Intrinsic dynamics enhance temporal stability of stimulus representation along a visual cortical hierarchy

Eugenio Piasini[1,#], Liviu Soltuzu[2,#,¶], Riccardo Caramellino[2], Vijay Balasubramanian[1] and Davide Zoccolan[2,*]

[1]*Computational Neuroscience Initiative, University of Pennsylvania, Philadelphia, PA 19104, United States*

[2]*Visual Neuroscience Lab, International School for Advanced Studies (SISSA), Trieste, 34136, Italy*

[#]*these authors contributed equally.*
[¶] *current address: Blue Brain Project, Ecole Polytechnique Fédérale de Lausanne, Geneva, Switzerland*

*Corresponding author:

Davide Zoccolan
International School for Advanced Studies (SISSA)
Via Bonomea 265, 34136 Trieste (TS) ITALY
Office: +39 040 3787 701 Fax: +39 040 3787 702
E-mail: zoccolan@sissa.it

Summary

Along the ventral stream, cortical representations of brief, static stimuli become gradually more invariant to identity-preserving transformations. In the presence of long, ethologically relevant dynamic stimuli, higher invariance should imply temporally persistent representations at the top of this functional hierarchy. However, such stimuli could engage adaptive and predictive processes, whose impact on neural coding dynamics is unknown. More generally, coding dynamics in the presence of temporally structured stimuli are not understood. By probing the rodent analogue of the ventral stream with movies, we uncovered a hierarchy of temporal scales along this pathway, with deeper areas encoding visual information more persistently. Furthermore, the impact of intrinsic dynamics on the stability of stimulus representations gradually grows along the hierarchy. These results suggest that feedforward computations in the cortical hierarchy build up invariance even for dynamic, temporally structured stimuli, and that intrinsic processing contributes to the stabilization of representations in noisy, changing environments.

Introduction

In the mammalian brain, object vision is supported by neural representations of visual objects that are invariant to identity-preserving transformations, such as translations and changes of scale. A popular conceptual framework posits that such invariance emerges progressively, as the outcome of largely feedforward processing, along a hierarchy of visual cortical areas that, in primates, is known as the ventral stream [1-3]. This invariance hypothesis has been extensively validated in primates [4-8] and, more recently, in rodents [9-13]. However, existing studies have mainly focused on brief, static visual stimuli, designed to engage feedforward processing. Thus, it is not clear to what extent this conceptual framework applies to the dynamic scenes that are processed in natural vision.

Various theories suggest that temporal continuity of the visual input is a key statistical structure that could be exploited by unsupervised learning mechanisms to build invariant representations [14-23]. Despite mathematical differences, these proposals share a key intuition: behaviorally relevant, high-order features of the stimulus, such as object identity/category, tend to be preserved over longer timescales than low-level features (e.g., pixel/retinal level representations). Therefore, useful representations can be discovered by learning to extract features that change slowly over time within a sensory stream. Overall then, if the ventral stream progressively computes representations of the visual input with increasing invariance, we hypothesize a matching increase of the persistence of neural responses along the hierarchy, driven mainly by the fact that higher areas represent features of the input that change more slowly in time. This standard account relies on feedforward processing, and generally does not attribute a major role to adaptive, recurrent and feedback (e.g., predictive) processes.

However, mechanisms beyond simple feedforward processing could significantly modify the temporal dynamics of population codes. In the early visual pathway, retina and LGN implement temporal decorrelation of input signals, which has been argued to be metabolically advantageous [24]. Adaptation has been shown to depress visual cortical responses to repeated presentation of perceptually similar visual stimuli in both monkeys and humans [25-27], resulting in a reduced

discriminability of the adapted stimuli [28]. fMRI evidence in humans suggests that the impact of adaptation mechanisms may increase along the ventral stream [29-31] and that neural responses in higher areas of the hierarchy may be driven more strongly by transient than sustained stimuli [30; 32]. Similarly, in rats, adaptation has been shown to increase in magnitude along the ventral stream [9], attenuating the responses to predictable stimuli [33]. Finally, in primates, adaptation is at least partially preserved across object transformations (e.g., a smaller object can adapt the response to a larger object), as shown in both neurophysiological [34-36] and behavioral studies [37; 38]. Together, these effects could arguably counteract any increase in response persistence resulting from increased invariance, as shapes that are temporally stable (e.g., a translating or expanding object) could strongly and continuously adapt neuronal responses until a new, surprising stimulus (i.e., a novel object) enters the neurons' receptive fields (RFs). More broadly, the predictive coding framework posits that, within certain cortical circuits, only error or surprise signals are encoded, their temporal dynamics naturally depending on a top-down signal carrying the input prediction [39-41]. This leads to the alternative hypothesis that the timescale and persistence of neural responses do not increase across the cortical hierarchy, as each cortical area encodes surprising features of the response of the previous area.

Recently it has been shown that models of the ventral stream based on deep convolutional neural networks can be improved in their predictive power for perception and neural activity by including adaptation mechanisms [42] and recurrent processing [43-45]. Moreover, a progressive increase of the importance of intrinsic processing along the ventral stream may be expected, given that intrinsic temporal scales increase along various cortical hierarchies in primates and rodents [46-49]. However, it is not known how the interaction between invariant encoding of object information and intrinsic processing unfolds along the ventral stream, and whether or not the net result is an ordered progression of temporal scales of neural processing.

Here, we addressed this question by performing electrophysiological recordings across the rat analogue of the ventral stream, while exposing the animals to dynamic visual stimuli. We found that neural activity displays a hierarchy of temporal scales, as predicted for a feedforward object-processing pathway. At the same time, we found that intrinsic processing also becomes slower and more important in determining the overall temporal persistency of the representation, as one progresses from primary visual cortex (V1) towards higher-level areas. This indicates that the progressive increase in the stability of visual representations along a ventral-like hierarchy relies on both feedforward and intrinsic processing.

Results

To investigate the temporal structure of neural representations of dynamic visual scenes, we used 32-channel silicon probes to record from V1 of anesthetized rats, as well as from three distinct areas within the lateral extrastriate cortex – the lateromedial (LM), lerointermediate (LI) and laterolateral (LL) areas (see [11] for the anatomical layout of these areas). These target regions were chosen because, according to earlier findings by us and other groups [9-13; 33], they are organized in a functional hierarchy (V1→LM→LI→LL) that is the homologous of the primate ventral stream. During the recordings, the animals were kept under anesthesia and were passively exposed to

repeated presentations of nine 20s-long movies, which included six natural dynamic scenes (see example frames in Figure 1A), as well as three synthetic movies – namely, a random sequence of white noise patterns and the Fourier phase-scrambled versions of two of the natural movies (see Methods and supplementary videos). In addition, drifting bars, displayed over a grid of visual field locations, were used to map the receptive fields (RFs) of the units recorded along a probe [10; 11; 50]. This allowed identifying the area from which each unit was recorded by tracking the polarity of the retinotopic map, which, in rodents, reverses at the borders between adjacent visual areas [11; 12; 51-53]. Overall, a total of 294 well-isolated single units were recorded in 18 rats: 168 from V1, 20 from LM, 36 from LI, and 70 from LL.

