

## ORIGINAL ARTICLE

# Neural Representations of Natural and Scrambled Movies Progressively Change from Rat Striate to Temporal Cortex

Kasper Vinken<sup>1,2</sup>, Gert Van den Bergh<sup>1</sup>, Ben Vermaercke<sup>1</sup>  
and Hans P. Op de Beeck<sup>1</sup>

<sup>1</sup>Laboratory of Biological Psychology and <sup>2</sup>Laboratory for Neuro- and Psychophysiology, KU Leuven, 3000 Leuven, Belgium

Address correspondence to Hans P. Op de Beeck, Tiensestraat 102, 3000 Leuven, Belgium. Email: hans.opdebeeck@ppw.kuleuven.be

## Abstract

In recent years, the rodent has come forward as a candidate model for investigating higher level visual abilities such as object vision. This view has been backed up substantially by evidence from behavioral studies that show rats can be trained to express visual object recognition and categorization capabilities. However, almost no studies have investigated the functional properties of rodent extrastriate visual cortex using stimuli that target object vision, leaving a gap compared with the primate literature. Therefore, we recorded single-neuron responses along a proposed ventral pathway in rat visual cortex to investigate hallmarks of primate neural object representations such as preference for intact versus scrambled stimuli and category-selectivity. We presented natural movies containing a rat or no rat as well as their phase-scrambled versions. Population analyses showed increased dissociation in representations of natural versus scrambled stimuli along the targeted stream, but without a clear preference for natural stimuli. Along the measured cortical hierarchy the neural response seemed to be driven increasingly by features that are not V1-like and destroyed by phase-scrambling. However, there was no evidence for category selectivity for the rat versus nonrat distinction. Together, these findings provide insights about differences and commonalities between rodent and primate visual cortex.

**Key words:** electrophysiology, extrastriate cortex, object representation, rodent, ventral stream

## Introduction

Visual perception is the end product of a series of computations that start in the retina and culminate in several cortical areas. Although we can readily experience this end product effortlessly, decades of intensive research still have not yielded a full picture about the computations taking place beyond the point where visual information first arrives at the cortex, the primary visual area (V1). Until a few years ago, the neural underpinnings of visual perception were mainly investigated in primates and cats. With the recent surge of rodent studies involving new techniques

which have proven to be of high value to disentangle the mechanisms of visual processing (Huberman and Niell 2011), questions concerning the functional properties and capabilities of areas in rodent extrastriate visual cortex have become highly relevant. Behavioral experiments have found evidence in rats for forms of higher level visual processing (Zoccolan et al. 2009; Tafazoli et al. 2012; Vermaercke and Op de Beeck 2012; Alemi-Neissi et al. 2013; Brooks et al. 2013; Vinken et al. 2014; Rosselli et al. 2015; for review, see Zoccolan 2015), fueling the idea that these animals might be useful as an alternative and experimentally more flexible model

to tackle certain questions related to these complex visual capabilities.

In primates, extrastriate visual areas further integrate visual features that are encoded in V1 into more complex representations (Orban 2008). These areas have traditionally been grouped into 2 anatomically and functionally distinct pathways: a dorsal stream and a ventral stream (Mishkin and Ungerleider 1982; Kravitz et al. 2011). The latter is responsible for the transformations that eventually produce the ingredients necessary for extraordinary abilities such as object recognition, namely high selectivity distinguishing between objects, combined with tolerance for a range of identity preserving transformations, such as changes in size, position, viewpoint, illumination, etc. (DiCarlo and Cox 2007; DiCarlo et al. 2012). The result is a high level representation that manifests itself in strong categorical responses in monkey and human ventral regions, with, for example, a high selectivity for the distinction between animal and nonanimal pictures (Kiani et al. 2007; Kriegeskorte, Mur and Bandettini 2008). This category selectivity comes on top of a general preference in primate occipitotemporal cortex for natural, intact images compared with scrambled versions of these stimuli. Thus, in primates the computations along the ventral pathway introduce a bias in favor of coherent stimuli containing surfaces and objects over random texture patterns. This preference for intact coherent images was found higher up in this pathway through human functional magnetic resonance imaging (fMRI, Grill-Spector et al. 1998), monkey fMRI (Rainer et al. 2002), and monkey single-neuron physiology (Vogels 1999). This bias does not exist in lower levels of the pathway where sometimes even a preference for scrambled images is found (Rainer et al. 2002), potentially depending upon the exact scrambling procedure (Stojanoski and Cusack 2014).

