Fig. 2C)

and, for each bin b in which the OSI axis had been divided (i.e., 10 equi-spaced bins of size = 0.1), we took as a reference the population with the lowest number of units N_b in that bin. For this population, all the N_b units were considered, while, for the other population, N_b units were randomly sampled (without replacement) from those with OSI falling in the bin b. Repeating this procedure for all the 10 bins, we obtained two down-sampled populations of control and experimental units, having all the same OSI distribution and the same number of units (n = 92). When considering separately the pools of simple and complex cells within such down-sampled populations, the resulting mean OSI were very similar (CS: 0.44 ± 0.04, n = 43; CC: 0.42 ± 0.03, n = 49; ES: 0.46 ± 0.03, n = 57; EC: 0.38 ± 0.04, n = 35) and not statistically different pairwise (two-tailed, unpaired t-test $p > 0.05$). Matching the four populations in terms of the OSI was essential, but not sufficient, to make sure that they had approximately the same power to support discrimination of the oriented gratings. In fact, the populations could still differ in terms of the distributions of orientation preference. To also equate them in this sense, and make sure that all possible orientations were equally discriminable, we replicated each unit 11 times, by circularly shifting its tuning curve of 11 incremental steps of 30º. This yielded four final pseudo-populations of 473 (CS), 539 (CC), 627 (ES) and 385 (EC) units, with matched orientation tuning and homogeneous orientation preference, to be used for the decoding analysis.

The latter worked as follows. From each pseudo-population, we sampled (without replacement) 300 units (referred to as *decoding pool* in what follows) and we built 300-dimensional population vectors having as components the responses (i.e., spike counts) of the sampled units in randomly selected presentations (i.e., trials) of either the 0°- or the 90°-oriented grating (drifting at 4 Hz), with each response computed in the same, randomly chosen 33 ms-wide time bin within the presentation epoch of the grating. More specifically, this time bin was

chosen under the constraint of being between 561 ms and 957 ms from the onset of stimulus presentation, so that the drifting grating continued for at least 2 full cycles (i.e., 561 ms) after the selected bin. The random sampling of the trial to be used in a given population vector was performed independently for each neuron (and without replacement), so as to get rid of any noise correlation among the units that were recorded in the same session. Given that 20 repeated trials were recorded per neuron and stimulus condition, a set of 20 population vectors was built for the 0°-oriented grating and another set for the 90°-oriented gratings. These vectors were used to train a binary logistic classifier to discriminate the two stimuli. The resulting classifier was then tested for its ability to discriminate the gratings in 33 ms-wide test bins that were increasingly distant (in time) from the training bin, covering two full cycles of the drifting gratings (i.e., from 33 to 561 ms following the training bin; see abscissa in Fig. 4B). This analysis was repeated for 50 random samplings (without replacement) of the decoding pools and, given a decoding pool, for 10 independent random draws (without replacement) of the training time bin. The resulting 500 accuracy curves were then averaged to yield the final estimate of the stability of the classification over time (solid curves in Fig. 4B).

To obtain 95% confidence intervals (shaded regions in Fig. 4B) for such average classification curves, we run a bootstrap analysis that worked as follows. For each of the four pseudo-populations, we sampled (with replacement) 50 surrogate populations and we used those to re-run the whole decoding analysis described in the previous paragraph. This yielded 50 bootstrap classification curves that were used to compute standard errors for the actual generalization curve. The standard errors were then converted into confidence intervals by multiplying them by the appropriate critical value 1.96 as described in (*93*).

LIST OF SUPPLEMENTARY MATERIALS

Fig. S1. The *virtual cage* system.

Fig. S2. The distributions of cortical depths and laminar locations did not differ between the populations of control and experimental units.

Fig. S3. Characterization of the response modulation of the control and experimental units using the $F1/F0$ ratio.

Fig. S4. Bootstrap analysis to verify the across-session consistency of the difference of complex cells found in the control and experimental rats.

Fig. S5. Postnatal rearing in temporally discontinuous visual environments leaves direction tuning of V1 neurons unaltered.

REFERENCES AND NOTES

1. E. P. Simoncelli, B. A. Olshausen, Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).

2. B. A. Olshausen, D. J. Field, Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14**, 481–487 (2004).

3. Y. Karklin, M. S. Lewicki, Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*. **457**, 83–86 (2009).

4. A. Hyvärinen, P. Hoyer, Emergence of Phase- and Shift-Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces. *Neural Comput.* **12**, 1705–1720 (2000).

5. B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. **381**, 607–9 (1996).

6. J. J. Hunt, P. Dayan, G. J. Goodhill, Sparse Coding Can Predict Primary Visual Cortex Receptive Field Changes Induced by Abnormal Visual Input. *PLOS Comput. Biol.* **9**, e1003005 (2013).

7. P. Foldiak, Learning invariance from transformation sequnces. *Neural Comput.* **3**, 194–200 (1991).

8. P. Berkes, L. Wiskott, Slow feature analysis yields a rich repertoire of complex cell properties. *J Vis*. **5**, 579–602 (2005).

9. L. Wiskott, T. J. Sejnowski, Slow feature analysis: unsupervised learning of invariances. *Neural Comput*. **14**, 715–70 (2002).

10. C. F. Cadieu, B. A. Olshausen, Learning Intermediate-Level Representations of Form and Motion from Natural Movies. *Neural Comput.* **24**, 827–866 (2011).

11. G. Wallis, Using Spatio-temporal Correlations to Learn Invariant Object Recognition. *Neural Netw.* **9**, 1513–1519 (1996).

12. G. Wallis, E. T. Rolls, Invariant face and object recognition in the visual system. *Prog Neurobiol*. **51**, 167–94 (1997).

13. W. Einhäuser, C. Kayser, P. König, K. P. Körding, Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur. J. Neurosci.* **15**, 475–486 (2002).

14. K. P. Körding, C. Kayser, W. Einhäuser, P. König, How Are Complex Cell Properties Adapted to the Statistics of Natural Stimuli? *J. Neurophysiol.* **91**, 206–212 (2004).

15. R. Wyss, P. König, P. F. M. J. Verschure, A Model of the Ventral Visual System Based on Temporal Stability and Local Memory. *PLOS Biol.* **4**, e120 (2006).

16. J. J. DiCarlo, D. Zoccolan, N. C. Rust, How Does the Brain Solve Visual Object Recognition? *Neuron*. **73**, 415–434 (2012).

17. D. H. Hubel, T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex. *J Physiol*. **195**, 215–43 (1968).

18. A. J. Bell, T. J. Sejnowski, The “independent components” of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338 (1997).

19. N. Berardi, T. Pizzorusso, L. Maffei, Critical periods during sensory development. *Curr. Opin. Neurobiol.* **10**, 138–145 (2000).

20. J. S. Espinosa, M. P. Stryker, Development and plasticity of the primary visual cortex. *Neuron*. **75**, 230–249 (2012).

21. C. Blakemore, G. F. Cooper, Development of the Brain depends on the Visual Environment. *Nature*. **228**, 477–478 (1970).