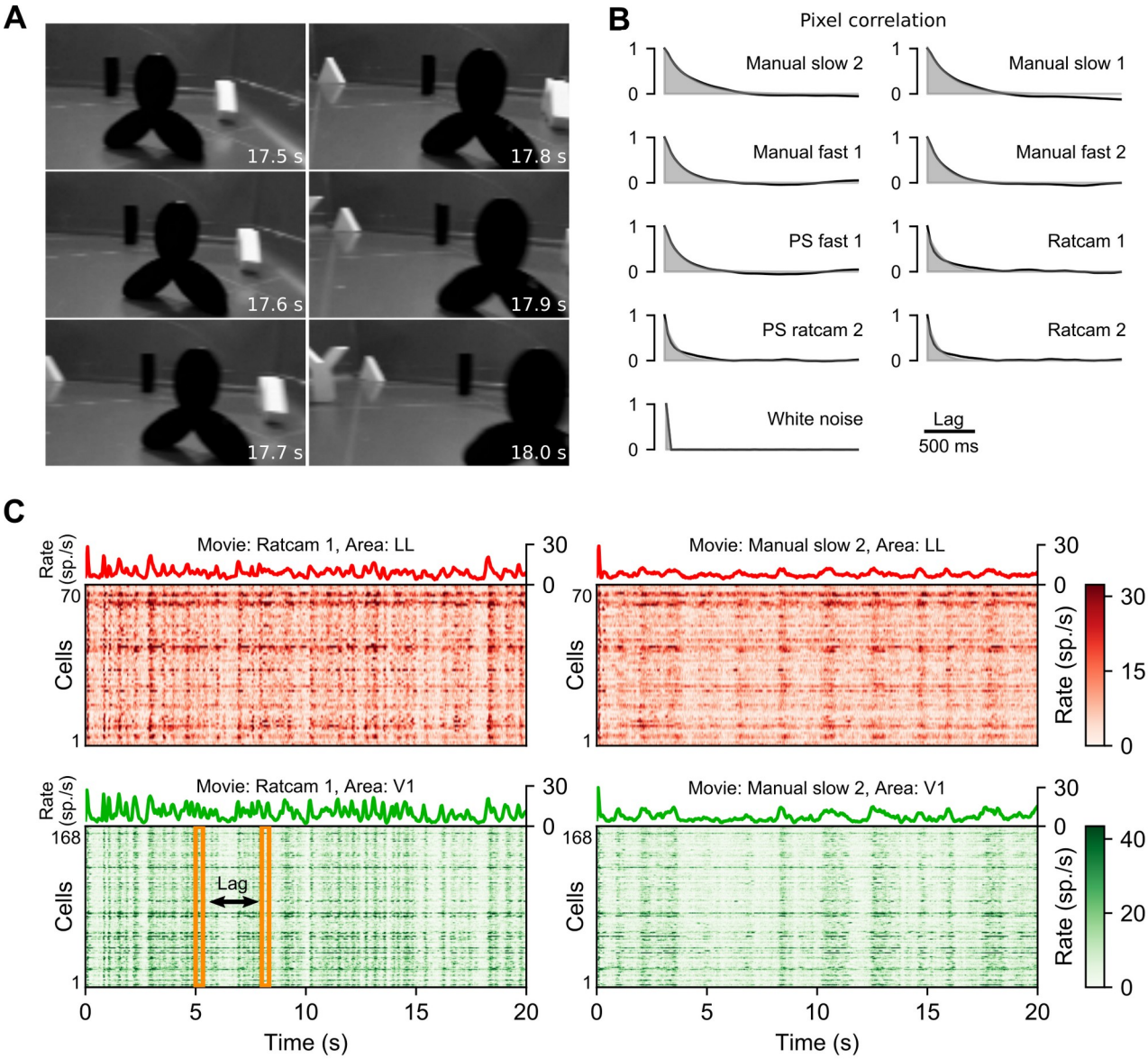

Figure 1. Visual stimuli and example neuronal recordings. (**A**) Six consecutive example frames from one of the movies used as visual stimuli in our study. The numbers on the bottom right of each panel give the time at which each frame appears. (**B**) Quantification of the characteristic timescale of the individual visual stimuli (movies). The *Manual* and *Ratcam* labels refer, respectively, to movies that were taken either by moving a hand-held camera across an arena filled with 3D-printed objects (some of them visible in A), or by installing a small camera over the head of a rat that was allowed to roam inside an arena containing the 3D-printed objects and another rat. The *PS* label refers to Fourier phase-scrambled versions of the corresponding movies (see Methods for details). In each panel, the black line shows the average correlation of pairs of image frames within a movie as a function of their temporal lag, while the shaded area shows the best exponential fit. (**C**) Examples of population responses to two of the movie stimuli. In the color matrices, each row color-codes the intensity of the trial-averaged response (computed in 33 ms time bins) of a neuron belonging to either the LL (red) or V1 (green) population. The color code is truncated to the 98th percentile of the firing rate distribution within each area for ease of display. The trace on top of each matrix shows the corresponding population-averaged response. The yellow frames illustrate the procedure to compute the characteristic timescale over which the population activity unfolded, in response to a given movie (see Figure 2).

Cortical representations unfold over timescales that are both stimulus- and area-dependent

As a first step in our analysis, we adapted an approach from [54] to characterize the temporal structure of the stimuli. Briefly, for each movie we computed the correlation coefficient between pixels belonging to image frames separated by a given lag. We then fitted the resulting dependence of the average correlation on the lag (Figure 1B, solid line) with a decaying exponential (shaded areas). The time constant of this exponential defined the timescale of the movie (see Methods for details). According to this analysis, the dynamics of our stimuli spanned a range of timescales, from ~30 ms for the white noise movie to ~600 ms for the slowest of the natural movies.

The activity in each population was stimulus-modulated, in a consistent way across the recorded neurons, for both fast and slow movies. This is shown in Figure 1C, where each line color-codes the response intensity of a LL (red) or V1 (green) unit to two example movies. As a result, the overall population-averaged activity (green and red traces) was also strongly stimulus modulated. To characterize how the temporal structure of the stimulus-locked responses depended on the visual input, we measured the population response timescale for each visual area and each movie by applying the same metrics defined above for the movie stimuli to the trial-averaged population response vectors [54]. That is, with reference to Figure 1C, we computed correlation coefficients between vertical slices (bins of spike counts) that were separated by a given lag (yellow frames), and then averaged the resulting coefficients to obtain a measure of signal correlation [55] as a function of lag (Figure 2A, solid lines). This curve was fitted with an exponentially damped oscillation or a simple exponential based on a systematic model selection procedure (dashed lines; see Methods for details), and the decay constant of the exponential envelope was taken as the timescale of the response modulation. Next, we regressed the characteristic timescale of the response against the characteristic timescale of the stimulus, separately for each area, using a linear model with a common slope across all areas and an area-dependent intercept (Figure 2B). The intercept is the baseline temporal timescale for stimulus-driven responses in each area. This baseline was clearly higher in the extrastriate areas (LM, LI, LL) than in V1, with smaller, statistically insignificant differences between LM, LI and LL ($p=0.4$ for both LM vs LI, and LI vs LL; two-tailed t-test, t=0.8,-0.8, df=31). Therefore, we repeated the regression analysis after lumping together these three areas (gray line). This revealed that the response timescales in all the areas depended strongly on the stimulus timescale (slope: 0.71±0.14, significantly different from 0 at p=1e-5, two-tailed t-test, t=5.2, df=33), and that the response timescale was significantly longer in extrastriate cortex than in V1 (intercept difference: 56±22ms; p=0.015, two-tailed t-test, t=2.6, df=33). Overall, these results show that the characteristic temporal scale of the population representation of dynamic stimuli depends on the temporal scale of the visual input, and that representations in extrastriate cortex unfold over longer temporal scales than in V1.