22. H. V. B. Hirsch, D. N. Spinelli, Visual Experience Modifies Distribution of Horizontally and Vertically Oriented Receptive Fields in Cats. *Science*. **168**, 869–871 (1970).

23. F. Sengpiel, P. Stawinski, T. Bonhoeffer, Influence of experience on orientation maps in cat visual cortex. *Nat. Neurosci.* **2**, 727–732 (1999).

24. S. Tanaka, J. Ribot, K. Imamura, T. Tani, Orientation-restricted continuous visual exposure induces marked reorganization of orientation maps in early life. *NeuroImage*. **30**, 462–477 (2006).

25. T. Yoshida, K. Ozawa, S. Tanaka, Sensitivity Profile for Orientation Selectivity in the Visual Cortex of Goggle-Reared Mice. *PLOS ONE*. **7**, e40630 (2012).

26. L. E. White, D. Fitzpatrick, Vision and Cortical Map Development. *Neuron*. **56**, 327–338 (2007).

27. Y. Li, S. D. Van Hooser, M. Mazurek, L. E. White, D. Fitzpatrick, Experience with moving visual stimuli drives the early development of cortical direction selectivity. *Nature*. **456**, 952–956 (2008).

28. Y. Li, D. Fitzpatrick, L. E. White, The development of direction selectivity in ferret visual cortex requires early visual experience. *Nat. Neurosci.* **9**, 676–681 (2006).

29. A. S. Hsu, P. Dayan, An unsupervised learning model of neural plasticity: Orientation selectivity in goggle-reared kittens. *Vision Res.* **47**, 2868–2877 (2007).

30. T. Ohshiro, S. Hussain, M. Weliky, Development of cortical orientation selectivity in the absence of visual experience with contour. *J. Neurophysiol.* **106**, 1923–1932 (2011).

31. J. N. Wood, A. Prasad, J. G. Goldman, S. M. W. Wood, Enhanced learning of natural visual sequences in newborn chicks. *Anim. Cogn.* **19**, 835–845 (2016).

32. J. N. Wood, S. M. W. Wood, The Development of Invariant Object Recognition Requires Visual Experience With Temporally Smooth Objects. *Cogn. Sci.* **42**, 1391–1406 (2018).

33. M. Cynader, G. Chernenko, Abolition of direction selectivity in the visual cortex of the cat. *Science*. **193**, 504–505 (1976).

34. M. Cynader, N. Berman, A. Hein, Cats Reared in Stroboscopic Illumination: Effects on Receptive Fields in Visual Cortex. *Proc. Natl. Acad. Sci.* **70**, 1353–1354 (1973).

35. C. R. Olson, J. D. Pettigrew, Single units in visual cortex of kittens reared in stroboscopic illumination. *Brain Res.* **70**, 189–204 (1974).

36. H. E. Pearson, N. Berman, E. H. Murphy, Critical periods in development for susceptibility to the effects of stroboscopic rearing in the rabbit visual cortex. *Exp. Brain Res.* **50**, 367–372 (1983).

37. A. L. Humphrey, A. B. Saul, Strobe Rearing Reduces Direction Selectivity in Area 17 by Altering Spatiotemporal Receptive-Field Structure. *J. Neurophysiol.* **80**, 2991–3004 (1998).

38. H. Kennedy, G. A. Orban, Response properties of visual cortical neurons in cats reared in stroboscopic illumination. *J. Neurophysiol.* **49**, 686–704 (1983).

39. T. Pasternak, R. A. Schumer, M. S. Gizzi, J. A. Movshon, Abolition of visual cortical direction selectivity affects visual behavior in cats. *Exp. Brain Res.* **61**, 214–217 (1985).

40. M. Fagiolini, T. Pizzorusso, N. Berardi, L. Domenici, L. Maffei, Functional postnatal development of the rat primary visual cortex and the role of visual experience: Dark rearing and monocular deprivation. *Vision Res.* **34**, 709–720 (1994).

41. R. A. Williams, C. H. Pollitz, J. C. Smith, T. P. Williams, Flicker detection in the albino rat following light-induced retinal damage. *Physiol. Behav.* **34**, 259–266 (1985).

42. A. Bartoletti, P. Medini, N. Berardi, L. Maffei, Environmental enrichment prevents effects of dark-rearing in the rat visual cortex. *Nat. Neurosci.* **7**, 215–216 (2004).

43. A. Sale, N. Berardi, L. Maffei, Environment and Brain Plasticity: Towards an Endogenous Pharmacotherapy. *Physiol. Rev.* **94**, 189–234 (2014).

44. G. Matteucci, R. B. Marotti, M. Riggi, F. B. Rosselli, D. Zoccolan, Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. *J. Neurosci.* **39**, 1649–1670 (2019).

45. S. Tafazoli, H. Safaai, G. De Franceschi, F. B. Rosselli, W. Vanzella, M. Riggi, F. Buffolo, S. Panzeri, D. Zoccolan, Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife*. **6**, e22794 (2017).

46. C. M. Niell, M. P. Stryker, Highly selective receptive fields in mouse visual cortex. *J Neurosci*. **28**, 7520–36 (2008).

47. M. Wypych, C. Wang, A. Nagy, G. Benedek, B. Dreher, W. J. Waleszczyk, Standardized F1 – A consistent measure of strength of modulation of visual responses to sine-wave drifting gratings. *Vision Res.* **72**, 14–33 (2012).

48. N. Li, J. J. DiCarlo, Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex. *Science*. **321**, 1502–1507 (2008).

49. N. Li, J. J. DiCarlo, Unsupervised Natural Visual Experience Rapidly Reshapes Size-Invariant Object Representation in Inferior Temporal Cortex. *Neuron*. **67**, 1062–1075 (2010).

50. Y. Miyashita, Inferior Temporal Cortex: Where Visual Perception Meets Memory. *Annu. Rev. Neurosci.* **16**, 245–263 (1993).

51. J. A. Movshon, I. D. Thompson, D. J. Tolhurst, Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol.* **283**, 79–99 (1978).

52. B. C. Skottun, R. L. De Valois, D. H. Grosof, J. A. Movshon, D. G. Albrecht, A. B. Bonds, Classifying simple and complex cells on the basis of response modulation. *Vision Res.* **31**, 1078–1086 (1991).

53. J. L. Hoy, C. M. Niell, Layer-Specific Refinement of Visual Cortex Function after Eye Opening in the Awake Mouse. *J. Neurosci.* **35**, 3370–3383 (2015).

54. Y. Chen, S. Anand, S. Martinez-Conde, S. L. Macknik, Y. Bereshpolova, H. A. Swadlow, J. M. Alonso, The linearity and selectivity of neuronal responses in awake visual cortex. *J. Vis.* **9**, 12–12 (2009).

55. F. Mechler, D. L. Ringach, On the classification of simple and complex cells. *Vision Res.* **42**, 1017–1033 (2002).

56. N. J. Priebe, F. Mechler, M. Carandini, D. Ferster, The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nat. Neurosci.* **7**, 1113–1122 (2004).

57. J. H. Siegle, X. Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T. K. Ramirez, H. Choi, J. A. Luviano, P. A. Groblewski, R. Ahmed, A. Arkhipov, A. Bernard, Y. N. Billeh, D. Brown, M. A. Buice, N. Cain, S. Caldejon, L. Casal, A. Cho, M. Chvilicek, T. C. Cox, K. Dai, D. J. Denman, S. E. J. de Vries, R. Dietzman, L. Esposito, C. Farrell, D. Feng, J. Galbraith, M. Garrett, E. C. Gelfand, N. Hancock, J. A. Harris, R. Howard, B. Hu, R. Hytnen, R. Iyer, E. Jessett, K. Johnson, I. Kato, J. Kiggins, S. Lambert, J. Lecoq, P. Ledochowitsch, J. H. Lee, A. Leon, Y. Li, E. Liang, F. Long, K. Mace, J. Melchior, D. Millman, T. Mollenkopf, C. Nayan, L. Ng, K. Ngo, T. Nguyen, P. R. Nicovich, K. North, G. K. Ocker, D. Ollerenshaw, M. Oliver, M. Pachitariu, J. Perkins, M. Reding, D. Reid, M. Robertson, K. Ronellenfitch, S. Seid, C. Slaughterbeck, M. Stoecklin, D. Sullivan, B. Sutton, J. Swapp, C. Thompson, K. Turner, W. Wakeman, J. D. Whitesell, D. Williams, A. Williford, R. Young, H. Zeng, S. Naylor, J. W. Phillips, R. C. Reid, S. Mihalas, S. R. Olsen, C. Koch, "A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas" (preprint, Neuroscience, 2019), , doi:10.1101/805010.

58. M. E. Garrett, S. Manavi, K. Roll, D. R. Ollerenshaw, P. A. Groblewski, J. Kiggins, X. Jia, L. Casal, K. Mace, A. Williford, A. Leon, S. Mihalas, S. R. Olsen, "Experience shapes activity dynamics and stimulus coding of VIP inhibitory and excitatory cells in visual cortex" (preprint, Neuroscience, 2019), , doi:10.1101/686063.

59. S. D. Van Hooser, A. Roy, H. J. Rhodes, J. H. Culp, D. Fitzpatrick, Transformation of Receptive Field Properties from Lateral Geniculate Nucleus to Superficial V1 in the Tree Shrew. *J. Neurosci.* **33**, 11494–11505 (2013).

60. M. Riesenhuber, T. Poggio, Models of object recognition. *Nat Neurosci*. **3 Suppl**, 1199–204. (2000).

61. D. J. Wallace, D. S. Greenberg, J. Sawinski, S. Rulla, G. Notaro, J. N. D. Kerr, Rats maintain an overhead binocular field at the expense of constant fusion. *Nature*. **498**, 65–69 (2013).

62. W. Vanzella, N. Grion, D. Bertolini, A. Perissinotto, D. Zoccolan, A passive, camera-based head-tracking system for real-time, 3D estimate of head position and orientation in rodents. *bioRxiv*, 599365 (2019).

63. A. F. Meyer, J. Poort, J. O'Keefe, M. Sahani, J. F. Linden, A Head-Mounted Camera System Integrates Detailed Behavioral Monitoring with Multichannel Electrophysiology in Freely Moving Mice. *Neuron*. **100**, 46-60.e7 (2018).

64. C. M. Niell, M. P. Stryker, Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*. **65**, 472–9 (2010).

65. A. B. Saleem, A. Ayaz, K. J. Jeffery, K. D. Harris, M. Carandini, Integration of visual motion and locomotion in mouse visual cortex. *Nat. Neurosci.* **16**, 1864–1869 (2013).

66. A. B. Saleem, E. M. Diamanti, J. Fournier, K. D. Harris, M. Carandini, Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature*. **562**, 124 (2018).

67. G. Guitchounts, J. Masis, S. B. Wolff, D. Cox, *bioRxiv*, in press, doi:10.1101/2020.01.16.909473.

68. J. H. van Hateren, D. L. Ruderman, Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* **265**, 2315–2320 (1998).

69. A. M. Zador, A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* **10**, 3770 (2019).

70. D. H. Hubel, T. N. Wiesel, RECEPTIVE FIELDS OF CELLS IN STRIATE CORTEX OF VERY YOUNG, VISUALLY INEXPERIENCED KITTENS. *J. Neurophysiol.* **26**, 994–1002 (1963).

71. D. H. Hubel, T. N. Wiesel, S. LeVay, Functional Architecture of Area 17 in Normal and Monocularly Deprived Macaque Monkeys. *Cold Spring Harb. Symp. Quant. Biol.* **40**, 581–589 (1976).

72. H. Sherk, M. P. Stryker, Quantitative study of cortical orientation selectivity in visually inexperienced kitten. *J. Neurophysiol.* **39**, 63–70 (1976).

73. N. L. Rochefort, M. Narushima, C. Grienberger, N. Marandi, D. N. Hill, A. Konnerth, Development of Direction Selectivity in Mouse Cortical Neurons. *Neuron*. **71**, 425–432 (2011).

74. K. M. Hagihara, T. Murakami, T. Yoshida, Y. Tagawa, K. Ohki, Neuronal activity is not required for the initial formation and maturation of visual selectivity. *Nat. Neurosci.* **18**, 1780–1788 (2015).

75. A. Tiriac, B. E. Smith, M. B. Feller, Light Prior to Eye Opening Promotes Retinal Waves and Eye-Specific Segregation. *Neuron*. **100**, 1059-1065.e4 (2018).

76. M. V. Albert, A. Schnabel, D. J. Field, Innate Visual Learning through Spontaneous Activity Patterns. *PLoS Comput. Biol.* **4**, e1000137 (2008).

77. J. B. Ackman, T. J. Burbridge, M. C. Crair, Retinal waves coordinate patterned activity throughout the developing visual system. *Nature*. **490**, 219–225 (2012).

78. T. McLaughlin, C. L. Torborg, M. B. Feller, D. D. M. O'Leary, Retinotopic Map Refinement Requires Spontaneous Retinal Waves during a Brief Critical Period of Development. *Neuron*. **40**, 1147–1160 (2003).

79. A. Thompson, A. Gribizis, C. Chen, M. C. Crair, Activity-dependent development of visual receptive fields. *Curr. Opin. Neurobiol.* **42**, 136–143 (2017).

80. G. J. Goodhill, Theoretical Models of Neural Development. *iScience*. **8**, 183–199 (2018).

81. J. J. Hunt, M. Ibbotson, G. J. Goodhill, Sparse Coding on the Spot: Spontaneous Retinal Waves Suffice for Orientation Selectivity. *Neural Comput.* **24**, 2422–2433 (2012).

82. S. Dähne, N. Wilbert, L. Wiskott, Slow Feature Analysis on Retinal Waves Leads to V1 Complex Cells. *PLoS Comput Biol*. **10**, e1003564 (2014).

83. F. Creutzig, H. Sprekeler, Predictive Coding and the Slowness Principle: An Information-Theoretic Approach. *Neural Comput.* **20**, 1026–1041 (2007).

84. M. Chalk, O. Marre, G. Tkačik, Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci.* **115**, 186–191 (2018).

85. B. Weghenkel, L. Wiskott, Slowness as a Proxy for Temporal Predictability: An Empirical Comparison. *Neural Comput.* **30**, 1151–1179 (2018).