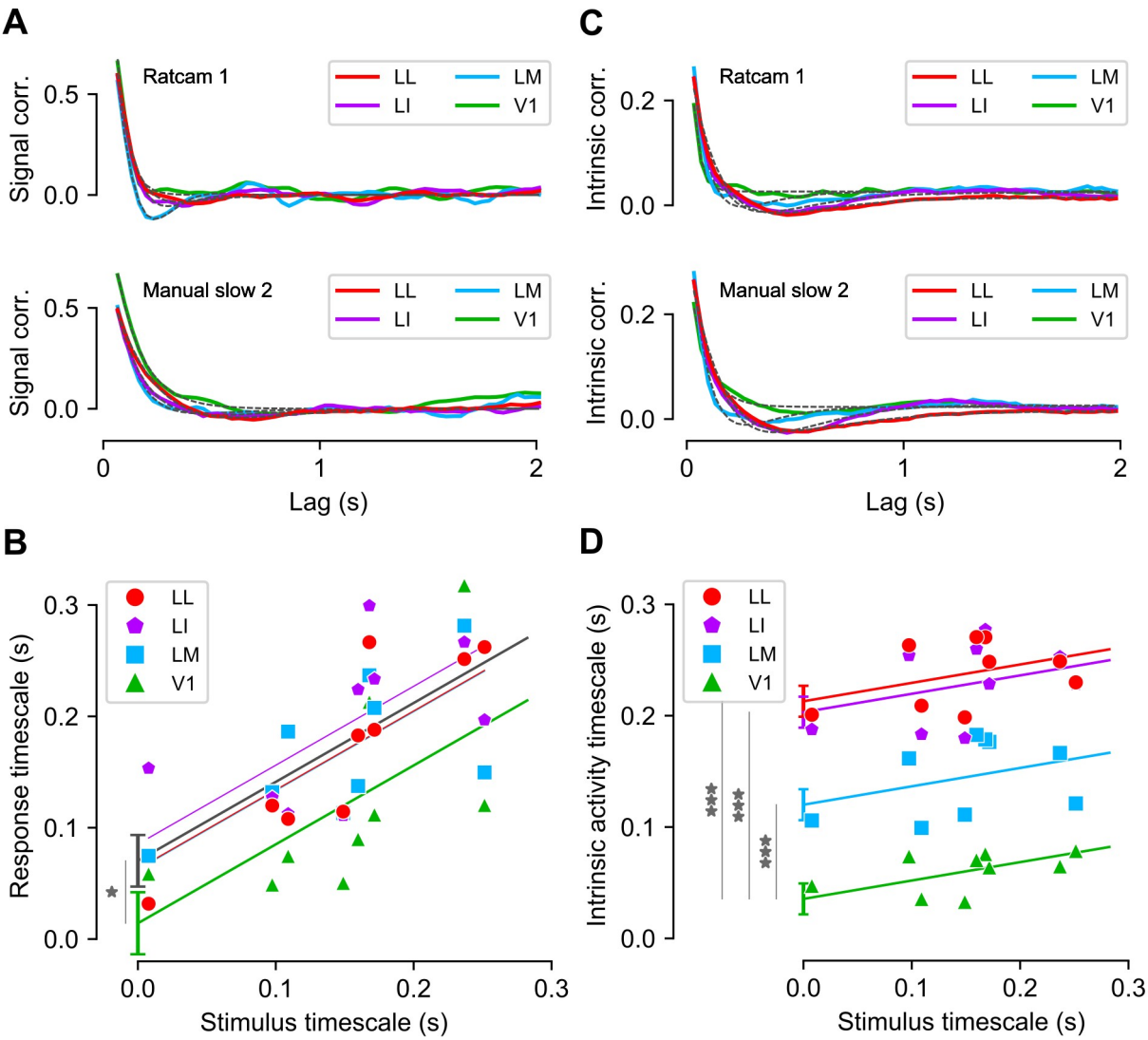

Figure 2. Characteristic timescale of the stimulus-driven population responses and of the intrinsic neuronal activity. (**A**) The correlation between pairs of population response vectors (as those highlighted by the yellow frames in Figure 1C) is plotted as a function of their temporal lag for two example movies (solid, colored lines). These curves were fitted with either an exponential decay or a damped oscillation function (fits are shown as black dashed lines). The time constants of these fits were taken as the characteristic timescales of the population responses. (**B**) The timescales of the population responses are plotted as a function of the timescales of the corresponding movie stimuli (colored markers). Each colored line is the linear fit prediction of the relationship between such timescales for a given area (same color code as in the key; note that the lines for LM and LL are partially overlapping). The gray line is the linear fit prediction obtained by pooling together the data of the three extrastriate areas (i.e., LM, LI and LL). The error bars show the standard errors of the intercept of the linear fits for V1 and the pooled extrastriate areas. (**C**) and (**D**) Same as A and B, but for the intrinsic timescales of neuronal activity (see main text for details).

The timescale of intrinsic processing increases along the ventral-like pathway

As illustrated in the previous section, our results suggest that the timescale of stimulus-locked, trial-averaged neural representations increases from V1 to extrastriate cortex. The temporal scale of intrinsic neural processing has also been suggested to increase across cortical hierarchies at the single cell [46] and at the population level [48]. We tested whether this applied to the rat ventral-like pathway by using the method described by [46], which is mathematically similar to the procedure used above to compute the timescale of the population response vectors, but considers the responses of a single cell across multiple trials, rather than the average responses of multiple neurons (see Methods). This allowed us to capture the stimulus-independent, within-trial temporal correlations in the spiking activity of a neuron, which were then averaged over all the units of a population. The time-dependence of the resulting correlation function (solid lines in Figure 2C) was fitted with an exponential or an exponentially damped oscillation, based on systematic model selection performed independently for each condition (black dashed lines; see Methods for details). The characteristic temporal scale was determined as the decay time constant of the exponential envelope of the fit. Finally, we linearly regressed the intrinsic timescale of neural activity with the timescale of the movie stimulus and the cortical area (Figure 2D), as we did above for the stimulus driven response (Figure 2B). The intercept values of the linear fit revealed a clear hierarchical organization of the timescale of intrinsic activity (V1: 36±14ms; LM: 120±14ms; LI: 203±14ms; LL: 213±14ms), with all three extrastriate areas being significantly different from V1 (p=5e-7, 1e-13, 2e-14 respectively for LM, LI, LL, two-tailed t-test, t=6.3, 12.6, 13.3, df=31). We also found a mild dependence on the timescale of the movie (slope: 0.17±0.07, p=0.02, two-tailed t-test, t=2.4, df=31), much weaker than that observed for the stimulus-driven response (compare Figures 2B and D). Interestingly, the range of intrinsic timescales we recorded in our experiment quantitatively matched that reported for sensory cortex in behaving monkeys by [46] and in behaving mice by [48]. Overall, these results show that the temporal scale of the intrinsic activity increases along rat ventral stream.

The temporal stability of neural representations increases along the ventral-like hierarchy

Temporal stability of a neuronal representation supports stable perception of the content of a visual input that unfolds smoothly over time. For instance, if an object (e.g., a triangle) sweeps across the visual field, a temporally stable representation should support discrimination of this object from other objects (e.g., squares) despite ongoing continuous changes in its appearance. Correlation between neural activity at different times does not by itself assess such stability of discrimination because population firing vectors can have a fixed amount of correlation and yet be discriminable to different degrees, depending on how they are organized with respect to a discrimination boundary. Moreover, a correlation measure assumes that deviations in the neural code are most important along directions that are orthogonal to the current population vector, as correlation coefficients behave like a dot product or a cosine similarity measure. This is not necessarily the right notion of similarity for discrimination problems. Thus, we sought a direct test of the temporal stability of discrimination based on neural population responses.