86. W. Bialek, R. R. D. R. V. Steveninck, N. Tishby, in *2006 IEEE International Symposium on Information Theory* (2006), pp. 659–663.

87. J.-P. Lies, R. M. Häfner, M. Bethge, Slowness and Sparseness Have Diverging Effects on Complex Cell Learning. *PLOS Comput. Biol.* **10**, e1003468 (2014).

88. J. N. Wood, Newborn chickens generate invariant object representations at the onset of visual object experience. *Proc. Natl. Acad. Sci.* **110**, 14000–14005 (2013).

89. G. Matteucci, M. Riggi, D. Zoccolan, A template-matching algorithm for laminar identification of cortical recording sites from evoked response potentials. *bioRxiv*, 749069 (2019).

90. C. Rossant, S. N. Kadir, D. F. M. Goodman, J. Schulman, M. L. D. Hunter, A. B. Saleem, A. Grosmark, M. Belluscio, G. H. Denfield, A. S. Ecker, A. S. Tolias, S. Solomon, G. Buzsáki, M. Carandini, K. D. Harris, Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).

91. O. Schwartz, J. W. Pillow, N. C. Rust, E. P. Simoncelli, Spike-triggered neural characterization. *J. Vis.* **6** (2006), doi:10.1167/6.4.13.

92. T. O. Sharpee, Computational Identification of Receptive Fields. *Annu. Rev. Neurosci.* **36**, 103–120 (2013).

93. B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall/CRC, Boca Raton, FL, 1998).

ACKNOWLEDGEMENT

General: We thank Mattia D'Andola, Nelly Redolfi and Margherita Riggi for their help with the rearing of rats in the visually controlled environments. We thank Mattia D'Andola for his support in data collection. We thank Daniele Bertolini for his support in designing and building the controlled-rearing apparatus. We thank Laurenz Wiskott, Tomaso Poggio, James DiCarlo, Matteo Carandini, Concetta Morrone and Andrea Benucci for their feedback on the manuscript. **Funding:** This work was supported by a European Research Council Consolidator Grant (DZ, project n. 616803-LEARN2SEE). **Author contributions:** GM and DZ designed the study. GM performed and analyzed the neuronal recordings. GM wrote the code for the analyses. GM and DZ wrote the manuscript. GM and DZ revised the manuscript. **Competing Interests:** The authors declare no competing interests. **Data and materials availability:** Code and data to reproduce the analysis will be made available upon reasonable request via e-mail to the corresponding author.

FIGURE LEGENDS

Fig. 1. Experimental design. (**A**) Two groups of rats, control (top) and experimental (bottom), were born in dark and housed in lightproof cabinets till eye opening (black bars). Afterward, the

control rats were subjected to daily, 4-hours long exposures to natural videos inside the virtual cages (blue bar), while the experimental rats were subjected to the frame-scrambled versions of the same movies (orange bar). Starting from P60, neuronal recordings from V1 of both the control and experimental rats were performed under anesthesia (gray bars), while the animals were exposed to drifting gratings and movies of spatially and temporally correlated noise. (**B**) Left: the mean correlation between the image frames of one of the natural movie (same as in A) is plotted as a function of their temporal lag (blue curve). The dashed line shows the best exponential fit to the resulting autocorrelation function (τ is the time constant of the fit). Right: The autocorrelation function obtained for the frame-scrambled version of the movie (orange curve) is shown along with its best exponential fit (dashed line). (**C**) Left: the autocorrelation functions of all the natural movies used during postnatal rearing of the control rats (blue curves) are shown along with the autocorrelation functions of their frame-scrambled versions (orange curves), used during postnatal rearing of the experimental rats. Right: the time constants of the best exponential fits to the autocorrelation functions of the natural movies (blue dots) and their frame-scrambled versions (orange dots) were significantly different ($p < 0.001$; 1-tailed, unpaired t-test).

Fig. 2. Postnatal rearing in temporally discontinuous visual environments results in an impoverished population of V1 complex cells, but spares the development of orientation tuning. (**A**) A representative V1 complex cell of the control group (Left: blue lines) is compared to a representative simple cell of the experimental group (Right: orange lines). For each neuron, the graph shows, from top/left to bottom: 1) the linear RF structure inferred through STA; 2) the direction tuning curve; 3) the raster plot, with the number of spikes (dots) fired across repeated

presentations of most effective grating stimulus; 4) the corresponding peri-stimulus time histogram (PSTH), computed in 10 ms-wide time bins; and 5) its power spectrum, with indicated its mean (dotted line), its mean plus SD (dashed line), and the 4 HZ frequency of the grating stimulus (vertical line). (**B**) Left: distributions of the modulation index (MI) used to distinguish the poorly phase-modulated complex cells (MI < 3; gray-shaded area) from the strongly modulated simple cells (MI > 3), as obtained for the control (blue; n = 50) and experimental (orange; n = 75) V1 populations (only units with OSI > 0.4 included). Both the distributions and their medians (dashed lines) were significantly different ($p < 0.02$, Kolmogorov-Smirnov test; $p < 0.001$, Wilcoxon test). Right: the fraction of units that were classified as complex cells (i.e., with MI < 3) was significantly larger for the control than for the experimental group ($p < 0.001$, Fisher exact-test). (**C**) Left: distributions of the orientation selectivity index (OSI), as obtained for the control (blue; n = 105) and experimental (orange; n = 158) V1 populations. No significant difference was found between the two distributions and their medians ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, Wilcoxon test). Right: the fraction of sharply orientation-tuned units (i.e., units with OSI > 0.6) did not differ between the two groups ($p > 0.05$, Fisher exact-test).

Fig. 3. Unaltered spatial tuning of V1 neurons following the controlled rearing in temporally discontinuous visual environments. (**A**) Examples of linear RFs inferred through STA for the control (blue frame) and experimental group (orange frame). In every STA image, each pixel intensity value was independently z scored, based on the null distribution of STA values obtained through a permutation test (see Materials and Methods). (**B**) Left: distributions of the contrast index (CI) used to measure the sharpness of the STA images, as obtained for the control (blue; n = 105) and experimental (orange; n = 158) V1 populations. No significant

difference was found between the two distributions and their medians ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, Wilcoxon test). Right: Mean values (± SEM) of the CIs, computed separately for the simple (dark bars; $n = 20$, control and $n = 49$, experimental) and complex (light bars; $n = 30$, control and $n = 26$, experimental) cells of the two groups (only units with OSI > 0.4 included). Within each group, the mean CI was significantly larger for the simple than for the complex cells (** $p < 0.01$, two-tailed, unpaired t test). (**C**) The color maps showing the distributions of lobe counts (abscissa) for the units with well-defined linear RFs (i.e., within the top quartiles of the CI distributions shown in B) in the control (blue; $n = 27$) and experimental (orange; $n = 37$) populations are plotted as a function of the binarization threshold (ordinate) used by the lobe-counting algorithm (see Materials and Methods). For every choice of the threshold, the control and experimental distributions were not significantly different ($p > 0.05$; Fisher exact test). (**D**) Same analysis as in C, but applied to the distributions of RF sizes obtained for the control (blue; $n = 27$) and experimental (orange; $n = 37$) populations, as a function of the binarization threshold. Again, no significant difference was found between the two distributions at any threshold level ($p > 0.05$; Fisher exact test).

Fig. 4. Postnatal rearing in temporally discontinuous visual environments leads to the development of complex cells with abnormally fast response dynamics. (**A**) Top: PSTHs showing the average responses of the two example neurons of Fig. 2A to the contrast-modulated noise movies (see Materials and Methods). For both neurons, the firing rate was strongly modulated at the frequency of variation of the contrast of the movies (i.e., 0.1 Hz). Bottom: distributions of inter-spike intervals (ISIs) of the spike trains evoked by the noise movies for the two example neurons. The resulting autocorrelograms were fitted with exponential decaying

functions (dashed lines) to measure the slowness (i.e., the time constant τ of the exponential fit) of the responses. In this example, the complex cell (blue curve) displays slower dynamics (i.e., larger τ) than the simple cell (orange curve). The two units also differ in the number of counts at low ISIs, which is much larger for the simple cell, as expected for a unit firing tightly packed trains of spikes (see Fig. 2A). (**B**) Mean values (± SEM) of the time constants τ, computed for the control (blue; n = 92) and experimental (orange; n = 143) populations (*** $p < 0.001$, two-tailed, unpaired t test). (**C**) Mean values (± SEM) of the time constants τ, computed separately for the simple (dark bars; n = 43, control and n = 89, experimental) and complex (light bars; n = 49, control and n = 54, experimental) cells of the two groups. While the simple cells had equally fast dynamics, the complex cells were significantly slower in the control than in the experimental group (** $p < 0.01$, two-tailed, unpaired t test).

Fig. 5. Postnatal rearing in temporally discontinuous visual environments reduces the ability of complex cells to support phase-tolerant discrimination of grating orientation. (**A**) The cartoon illustrates the expected outcome of the decoding analysis to test the ability of simple and complex cells to support phase-tolerant discrimination of grating orientation. In the case of simple cells (top), a linear classifier built at time t_0 (middle: light gray shading) to successfully discriminate a vertical from a horizontal drifting grating (left: the filled and empty dots are well separated, within the neuronal representational space, by the linear decision boundary) will generalize poorly, when tested at a later time t_1 (middle: dark gray shading), with the accuracy dropping even below chance (right: the filled and empty dots swap sides of the linear decision boundary), because of the strong phase dependency of the responses r_i (middle: some neurons firing at t_0 will stop firing at t_1, while some other units that are silent at t_0 will respond at t_1).

By contrast, for a population of complex cells (bottom), given the larger stability of the responses r_i (middle), the decision boundary resulting from the training at t_0 (left) will generalize better at t_1 (right: the filled and empty dots are still mostly on the original side of the decision boundary). (**B**) Decoding accuracy yielded by the four populations of simple control (dark blue), complex control (light blue), simple experimental (brown) and complex experimental (orange) cells, in the vertical (i.e., 90º) vs. horizontal (i.e., 0º) grating discrimination task, in 33 ms-wide test bins located at increasingly larger time lags from the training bin (i.e., bin with lag = 0). The solid curves are the averages of many resampling loops of the neuronal population vectors and the training bins (see Materials and Methods). The shaded regions are the bootstrap-estimated 95% confidence intervals of the solid lines (see Materials and Methods).

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. The *virtual cage* system. CAD drawing of a *virtual cage* – the custom apparatus used to rear newborn rats in visually controlled environments. The cage consists of a metal, box-like structure (light gray) holding 4 computer monitors (black/dark gray) that fully surround a transparent basin (light blue), where a newborn rat can be placed for immersive exposure to the visual scenes/movies displayed on the monitors.

Fig. S2. The distributions of cortical depths and laminar locations did not differ between the populations of control and experimental units. (**A**) The distributions of cortical depths (left) and their means (right) did not significantly differ between the populations of control (blue;

n = 105) and experimental (orange; n = 158) units ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, two-tailed, unpaired t-test). Error bars are SEMs. (**B**) The distributions of laminar locations were also not significantly different between control (blue) and experimental (orange) units ($p > 0.05$, χ^2 test). Both the cortical depths and laminar locations were estimated from the patterns of visually evoked potentials recorded along the silicon probes (see Materials and Methods).

Fig. S3. Characterization of the response modulation of the control and experimental units using the *F*1/*F*0 ratio. (**A**) Left: distributions of the $F1/F0$ ratio used, as an alternative to the MI index of Fig. 2B, to quantify the level of response modulation of the control (blue; n = 50) and experimental (orange; n = 75) V1 populations (only units with OSI > 0.4 included). Following the convention in the literature, the $F1/F0 = 1$ threshold was used to distinguish complex (gray-shaded area) from simple cells. The medians of the distributions (dashed lines) were significantly different ($p < 0.05$, Wilcoxon test). Right: the fraction of units that were classified as complex cells (i.e., with $F1/F0 < 1$) was significantly larger for the control than for the experimental group ($p < 0.05$, Fisher exact-test). (**B**) Same analysis as in A, but applied to a modified version of the $F1/F0$ ratio (referred to as $F1/F0^*$ ratio; see Materials and Methods for a definition). Same significance levels/tests as in A.

Fig. S4. Bootstrap analysis to verify the across-session consistency of the difference of complex cells found in the control and experimental rats. (**A**) Distributions of the fraction of complex cells found in the control (blue) and experimental (orange) groups, as obtained by sampling (with replacement) 100 times the available sessions for the two groups (see Results for details). The differences between the pairs of complex cells' fractions (i.e., control minus

experimental) obtained across the 100 bootstrap samplings were used to compute the confidence interval shown in B. (**B**) The difference between the fractions of complex cells found in the control and experimental groups (dot) is shown along the 95% confidence interval of such difference (bar), as obtained by the bootstrap analysis described in A.

Fig. S5. Postnatal rearing in temporally discontinuous visual environments leaves direction tuning of V1 neurons unaltered. Left: distributions of the direction selectivity index (DSI) used to measure the sharpness of orientation tuning, as obtained for the control (blue; n = 105) and experimental (orange; n = 158) V1 populations. No significant difference was found between the two distributions and their medians ($p > 0.05$, Kolmogorov-Smirnov test; $p > 0.05$, Wilcoxon test). Right: the fraction of sharply orientation-tuned units (i.e., units with DSI > 0.6) did not differ between the two groups ($p > 0.05$, Fisher exact-test).