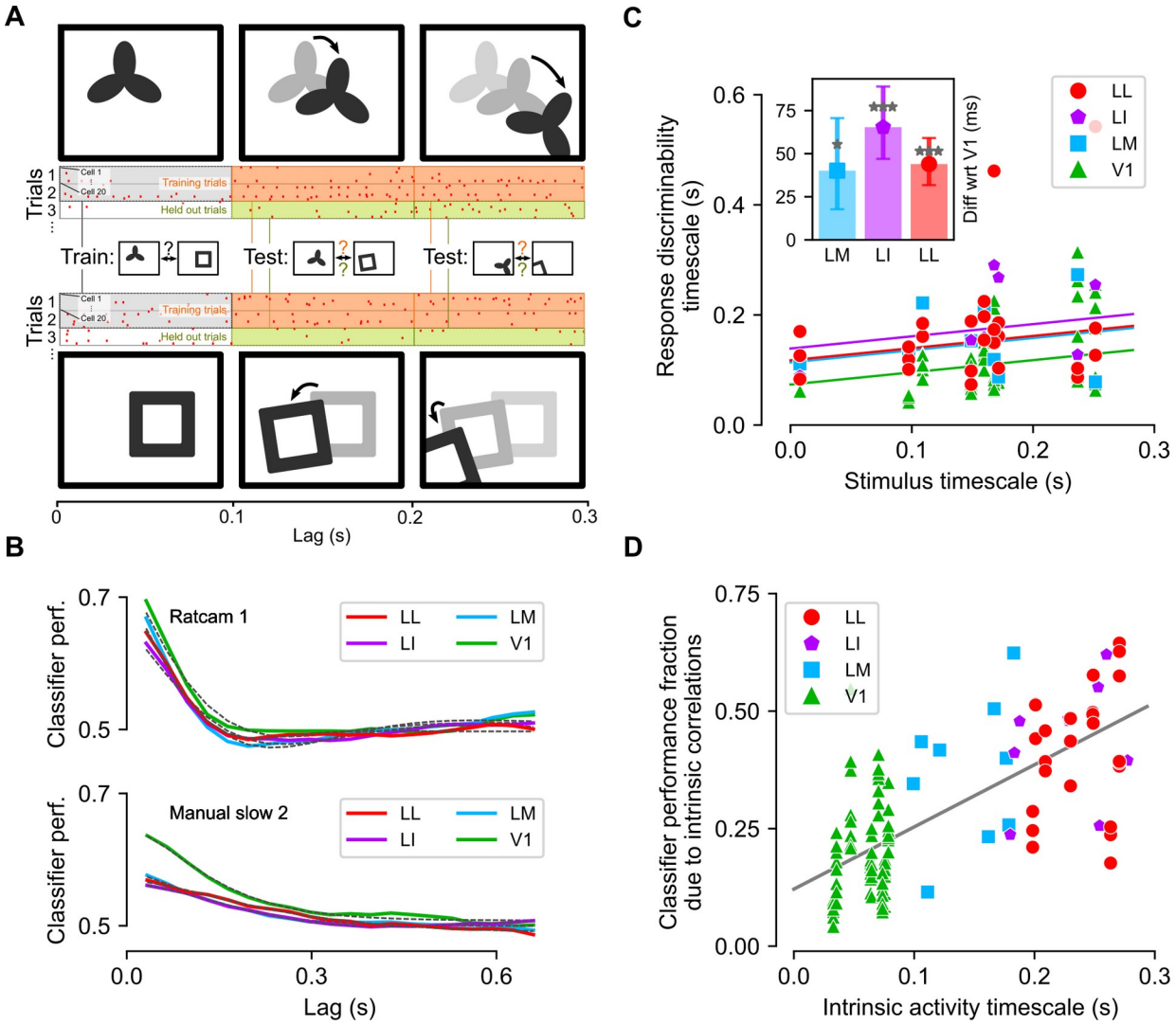

Figure 3. Temporal stability of neuronal representations, as assessed through decoding analysis. (**A**) Schematic of the classifier analysis, aimed at assessing the temporal stability of the population representation. A linear classifier is trained to distinguish the population activity recorded at two different time points of the movie (spike counts computed from the gray time bins) and tested at other time points that follow (or precede, not shown) the training points by a certain lag. The classifier is tested on a held-out set of trials (green boxes), as well as on the same experimental trials that were used for training (orange boxes). The cartoons of the movie frames shown on top or below each time bin are for illustration purpose only: they highlight the fact that, in general, two non-overlapping segments of a movie (from which the population activity is decoded) will contain different, transforming (e.g., translating, rotating, etc.) objects. For clarity, in this illustration, the size of the spike count windows (i.e., gray, orange and green boxes) was set to 100 ms, while 33 ms wide bins were used in the actual analysis. The spike patterns shown in the cartoon correspond to those actually fired by a randomly selected pseudopopulation of 20 units in V1 at two time points during the presentation of one of the movies (only three trials shown out of 30). (**B**) Classifier performance on held-out trials for two example movie stimuli and one example pseudopopulation per area (colored curves). The performance is plotted as a function of the lag between training bin (gray boxes in A) and test bin (green boxes in A), and is fitted with either an exponential decay or a damped oscillation function (fits are shown as black dashed lines). (**C**) The timescale of response discriminability, measured as the time constant of the exponential decay of the classifier performance on the held out trial set, is plotted as a function of the timescale of the movies. Each dot corresponds to a distinct pseudopopulation of 20 units. The solid lines are linear regressions with common slope and different intercept across the four areas. The inset shows the difference of the intercept for areas LM, LI and LL vs. area V1 (* $p < 0.05$, *** $p < 0.001$, one-tailed bootstrap test). (**D**) The fraction of classifier performance due to intrinsic correlations is plotted as a function of the timescale of such correlations (i.e., the intrinsic timescale of neuronal activity shown in Figure 2D). If p_{train} and p_{test} are, respectively, the performance of the classifier on the train set (orange boxes in A) and the held-out trial set (green boxes in A) at lag = 33ms (=1 time bin), the fraction of classifier performance due to intrinsic correlations is defined as $(p_{train}-p_{test})/p_{train}$.

To this end, we imagined a task where an observer learns to discriminate between the visual input appearing at a time t_A and that presented at a much later time t_B. If this task is performed on the basis of the activity in a given cortical area, we should be able to train a classifier to reliably discriminate population response vectors occurring at these two times in the movie. Now consider responses at a pair of time points shifted by a small lag Δt, namely $t_A+\Delta t$ and $t_B+\Delta t$. If the relevant part of the representation of the visual input is temporally stable over this lag, the trained classifier should perform as well on responses at the lagged time points as it did at the original time points (schematic in Figure 3A). Averaging over all t_A and t_B for each lag Δt will assay how well the population responses support temporally stable discrimination of visual inputs.

Following this strategy, we implemented a linear classifier on pseudopopulations of 20 randomly selected units (see Methods). Results for larger populations (50 units) from areas where these were available (V1 and LL) are reported in Supplementary Figure 1). The classifier was trained to distinguish population activities, represented as 20-dimensional spike count vectors (bin width 33ms), evoked by a movie at two different time points. Critically, only a subset of trials was used for training (gray boxes in Figure 3A). The temporal stability of the population code was then assessed by testing the performance of the decoder at time points shifted by a lag Δt (Figure 3A). In order to isolate the contribution of the direct drive from the visual stimulus to the overall temporal stability, we tested the classifier only on held-out experimental trials that were not used in training (green boxes). This ensured that intrinsic temporal correlations (Figure 2D) between lagged frames would not affect the similarity between activity vectors used in training and testing. This analysis was repeated for each population and each movie, and for all possible choices of pairs of training time points (details in Methods). The results were averaged over these pairs to measure the mean discrimination performance as a function of the lag (Figure 3B; colored lines), and then fitted with the same functional form as used above for the temporal correlations of the neural responses (black dashed lines). A linear regression revealed that the decoder's performance depended significantly on the timescale of the movie stimulus (Figure 3C; common slope: 0.22, 68%CI [0.15-0.32], nonzero at p=8e-4, one-tailed bootstrap test). In addition, the intercept of the fit was significantly larger in the three extrastriate areas than in V1 (V1 intercept: 74[55-82]ms; difference with respect to V1 for the other areas: LM 40[17-70]ms, nonnegative p=0.03; LI 65[48-89]ms, p<1e-4, LL 44[32-59]ms, p=2e-4, one-tailed bootstrap test). Such a difference indicates that neural representations in higher areas of the ventral hierarchy support percepts of dynamic stimuli that are stable over longer timescales as compared to V1 representations.

Intrinsic correlations extend the within-trial temporal stability of neural representations

During behavior, sensory perception unfolds continuously on the basis of the ongoing activity in neural populations. Intrinsic temporal correlations in this activity (Figure 2D) could stabilize the neural code against changes in the visual input, increasing the reliability with which objects moving across the visual field can be discriminated based on neuronal responses. To test for this possibility, we repeated the classifier analysis described above, now measuring the performance of the trained

classifier in the same trials that were used for training, but at lagged time points (orange boxes in Figure 3A; the resulting performance curves are shown in Supplementary Figure 2A). In this case, fluctuations in test response vectors at times $t_A+\Delta t$ and t_B will be correlated with fluctuations in the training vectors at t_A and $t_B+\Delta t$. To quantify the contribution of these intrinsic correlations to the stability of discrimination we focused on classifier performance at a lag of 33ms (one video frame), and defined the fraction of performance due to intrinsic correlation as follows:

$$f=\left(p_{train}-p_{test}\right)/p_{train}$$

Here p_{train} and p_{test} are the performance of the classifier at lag 33ms when evaluated on the training trials or on the held out trials, respectively. f increased along the hierarchy, revealing that intrinsic correlations play an increasingly important role in determining the overall temporal stability of the neural code along the cortical progression (Figure 3D. Slope of the fit: 1.2±0.1, nonzero *p=8e-19*, two-tailed t-test, t=10.6, df=115; intercept: 0.12±0.02, p=9e-10, two-tailed t-test, t=6.7, df=115). Overall, these findings show that the relative importance of intrinsic dynamics for population codes in visual cortex increases along a ventral-like, object-processing hierarchy.

Discussion

Most popular models of the ventral stream are purely feedforward, and view the sensory hierarchy as implementing bottom-up computations that are driven primarily by the visual input and that extract representations that are increasingly invariant to image transformations [2; 3; 8; 23; 56]. This picture of the ventral stream suggests that responses in deeper areas will be more invariant to differences between static images (transformations) of the same object, as observed in several experiments in both monkeys and rats [4; 5; 7; 11; 12; 57]. As a result, responses to dynamic movies should be more temporally stable in deeper areas. In other words, continuous changes that occur from frame to frame in a dynamic visual input should elicit smaller changes in the neural code of deeper areas compared to peripheral areas, resulting in progressively slower dynamics. On the other hand, the temporal stability of a code has been shown to be at odds with its information theoretical efficiency in encoding the past of the stimulus, as this involves temporal decorrelation [24; 58]. Moreover, adaptive and top-down mechanisms (e.g., prediction error signals) have been identified in visual cortex [27; 59] that favor the encoding of surprising or transient inputs over predictable or sustained ones [30; 33; 40; 60; 61]. This would enforce a common, similarly fast stimulus-driven dynamics throughout visual cortex.

In our study, we tested which of these scenarios is more consistent with the dynamics of neuronal representations recorded along the rodent analogue of the ventral stream, while anesthetized rats were subjected to dynamic visual stimulation with naturalistic and synthetic movies. We found that, in all tested visual areas, the temporal scale of visual encoding depends on the characteristic timescale of the dynamic stimuli, yet it significantly increases from V1 to extrastriate ventral cortex (Figures 2A,B and 3B,C). This implies that the increase of invariance afforded by feedforward computations along the rat ventral pathway is not completely washed out by adaptative, recurrent and top-down processes. As a result, when the pathway is probed with dynamic stimuli, a functional hierarchy of representational timescales is observed that parallels the hierarchical buildup of

invariance revealed by brief presentation of static stimuli [11; 12].

A second key question addressed by our study is the role of intrinsic processing in establishing the persistency of neuronal representations along an object-processing pathway. In both primates and rodents, intrinsic temporal scales have been found to increase along various cortical hierarchies [46-49]. However, it is unknown whether a similar increase takes place along the ventral stream and, if so, how such increase may combine with the growing stability of stimulus-driven responses we have discussed in the previous section. Recently, this topic has received a lot of attention in the computational neuroscience literature. In fact, deep convolutional neural networks have emerged as successful models of the ventral stream [8], and authors investigating the limitations of purely feedforward architectures within this family have proposed including temporal dynamics and adaptive mechanisms [42], or recurrent computations [43-45]. Indeed, it has been suggested that convolutional networks that excel in object recognition need to be very deep simply to approximate operations that could be implemented more efficiently by recurrent architectures [44; 62; 63]. These theoretical developments point to the potential importance of intrinsic dynamics for the cortical representation of visual stimuli. Experimentally, however, most work has focused on static stimuli, and thus it is not known whether intrinsic dynamics contribute to the structure of population codes for stimuli that are themselves varying in time. Our experiments addressed this question. We found strong evidence that the temporal scale of intrinsic processing increases along rat lateral extrastriate areas (Figures 2C,D), and that such increase plays an incrementally important role in stabilizing the neural representation of the visual input along the cortical hierarchy (Figure 3D).