FIGURES

Figure 1

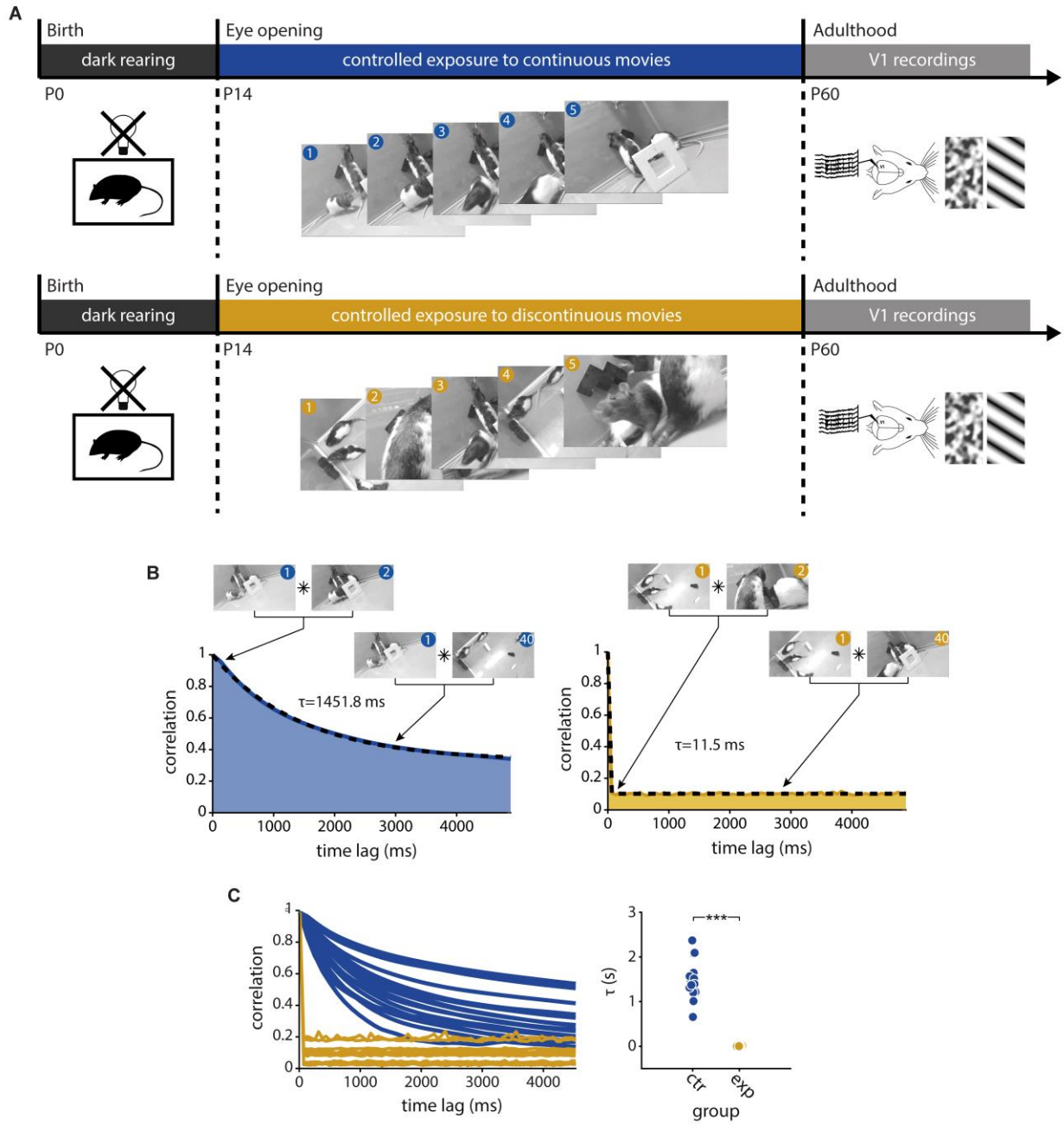


Figure 2

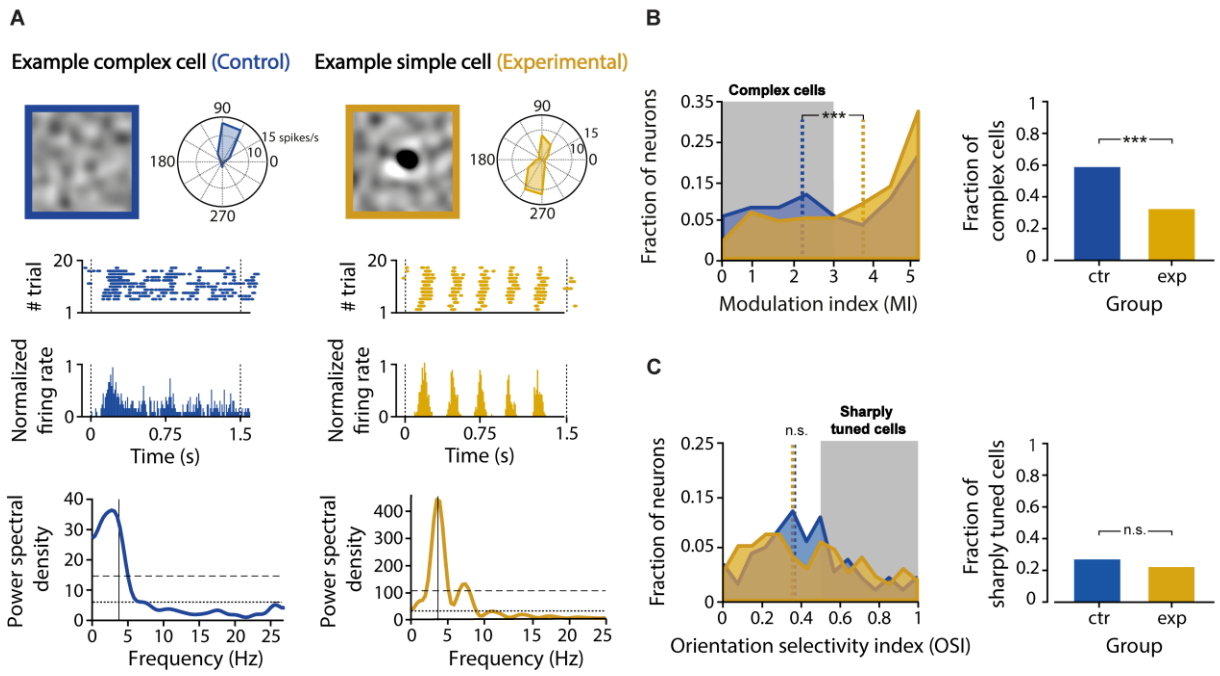

A
Example complex cell (Control)
Example simple cell (Experimental)
90
180
270
15 spikes/s
10
0
15
trial
Normalized firing rate
Time (s)
Power spectral density
Frequency (Hz)
B
Complex cells
***
Fraction of neurons
Modulation index (MI)
Fraction of complex cells
ctr
exp
Group
C
Sharply tuned cells
n.s.
Orientation selectivity index (OSI)
Fraction of sharply tuned cells