To achieve precise control and repeatability of stimuli falling within each neuron's receptive field over an extended period (1.5/2 hours) we performed acute recordings under anesthesia. We expect, in view of previous work in anesthetized animals [54; 64], that the spatiotemporal characteristics of population activity that we measured would be similar in awake animals. Indeed, our estimates of single-cell intrinsic processing timescales fall squarely within the ranges reported for behaving mice and monkeys [46; 48; 61]. On the other hand, our preparation may have suppressed top-down effects that could appear in awake animals. Such effects would add to the intrinsic processing, whose importance, relative to the feedforward drive, increases in higher areas of the ventral stream, as already discussed.

We tested the temporal stability of discrimination based on neural population responses by first training a classifier to distinguish between responses at times t_A and t_B (which correspond to different visual inputs) and then testing the classifier's ability to discriminate responses at the lagged times $t_A+\Delta t$ and $t_B+\Delta t$. Discrimination performance decayed with the lag, but the decay was slower in higher visual areas, indicating greater stability of the visual representation (Figure 3B,C). Interestingly, however, the *absolute* discrimination performance was highest in V1, the lowest visual area, at least at short lags (Figure 3B). This may seem surprising in light of popular conceptual frameworks advocating for "untangling" [2] and "straightening" [65] of neural representations along the ventral stream. These frameworks suggest that deeper areas, carrying untangled representations of objects, should support better discrimination by a linear classifier, as demonstrated in several monkey studies [4; 5; 7; 57]. However, in rats, lower-order areas along the hierarchy have been shown to encode prominent low-level features (such as luminance and contrast) [11; 66]. This leads

to higher absolute discriminability for scenes where, as in our stimuli, such features are not matched [11]. Our analyses accounted for this by quantifying temporal scales in terms of relative, rather than absolute, magnitude of correlations and decoding performance as a function of the lag (Figures 2 and 3). Moreover, when we did analyze absolute decoding performance, we found that a larger proportion of within-trial temporal stability is due to intrinsic processing at the top of the hierarchy (area LL) than at the bottom (area V1) (Figure 3D).

What roles could these intrinsic dynamics play in behavior? Intrinsic processing may be crucial for perception in a noisy, changing environment. For example, maintenance of sensory information by stimulus-independent temporal correlations in population activity can lead to better behavioral performance, when consistent estimates of a quantity are needed [48]. Alternatively, intrinsic processing may support predictive coding [39; 41], allowing neural circuits to use feedforward inputs to predict and represent what will happen next, an ability with obvious utility for behavior. In order to play these roles effectively, the dynamics of intrinsic processing should be adapted to the temporal structures that are typically encountered in natural environments. Previous work has shown (or at least suggested) that many aspects of spatial and temporal processing in the early visual system are indeed adapted to the structure of natural scenes. Examples are histogram equalization [67] and On-Off circuit asymmetries in the retina [68], development of V1 simple and complex cells [69-73], texture perception [74], as well as eye movements [75]. The idea that intrinsic processing could be similarly adapted deep into the cortical hierarchy could be causally tested by altering the animals' visual environment during development or during a learning task, and then measuring neural population activity across the cortical hierarchy with ethologically relevant, dynamic stimuli.

Methods

Animal preparation and surgery

All animal procedures were in agreement with international and institutional standards for the care and use of animals in research and were approved by the Italian Ministry of Health: project N. DGSAF 22791-A, submitted on Sep. 7, 2015 and approved on Dec. 10, 2015 (approval N. 1254/ 2015-PR). 18 male Long Evans rats (Charles River Laboratories) with age 3-12 months and weight 300-700 grams were anesthetized with an intraperitoneal (IP) injection of a solution of 0.3 mg/kg of fentanyl (Fentanest®, Pfizer) and 0.3 mg/kg of medetomidine (Domitor®, Orion Pharma). Body temperature was kept constant at ~37° with a warming pad, and a constant flow of oxygen was delivered to the rat to prevent hypoxia. The level of anesthesia was monitored by checking the absence of tail, ear and hind paw reflex, as well as monitoring blood oxygenation, heart and respiratory rate through a pulse oximeter (Pulsesense-VET, Nonin), whose sensor was attached to one of the hind paws. After induction, the rat was placed in 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 over the left and posterior parietal bones, a craniotomy was performed over the target area in the left hemisphere (typically, a 2x2 mm^2 window). Dura was also removed to ease the insertion of the electrode array. Stereotaxic coordinates for V1 recordings ranged from -7.49 to -8.36 mm anteroposterior (AP), with reference to bregma; for extrastriate areas

(LM, LI and LL), they ranged from -6.64 to -7.74 mm AP.

Once the surgical procedure was completed, the stereotaxic apparatus was moved on top of an elevated rotating platform. The right eye was immobilized with an eye ring anchored to the stereotaxic apparatus, and the left one was covered with opaque tape. The platform was then rotated, so as to align the right eye with the center of the stimulus display and bring the binocular portion of its visual field to cover the left side of the display. Throughout the experiment, ophthalmic solution Epigel (Ceva Vetem) was regularly applied onto the right eye and the exposed brain surface was covered with saline to prevent drying. In addition, a constant level of anesthesia was maintained through continuous IP infusion of the same anesthetic 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-500; New Era Pump Systems).

Neuronal recordings

Extracellular recordings were performed with 32-channel silicon probes (Neuronexus Technologies), following the same procedure used in [11]. Briefly, to maximize the receptive field coverage, recordings in V1 were performed with 4- or 8-shanks probes, which were inserted perpendicularly into the cortex. Recordings from lateral extrastriate areas were performed using single-shank probes, which were inserted diagonally into the cortex with an angle of ~30°, in order to map the progression of receptive fields's centers along the probe and track the reversal of retinotopy between adjacent areas (see [10; 11]). The space between recording sites on each shank ranged from 25 to 200μm; the distance between shanks (when more than one) was 200μm; the surface of recording sites was either 177 or 413 μm^2. Extracellular signal was acquired with a TDT system 3 workstation (Tucker-Davis Technologies) at a sampling rate of ~24kHz.

As mentioned above, 18 rats were used for this study. Since, in some occasions, multiple recording sessions were performed from the same animal, this yielded a total of 22 sessions. In some of these sessions, multiple areas were targeted (e.g., neurons could be recorded simultaneously from LI and LL). Overall, we performed neuronal recordings from: 1) V1 of 5 different rats, for a total of 7 recording sessions, yielding 168 units; 2) LM of 4 rats (4 sessions), yielding 20 units; 3) LI of 8 rats (8 sessions), yielding 36 units; and 4) LL of 11 rats (12 sessions), yielding 70 units.

Visual stimuli

Stimuli were presented full-field to the right eye at a distance of 30 cm on a 47-inch LCD monitor (SHARP PN-E471R), with 1920 × 1080 resolution, 60Hz refresh rate, 9ms response time, 700cd/m^2 maximum brightness, 1200:1 contrast ratio, spanning a field of view of ~120° azimuth and ~89° elevation. The stimuli were presented using Psychtoolbox [76] in MATLAB in a pseudorandom order. Between each stimulus (movie or drifting bar) a black screen was shown for at least 200 ms.

The receptive fields of the neurons recorded along a probe were estimated by showing drifting oriented bars over a grid of 73 locations on the stimulus display, covering the central 110° azimuth span and central-upper 70° elevation span of the total field of view of the monitor. The bars were 10º long and drifted 7 times along four different directions (0°, 45°, 90°, and 135°). A spherical correction (as described in [53]) was applied to each bar to compensate for shape distortions at large

eccentricities. Mapping the inversion of the retinotopic map at each area boundary allowed identifying the visual area each neuron was recorded from [10-12].

The main stimulus set consisted of 9 movies: two fast and two slow *manual* movies, two *ratcam* movies, a phase-scrambled version of one of the fast movies and a phase-scrambled version of one of the ratcam movies (see next paragraph for details). The resolution of these movies was 720 × 1280 pixels and they were presented for 20 seconds at a rate of 30 frames per second (fps). An additional white-noise movie with a resolution of 45 × 80 pixels was shown at 20 fps for 20 seconds. All movies were converted to grayscale and were gamma-corrected offline with a lookup table calculated for the monitor used for stimulus presentation. During the course of the experiment, each movie was presented 30 times. The videos can be found in the Supplementary Information.

The manual movies were designed to reproduce the continuous flow of visual information typically experienced by an observer that explores an environment containing a number of distinct visual objects, placed in different locations. To this aim, we 3D-printed various geometrical objects, we painted them black or white (see examples in Figure 1A), we placed them inside an arena and we simulated an observer roaming through such an environment by smoothly moving a hand-held camera across the arena for 20 s. The camera was moved at two different speeds to obtain slower and faster movies. The *ratcam* movies were intended to better simulate the visual input of a rat exploring a natural environment [64]. They were obtained by placing a small web camera (Microsoft Lifecam Cinema HD) on the head of a rat, while it was running freely inside the arena that contained some of the 3D-printed objects and another rat.

The phase-scrambled movies were obtained by performing a spatiotemporal fast Fourier transform (FFT) over the standardized 3D array of pixel values, obtained by stacking the consecutive frames of a movie. The phase spectrum was extracted and shuffled, then merged with the unaltered amplitude in such a way to preserve the conjugate symmetry of the initial transform. An inverse FFT was then applied on the resulting array. The imaginary parts were discarded, then the original mean and standard deviation were restored. Values outside the 0-255 range were clipped in order to obtain a valid movie. The white noise movie was generated by randomly setting each pixel in the image plane either to white or black.

Data analysis

Spike sorting and selection of the units included in the analyses

Data were spike-sorted offline with the spike sorting package KlustaKwik-Phy [77] in two steps: automated spike detection, feature extraction and expectation maximization (EM) clustering were followed by manual refinement of the sorting using a customized version of the GUI. The last step was performed by taking into account many features of the candidate clusters: 1) 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); 2) the shape of the auto- and cross-correlograms (important to decide whether to merge two clusters or not); 3) the variation, over time, of the principal component coefficients of the waveform (important to detect and take into account possible electrode drifts); and 4) the shape of the average waveform (to exclude, as artifacts, clearly

non-physiological signals). Clusters suspected to contain a mixture of one or more units were separated using the "reclustering" feature of the GUI (i.e., by rerunning the EM clustering algorithm on the spikes of these clusters only).

At the end of the manual refinement step, we further screened the resulting well-isolated single units to make sure that their firing was reproducible across repeated stimulus presentations. The selection was based on a reproducibility index that quantified, for each neuron, how reliable its responses were to its preferred frames within a movie (see definition below). This metric was chosen over other approaches that rely on computing the correlation coefficient between responses to repetitions of the same stimulus [13; 78], because we wanted to avoid the inclusion of silent neurons that would have yielded high correlations due to lack of activity. To calculate the reproducibility index for a given unit and movie, we included in the metric the top 10% time bins with the highest neuronal response (given by the firing rate of the neuron in the bins). Let $X_t = \{x_k^{(t)}\}$ be the set of responses of the neuron $x_k^{(t)}$ in one of such preferred time bins t (with k denoting the stimulus repetition). Let N be the number of repetitions. Then, the reproducibility index is calculated as $1 - \left\langle \frac{\sigma[X_t]}{\langle X_t \rangle} \right\rangle / \sqrt{N}$. The resulting metric ranges from 0 to 1 [79], where 1 corresponds to ideal responses with perfectly reproducible trials. A cell was considered reproducible if, for at least one of the nine movies, the reproducibility index was 0.7. This threshold was arbitrarily set following an extensive visual inspection of raster plots from different movies to ensure that clearly responsive and reproducible cells were included. For all the analyses described in this study, the responses of these units during the presentation of the movie stimuli were converted to a spike-count representation by binning spike trains in 33ms bins.

Characteristic timescale of the movie stimuli

To quantify how fast the pixel-level content of the movies changed over time, we computed the following metric. Given a movie, for a given time lag Δ, we computed the average correlation coefficient between all movie frame pairs separated by that lag, on a pixel-wise basis, i.e.:

$$C(\Delta) = \frac{1}{T-\Delta} \sum_{t=1}^{T-\Delta} \frac{\mathrm{Cov}\,[X_t, X_{t+\Delta}]}{\mathrm{Var}\,[X_t]\,\mathrm{Var}\,[X_{t+\Delta}]} \qquad \text{(Equation 1)}$$

where T is the total number of frames in the movie, $X_t = \{x_{w,h}^{(t)}\}$ is the movie frame t and expectations are taken over pixel positions, indexed by w, h (for width and height). We then plotted the correlation $C(\Delta)$ as a function of Δ (Figure 1B). To fit the decay of the correlation with Δ we considered two possible functional forms: a decaying exponential of the form $y(t) = a \exp(-t/\tau)$ and a damped oscillation function of the form $y(t) = a \exp(-t/\tau) \cos(\omega t + \phi)$. Fitting was performed with the basin-hopping algorithm [80] coupled with L-BFGS-B [81]. Model selection was performed independently for each fit with an Extra Sum of Squares test [82], choosing the more complex model whenever the p value from the test resulted below the threshold of 0.05. For the stimulus timescale, the model selection procedure selected the simple exponential form in all cases. The time constant τ of the exponential decay was taken as the characteristic timescale of the stimulus.

Characteristic timescale of stimulus-driven neuronal population responses

To measure the characteristic timescale of neural stimulus correlations from population peri-stimulus time histograms (PSTHs), we applied a very similar procedure to that described above for the visual input. We used the same expression as above (Equation 1) to compute $C(\Delta)$, but now we set $X_t = \left\{ x_n^{(t)} - \langle x_n \rangle \right\}$, where $x_n^{(t)}$ is the trial-averaged spike count of cell number n at time frame t, and subtracting $\langle x_n \rangle$ centers $x_n^{(t)}$ around its temporal average ($x_n^{(t)}$ as a function of t is color-coded in Figure 1C for the V1 and LL populations, in response to two example movies). A time constant was extracted by fitting either an exponential decay of the form $y(t) = a \exp(-t/\tau)$ or a damped oscillation function of the form $y(t) = a \exp(-t/\tau) \cos(\omega t + \phi) + b$ for $t > 0$ (see Figure 2A). Model selection was carried out as described in the previous section. The time constant τ of the exponential decay or of the exponential envelope of the damped oscillation was taken as the characteristic timescale of the trial-averaged population response (PSTH). We considered one population PSTH for each cortical area we recorded from, pooling together all the available sessions, as illustrated in Figure 1C.

Intrinsic timescale of neuronal activity

Intrinsic timescales of neuronal activity were computed in a way that mirrors the stimulus-driven correlation timescales defined in the previous section and that is a simple extension of the definition given by [46]. To compute $C(\Delta)$ for a given cell, we set $X_t = \{ x_k^{(t)} \}$ in Equation 1, where $x_k^{(t)}$ is the spike count of the cell at time bin t in trial k. Following [46], we then averaged $C(\Delta)$ across all units belonging to the same area before extracting a time constant (see example in Figure 2C). The characteristic timescale of intrinsic correlation decay was computed as above for the PSTH, using the same fitting functions and the same model selection procedures.

Characteristic timescale of decoding performance

After computing the average classifier performance $p(\Delta)$ as a function of the lag Δ (see next section), a characteristic decay time constant was extracted by the same procedure previously used to compute the timescales of the movies, of the stimulus-driven responses and of the intrinsic activity. Namely, either an exponential decay of the form $y(t) = a exp(-t/\tau) + b$ or a damped oscillation function of the form $y(t) = a exp(-t/\tau) \cos(\omega t + \phi) + b$ were fitted to $p(\Delta)$ for $\Delta > 0$ (see example in Figure 3B). Model selection was performed as described above, and the exponential decay time constant τ was extracted as the quantity of interest.

Classifier analysis

To assess the discriminability of population activity at different points in time during the presentation of a given movie, we proceeded as follows. For each recording session, we divided the 30 available recording trials in a training set of 20 trials and a validation set of 10 trials. We pooled all units that were recorded from a given area across all sessions to obtain a total number of units N per area. We generated pseudopopulations of $K < N$ units (with $K = 20$ for most of the analyses) by selecting as many random non-overlapping subpopulations of size K as possible. For instance, in V1

K=8 as N=168. These subpopulations were the same in the training and in the validation set. Following standard procedure when working with pseudopopulations, for each trial set (training and validation) we shuffled cell activity across trials to destroy cross-cell noise correlations.

We then considered all pairs of time bins along a trial that were separated by at least 40 time bins (i.e., by 1320 ms, as bin size was 33 ms). For each of these "reference" pairs of time bins (gray boxes in Figure 3A), we trained a linear support vector classifier (provided by liblinear [83] via scikit-learn [84]) to discriminate population activity samples from one of the element of the pair vs. the other. The penalty hyperparameter was chosen by 3-fold cross-validation within the training set, performing a grid search over candidate values $10^{-2,-1,0,1,2}$. After selecting the best value of the hyperparameter, the classifier was re-trained on the full training set.

To assess the temporal stability of the population activity, the trained classifier was then tested on samples of population activity coming from different time bins than those it was trained on (orange and green boxes in Figure 3A). More specifically, if a classifier was trained on the pair of time bins at times t_1 and t_2, it was then tested on pairs of time point at times $t_1+\Delta$ and $t_2+\Delta$, where Δ took on values {-20,...,-1,0,1,...,20} bins * 33 ms/bin. The fraction of correctly decoded trials at negative and positive time increments was then averaged to yield a classifier performance curve $p(\Delta)$ with Δ ranging from 0 to 20 bins*33 ms/bin = 660 ms. This performance curve was computed for all possible reference pairs of time bins, and averaged over all pairs. The resulting average performance decay curve (for each individual movie stimulus and pseudopopulation) was used to compute the typical timescale of self-similarity for population activity as described above. Examples of such performance decay curves are given in Figure 3B for the case in which the test was performed on held-out trials (i.e., green boxes in Figure 3A).