Figure 3

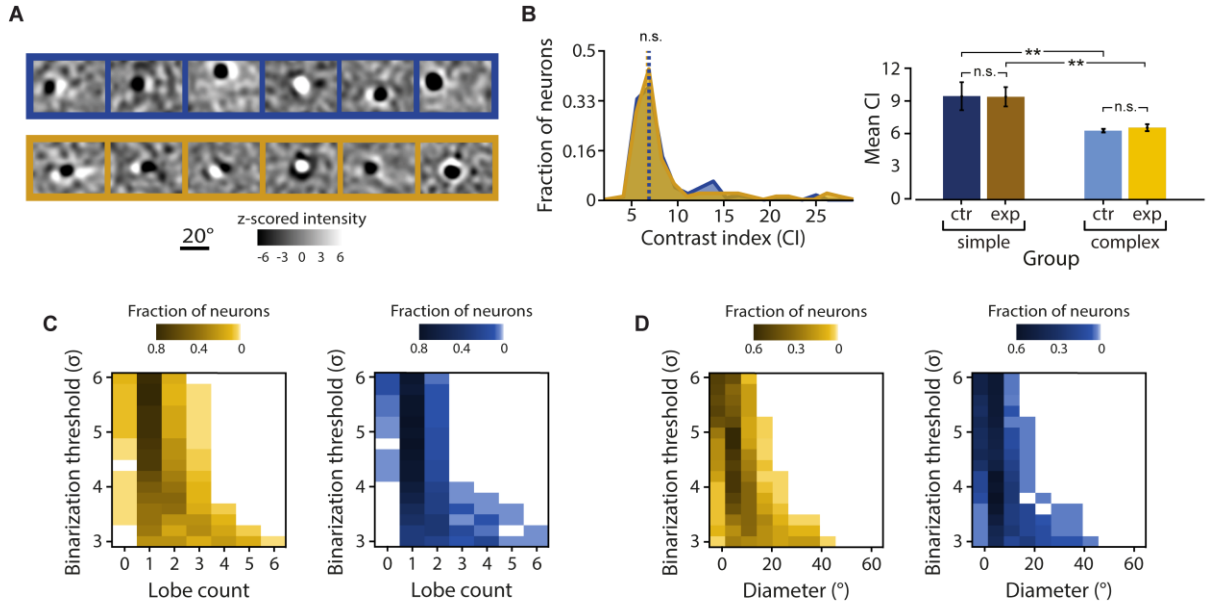

A
20°
z-scored intensity
-6 -3 0 3 6
B
Fraction of neurons
0.5
0.33
0.16
0
n.s.
5 10 15 20 25
Contrast index (CI)
Mean CI
12
9
6
3
0
n.s.
**
**
n.s.
ctr exp
simple
ctr exp
complex
Group
C
Fraction of neurons
0.8 0.4 0
Binarization threshold (σ)
6 5 4 3
0 1 2 3 4 5 6
Lobe count
Fraction of neurons
0.8 0.4 0
Binarization threshold (σ)
6 5 4 3
0 1 2 3 4 5 6
Lobe count
D
Fraction of neurons
0.6 0.3 0
Binarization threshold (σ)
6 5 4 3
0 20 40 60
Diameter (°)
Fraction of neurons
0.6 0.3 0
Binarization threshold (σ)
6 5 4 3
0 20 40 60
Diameter (°)

Figure 4

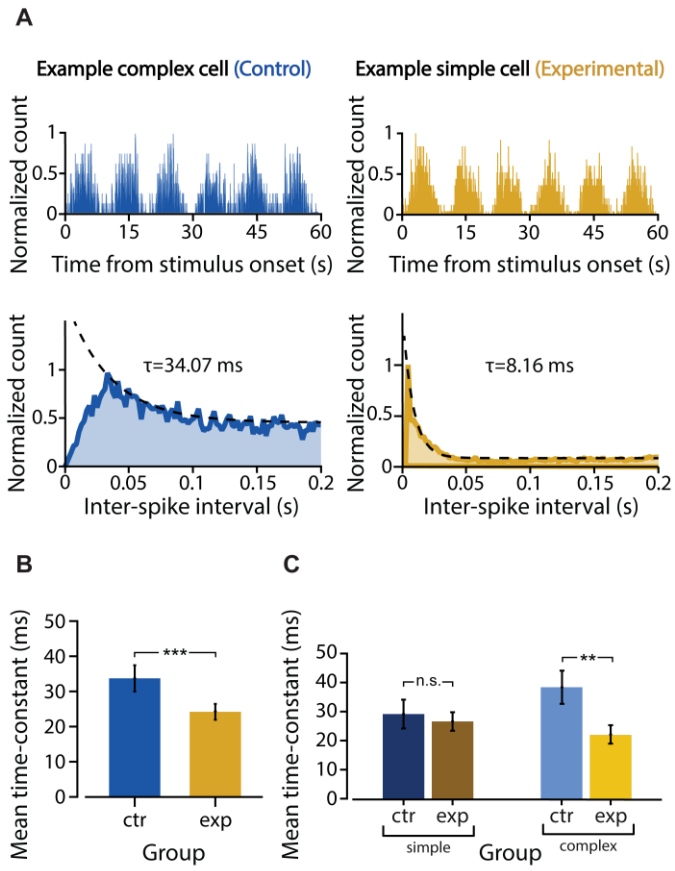
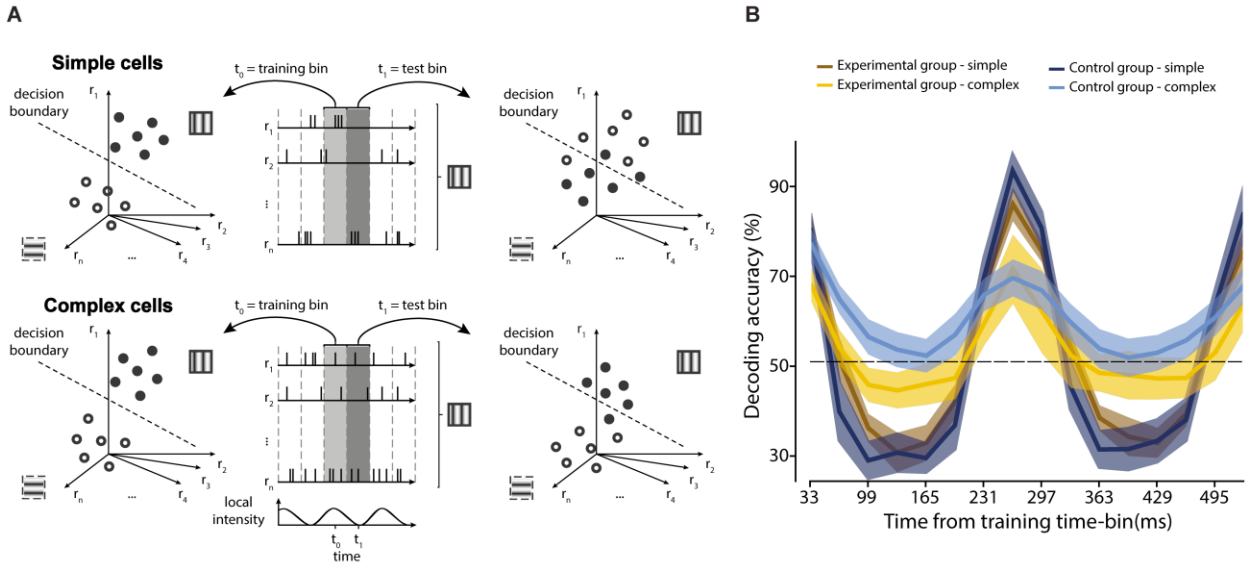