This analysis was performed separately on the training and the held-out trial sets (respectively orange and green boxes in Figure 3A), while keeping the trained decoders fixed (see performance decay curves in Figure 3B and Supplementary Figure 2A). When computed on the training set, p(0) corresponds to the performance of the classifier on its own training data, but p(33ms) is the performance of the classifier on an entirely new set of data, which happens to come from the same set of experimental trials as the training data. When compared to the results on the held-out set, this allowed quantifying the extent to which the self-similarity of the population activity in time was due to the stimulus-locked representational structure of the population code rather than to intrinsic, within-trial, temporal correlations in the activity of each neuron. Intuitively, population activity at time t in trial i can be expected to be more similar to population activity at time $t+\Delta$ in the same trial than in a different trial, precisely because of the existence of within-trial, stimulus-independent correlations in the activity of each neuron. Comparing the performance of the classifier on the training set vs. the held-out validation set, as done in Figure 3D, allowed therefore to assess the importance of this correlational structure in enhancing the self-similarity of population activity over time.

Regression analysis and hypothesis testing

Linear regression for the dependence of the timescale of stimulus-driven responses and intrinsic activity on the stimulus timescale (Figure 2B,D), as well as for the dependence of the fraction of

classifier performance on the timescale of intrinsic activity (Figure 3D and Supplementary Figure 1C) was computed by ordinary least squares. In each case, residuals were not incompatible with a normal distribution (Jarque-Bera, $p > 0.05$). Regression and tests were performed with *statsmodels* [85].

The distribution of the data was less regular for the case of the classifier performance vs. the timescale of the movie (Figure 3C and Supplementary Figure 2B). In particular, there were a few outliers with very long decoding timescales, especially for LL, which could have biased our conclusion in favor of the hypothesis that higher areas in the hierarchy have longer processing timescales. To prevent our conclusions from relying disproportionately on these points, we performed the corresponding linear regressions using a robust estimator (Theil-Sen estimator; [86]). The confidence intervals were determined by percentile bootstrap (10000 samples; [87]) stratified by cortical area. This combination of estimator and CI determination procedure has been shown to provide reasonable coverage probability in the face of model misspecification and data contamination by outliers [86]. Following [87], p-values were computed by subtracting the bootstrap distribution of any given parameter from its estimated value to obtain a surrogate for the null distribution. The p-value was then obtained as the mass of the surrogate distribution for values larger than the estimate. This yielded a one-tailed test where the null hypothesis was that the parameter value was not larger than zero. Regression and tests were performed with custom R code [88] using the boot [89] and WRS [86] libraries.

Data and code availability

The datasets and code generated during this study have been made available confidentially for peer review. They will be made public and permanently reachable upon publication of the paper. The movie stimuli can be found in the Supplementary Information.

References

[1] DiCarlo, J. J. and Cox, D. D. (2007). Untangling invariant object recognition. Trends in cognitive sciences 11, 333-341.

[2] DiCarlo, J. J.; Zoccolan, D. and Rust, N. C. (2012). How does the brain solve visual object recognition? Neuron 73, 415-434.

[3] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature neuroscience 2, 1019-1025.

[4] Hong, H.; Yamins, D. L. K.; Majaj, N. J. and DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. Nature neuroscience 19, 613-622.

[5] Li, N.; Cox, D. D.; Zoccolan, D. and DiCarlo, J. J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? Journal of neurophysiology 102, 360-376.

[6] Pagan, M.; Urban, L. S.; Wohl, M. P. and Rust, N. C. (2013). Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. Nature neuroscience 16, 1132-1139.

[7] Rust, N. C. and Dicarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. The Journal of neuroscience : the official journal of the Society for Neuroscience 30, 12978-12995.

[8] Yamins, D. L. K.; Hong, H.; Cadieu, C. F.; Solomon, E. A.; Seibert, D. and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences of the United States of America 111, 8619-8624.

[9] Kaliukhovich, D. A. and Op de Beeck, H. (2018). Hierarchical stimulus processing in rodent primary and lateral visual cortex as assessed through neuronal selectivity and repetition suppression. Journal of neurophysiology 120, 926-941.

[10] Matteucci, G.; Bellacosa Marotti, R.; Riggi, M.; Rosselli, F. B. and Zoccolan, D. (2019). Nonlinear Processing of Shape Information in Rat Lateral Extrastriate Cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 39, 1649-1670.

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

[12] Vermaercke, B.; Gerich, F. J.; Ytebrouck, E.; Arckens, L.; Op de Beeck, H. P. and Van den Bergh, G. (2014). Functional specialization in rat occipital and temporal visual cortex Journal of neurophysiology 112, 1963-1983.

[13] Vinken, K.; Van den Bergh, G.; Vermaercke, B. and Op de Beeck, H. P. (2016). Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to

Temporal Cortex. Cerebral cortex (New York, N.Y. : 1991) 26, 3310-3322.

[14] Becker, S. and Hinton, G. E. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms Nature 355, 161-163.

[15] Berkes, P. and Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties Journal of Vision 5, 9.

[16] Cadieu, C. F. and Olshausen, B. A. (2011). Learning Intermediate-Level Representations of Form and Motion from Natural Movies Neural Computation 24, 827-866.

[17] Einhäuser, W.; Kayser, C.; König, P. and Körding, K. P. (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. The European journal of neuroscience 15, 475-486.

[18] Földiák, P. (1991). Learning Invariance from Transformation Sequences Neural Computation 3, 194-200.

[19] Körding, K. P.; Kayser, C.; Einhäuser, W. and König, P. (2004). How Are Complex Cell Properties Adapted to the Statistics of Natural Stimuli? Journal of Neurophysiology 91, 206-212.

[20] Wallis, G. (1996). Using Spatio-temporal Correlations to Learn Invariant Object Recognition. Neural networks : the official journal of the International Neural Network Society 9, 1513-1519.

[21] Wallis, G. and Rolls, E. T. (1997). IInvariant face and object recognition in the visual system Progress in Neurobiology 51, 167-194.

[22] Wiskott, L. and Sejnowski, T. J. (2002). Slow Feature Analysis: Unsupervised Learning of Invariances Neural Computation 14, 715-770.

[23] Wyss, R.; König, P. and Verschure, P. F. M. J. (2006). A Model of the Ventral Visual System Based on Temporal Stability and Local Memory PLOS Biology 4, e120+.

[24] Dan, Y.; Atick, J. J. and Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. The Journal of neuroscience : the official journal of the Society for Neuroscience 16, 3351-3362.

[25] Grill-Spector, K.; Henson, R. and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends in cognitive sciences 10, 14-23.

[26] Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. Journal of neurophysiology 97, 3155-3164.

[27] Webster, M. A. (2015). Visual Adaptation. Annual review of vision science 1, 547-567.

[28] Kaliukhovich, D. A.; De Baene, W. and Vogels, R. (2013). Effect of adaptation on object representation accuracy in macaque inferior temporal cortex. Journal of cognitive neuroscience 25, 777-789.

[29] Fritsche, M.; Lawrence, S. J. D. and de Lange, F. P. (2019). Temporal tuning of repetition suppression across the visual cortex biorxiv , .

[30] Stigliani, A.; Jeska, B. and Grill-Spector, K. (2019). Differential sustained and transient temporal processing across visual streams. PLoS computational biology 15, e1007011.

[31] Zhou, J.; Benson, N. C.; Kay, K. and Winawer, J. (2017). Unifying Temporal Phenomena in Human Visual Cortex biorxiv , .

[32] Stigliani, A.; Jeska, B. and Grill-Spector, K. (2017). Encoding model of temporal processing in human visual cortex. Proceedings of the National Academy of Sciences of the United States of America 114, E11047-E11056.

[33] Vinken, K.; Vogels, R. and Op de Beeck, H. (2017). Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement. Current biology : CB 27, 914-919.

[34] Andrews, T. J. and Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. NeuroImage 23, 905-913.

[35] De Baene, W. and Vogels, R. (2010). Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. Cerebral cortex (New York, N.Y. : 1991) 20, 2145-2165.

[36] Lueschow, A.; Miller, E. K. and Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. Cerebral cortex (New York, N.Y. : 1991) 4, 523-531.

[37] Afraz, A. and Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. Journal of vision 9, 10.1-1017.

[38] Afraz, S.-R. and Cavanagh, P. (2008). Retinotopy of the face aftereffect. Vision research 48, 42-54.

[39] Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience 2, 79-87.

[40] Issa, E. B.; Cadieu, C. F. and DiCarlo, J. J. (2018). Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. eLife 7, .

[41] Keller, G. B. and Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. Neuron 100, 424-435.

[42] Vinken, K.; Boix Xavier and Kreiman, G. (2019). Incorporating adaptation in object recognition models captures temporal dynamics in neurophysiology and perception biorxiv , .

[43] Tang, H.; Schrimpf, M.; Lotter, W.; Moerman, C.; Paredes, A.; Ortega Caro, J.; Hardesty, W.; Cox, D. and Kreiman, G. (2018). Recurrent computations for visual pattern completion. Proceedings of the National Academy of Sciences of the United States of America 115, 8835-8840.

[44] Kar, K.; Kubilius, J.; Schmidt, K.; Issa, E. B. and DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nature neuroscience 22, 974-983.

[45] Kietzmann, T. C.; Spoerer, C. J.; Sörensen, L. K. A.; Cichy, R. M.; Hauk, O. and Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. Proceedings of the National Academy of Sciences of the United States of America , .

[46] Murray, J. D.; Bernacchia, A.; Freedman, D. J.; Romo, R.; Wallis, J. D.; Cai, X.; Padoa-Schioppa, C.; Pasternak, T.; Seo, H.; Lee, D. and Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. Nature neuroscience 17, 1661-1663.

[47] Chaudhuri, R.; Knoblauch, K.; Gariel, M.-A.; Kennedy, H. and Wang, X.-J. (2015). A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. Neuron 88, 419-431.

[48] Runyan, C. A.; Piasini, E.; Panzeri, S. and Harvey, C. D. (2017). Distinct timescales of population coding across cortex. Nature 548, 92-96.

[49] Himberger, K. D.; Chien, H.-Y. and Honey, C. J. (2018). Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience 389, 161-174.

[50] Niell, C. M. and Stryker, M. P. (2008). Highly selective receptive fields in mouse visual cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 28, 7520-7536.

[51] Wang, Q. and Burkhalter, A. (2007). Area map of mouse visual cortex. The Journal of comparative neurology 502, 339-357.

[52] Andermann, M. L.; Kerlin, A. M.; Roumis, D. K.; Glickfeld, L. L. and Reid, R. C. (2011). Functional specialization of mouse higher visual cortical areas. Neuron 72, 1025-1039.

[53] Marshel, J. H.; Garrett, M. E.; Nauhaus, I. and Callaway, E. M. (2011). Functional specialization of seven mouse visual cortical areas. Neuron 72, 1040-1054.

[54] Fiser, J.; Chiu, C. and Weliky, M. (2004). Small modulation of ongoing cortical dynamics by sensory input during natural vision. Nature 431, 573-578.

[55] Averbeck, B. B.; Latham, P. E. and Pouget, A. (2006). Neural correlations, population coding and computation. Nature reviews. Neuroscience 7, 358-366.

[56] Serre, T.; Oliva, A. and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Sciences of the United States of America 104, 6424-6429.

[57] Hung, C. P.; Kreiman, G.; Poggio, T. and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science (New York, N.Y.) 310, 863-866.

[58] Chalk, M.; Marre, O. and Tkačik, G. (2018). Toward a unified theory of efficient, predictive, and sparse coding. Proceedings of the National Academy of Sciences of the United States of America 115, 186-191.

[59] Gilbert, C. D. and Li, W. (2013). Top-down influences on visual processing. Nature reviews. Neuroscience 14, 350-363.

[60] Homann, J.; Koay, S. A.; Glidden, A. M.; Tank, D. W. and Berry, M. J. (2017). Predictive

Coding of Novel versus Familiar Stimuli in the Primary Visual Cortex biorxiv , .

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

[62] Liao, Q. and Poggio, T. (2016). Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex arxiv , .

[63] Kubilius, J.; Schrimpf, M.; Nayebi, A.; Bear, D.; Yamins, D. L. K. and DiCarlo, J. J. (2018). CORnet: Modeling the Neural Mechanisms of Core Object Recognition biorxiv , .

[64] Froudarakis, E.; Berens, P.; Ecker, A. S.; Cotton, R. J.; Sinz, F. H.; Yatsenko, D.; Saggau, P.; Bethge, M. and Tolias, A. S. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. Nature neuroscience 17, 851-857.

[65] Hénaff, O. J.; Goris, R. L. T. and Simoncelli, E. P. (2019). Perceptual straightening of natural videos. Nature neuroscience 22, 984-991.

[66] Vascon, S.; Parin, Y.; Annavini, E.; D'Andola, M.; Zoccolan, D. and Pelillo, M. (2018). Characterization of Visual Object Representations in Rat Primary Visual Cortex , .

[67] Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. Zeitschrift fur Naturforschung. Section C, Biosciences 36, 910-912.

[68] Tkacik, G.; Prentice, J. S.; Victor, J. D. and Balasubramanian, V. (2010). Local statistics in natural scenes predict the saliency of synthetic textures. Proceedings of the National Academy of Sciences of the United States of America 107, 18149-18154.

[69] Hunt, J. J.; Dayan, P. and Goodhill, G. J. (2013). Sparse coding can predict primary visual cortex receptive field changes induced by abnormal visual input. PLoS computational biology 9, e1003005.

[70] Li, Y.; Fitzpatrick, D. and White, L. E. (2006). The development of direction selectivity in ferret visual cortex requires early visual experience. Nature neuroscience 9, 676-681.

[71] Li, Y.; Van Hooser, S. D.; Mazurek, M.; White, L. E. and Fitzpatrick, D. (2008). Experience with moving visual stimuli drives the early development of cortical direction selectivity. Nature 456,

952-956.

[72] Matteucci, G. and Zoccolan, D. (2019). Causal adaptation to visual input dynamics governs the development of complex cells in V1 biorxiv , .

[73] Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607-609.

[74] Hermundstad, A. M.; Briguglio, J. J.; Conte, M. M.; Victor, J. D.; Balasubramanian, V. and Tkačik, G. (2014). Variance predicts salience in central sensory processing. eLife 3, .

[75] Kuang, X.; Poletti, M.; Victor, J. D. and Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. Current biology : CB 22, 510-514.

[76] Kleiner, M.; Brainard, D.; Pelli, D.; Ingling, A.; Murray, R. and Broussard, C. (2007). What's new in psychtoolbox-3? Perception 36, 1-16.

[77] Rossant, C.; Kadir, S. N.; Goodman, D. F. M.; Schulman, J.; Hunter, M. L. D.; Saleem, A. B.; Grosmark, A.; Belluscio, M.; Denfield, G. H.; Ecker, A. S.; Tolias, A. S.; Solomon, S.; Buzsaki, G.; Carandini, M. and Harris, K. D. (2016). Spike sorting for large, dense electrode arrays. Nature neuroscience 19, 634-641.

[78] Rikhye, R. V. and Sur, M. (2015). Spatial Correlations in Natural Scenes Modulate Response Reliability in Mouse Visual Cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 35, 14661-14680.

[79] Katsnelson, J. and Kotz, S. (1957). On the upper limits of some measures of variability Archiv für Meteorologie, Geophysik und Bioklimatologie Serie B 8, 103-107.

[80] Wales, D. J. and Scheraga, H. A. (1999). Global Optimization of Clusters, Crystals, and Biomolecules Science 285, 1368-1372.

[81] Byrd, R. H.; Lu, P.; Nocedal, J. and Zhu, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization SIAM Journal on Scientific Computing 16, 1190-1208.

[82] Bates, D. M. and Watts, D. G. (1988). Nonlinear regression analysis and its applications. Wiley New York, .

[83] Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R. and Lin, C.-J. (2008). LIBLINEAR: A Library for Large Linear Classification J. Mach. Learn. Res. 9, 1871-1874.

[84] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M. and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python Journal of Machine Learning Research 12, 2825-2830.

[85] Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python , .

[86] Wilcox, R. R. (2011). Introduction to robust estimation and hypothesis testing. Academic press, .

[87] Efron, B. and Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press, .

[88] R Core Team (2019). R: A Language and Environment for Statistical Computing, .

[89] Canty, A. and Ripley, B. D. (2019). boot: Bootstrap R (S-Plus) Functions, .

Acknowledgements

EP thanks Stefano Panzeri for many discussions, prior to this work, on temporally persistent neural codes and in particular on the topic of temporal noise correlations. We thank Walter Vanzella for his help in setting up the first version of the decoding analysis. This work was supported by a European Research Council (ERC) Consolidator Grant (DZ, project n. 616803-LEARN2SEE) and NIH grant R01EY07977 (VB).

Author Contributions

Conceptualization, E.P., L.S., V.B. and D.Z.; Data Curation, E.P. and L.S.; Formal Analysis, E.P. and L.S.; Funding Acquisition, V.B. and D.Z; Investigation, L.S.; Methodology, E.P., L.S., V.B and D.Z.; Project Administration, D.Z.; Resources, D.Z.; Software, E.P., L.S. and R.C.; Supervision, D.Z.; Writing – Original Draft Preparation, E.P.; Writing – Review & Editing, E.P., L.S., V.B. and D.Z.

Competing Interests

The authors declare no competing interests.