A
Example complex cell (Control)
Example simple cell (Experimental)
Normalized count
Time from stimulus onset (s)
τ=34.07 ms
τ=8.16 ms
Inter-spike interval (s)
B
C
Mean time-constant (ms)
***
n.s.
**
ctr
exp
Group
simple
complex

Figure 5


A
Simple cells
decision boundary
t_0 = training bin
t_1 = test bin
Complex cells
local intensity
time
B
Experimental group - simple
Experimental group - complex
Control group - simple
Control group - complex
Decoding accuracy (%)
Time from training time-bin(ms)

SUPPLEMENTARY FIGURES

Figure S1

Figure S2

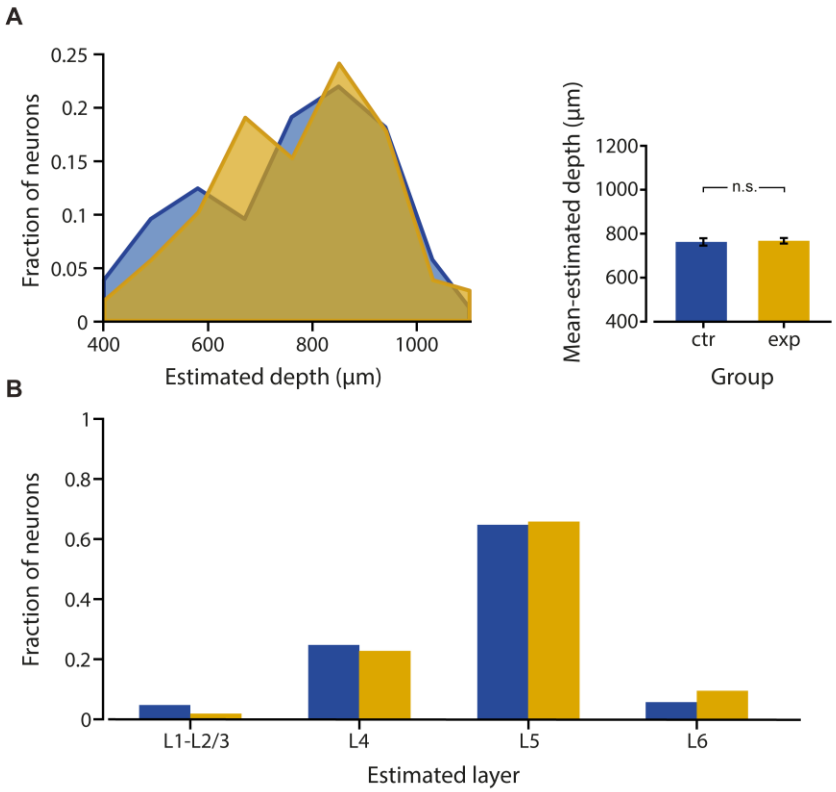

A
Fraction of neurons
0.25
0.2
0.15
0.1
0.05
0
400
600
800
1000
Estimated depth (µm)
Mean-estimated depth (µm)
1200
1000
800
600
400
n.s.
ctr
exp
Group
B
Fraction of neurons
1
0.8
0.6
0.4
0.2
0
L1-L2/3
L4
L5
L6
Estimated layer

Figure S3

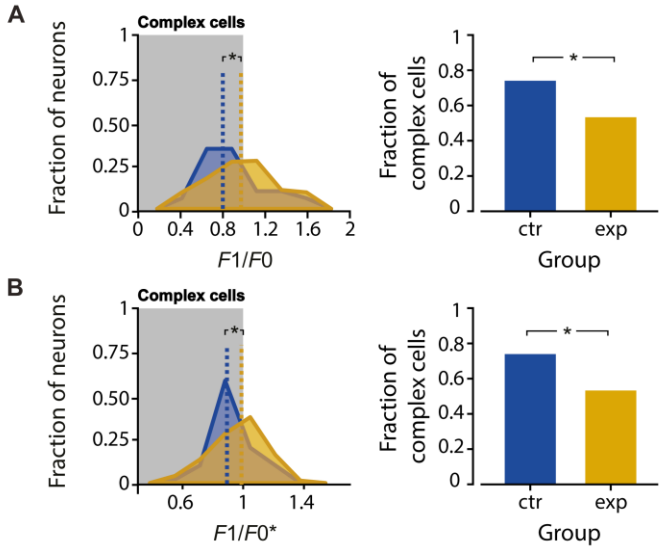

A
Complex cells
Fraction of neurons
1
0.75
0.50
0.25
0
0
0.4
0.8
1.2
1.6
2
F1/F0
Fraction of complex cells
1
0.8
0.6
0.4
0.2
0
*
ctr
exp
Group
B
Complex cells
Fraction of neurons
1
0.75
0.50
0.25
0
0.6
1
1.4
F1/F0*
Fraction of complex cells
1
0.8
0.6
0.4
0.2
0
*
ctr
exp
Group

Figure S4

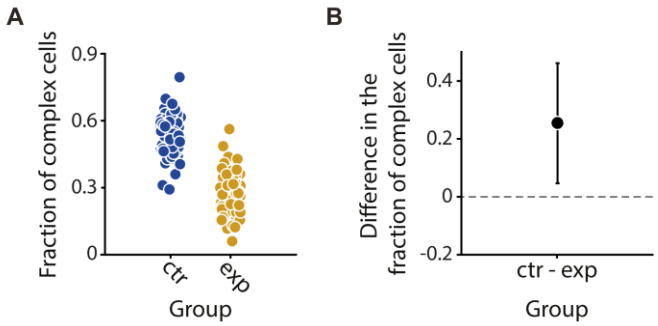
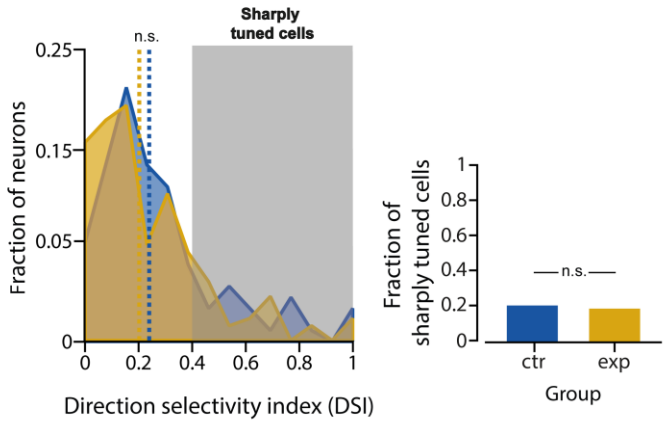

A
Fraction of complex cells
0.9
0.6
0.3
0
ctr
exp
Group
B
Difference in the fraction of complex cells
0.4
0.2
0
-0.2
ctr - exp
Group

Figure S5


Sharply tuned cells
n.s.
0.25
0.15
0.05
0
Fraction of neurons
0
0.2
0.4
0.6
0.8
1
Direction selectivity index (DSI)
1
0.8
0.6
0.4
0.2
0
Fraction of sharply tuned cells
n.s.
ctr
exp
Group