Can we find evidence for similar computations being performed in the rodent brain? Previous research has suggested that anatomically the rodent visual cortex consists of 2 streams resembling the dorsal and ventral pathways in primates (Niell 2011; Wang et al. 2012). Already some steps have been taken to investigate the functional properties of rodent extrastriate cortex using drifting bars and gratings (Andermann et al. 2011; Marshel et al. 2011) and simple shapes (Vermaercke et al. 2014). Marshel et al. (2011) reported that mouse latero-intermediate area (LI) prefers high spatial frequencies, which might indicate a role in the analysis of structural detail and form. Vermaercke et al. (2014) report an increase in position tolerance, consistent with the primate ventral visual stream, along a progression of 5 cortical areas starting in V1 and culminating via LI in recently established lateral occipitotemporal area TO (Vermaercke et al. 2014). This increased position tolerance paralleled a gradual transformation of the selectivity for the simple shapes used in the study. However, these areas were hardly selective to stationary shapes and were more responsive to moving stimuli, which contrasts with the primate ventral visual stream. More complex stimuli such as natural movies have rarely been used in rodents, with 2 recent exceptions (Kampa et al. 2011; Froudarakis et al. 2014). In those studies the focus was primarily on the population code in primary visual cortex. Kampa et al. (2011) measured responses of V1 layer 2/3 populations to dynamic stimuli (including natural movies), showing reliable stimulus-specific tuning and evidence for functional sub-networks (despite the lack of orientation columns in rodent V1). Froudarakis et al. (2014) found that natural scenes evoke a sparser population response compared with phase-scrambled movies, leading to an improved scene discriminability that also depended on cortical state. Both studies focused on primary visual cortex and not explicitly on coding of movie content.

Here we investigated whether the 2 most salient functional hallmarks of neural object representations in primates might also exist in rodents: preference for intact versus scrambled stimuli and category-selective responses. To achieve this, we recorded action potential activity in 3 areas belonging to this putative ventral stream in rats with the aim of systematically comparing how stimulus representations change across areas: V1, LI, and TO. LI is the most downstream area in the putative ventral visual pathway which has been identified in both mice and rats (Espinoza and Thomas 1983; Wang and Burkhalter 2007); TO extends even further to rat temporal cortex and its responses to simple grating stimuli and shapes already suggested a higher-order processing compared with the other areas (Vermaercke et al. 2014). While recording neural responses, we presented natural movies belonging to different categories, as well as phase-scrambled versions of these movies. Based on the primate research, we would expect very different results in higher stages of the cortical processing hierarchy. First, a functional hierarchy would be supported by a systematic and gradual change in population representation of scrambled versus natural movies across areas. Second, a change culminating in a preference for natural movies would show that this functional hierarchy is comparable to the primate ventral visual stream in this respect, a notion that would even be more supported by a categorical representation toward the most downstream area TO.

## Materials and Methods

Much of the materials and methods have been described previously in detail (for descriptions of the apparatus, methodological details, and functional criteria, see Vermaercke et al. 2014; for a description of the stimuli, see Vinken et al. 2014). There was however no overlap and animals were completely naïve with respect to the stimulus set. Here we focus upon the details which are most important and most relevant in the context of the present study.

### Animals

Experiments were conducted with 7 male FBNF1 rats, aged 14–30 months (21 on average) at the start of the study. This specific breed was chosen for their relatively high visual acuity of 1.5 cycles per degree (Prusky et al. 2002). Surgery was performed to implant a head post and a recording chamber. The craniotomy was centered  $-7.9$  mm anteroposterior and  $-2.5$  mm lateral from bregma. This location allowed the electrode to pass through 5 different visual areas, including our 3 target areas, when entering at an angle of  $45^\circ$ : V1, latero-medial area, LI, latero-lateral area, and TO (Vermaercke et al. 2014). As described in Vermaercke et al. (2014), we performed histology in 5 out of the 7 rat brains, confirming that the electrode tracks followed trajectories similar to that study (therefore also confirming a sampling bias toward upper layers in V1). After recovery, the animals were water deprived and had ad libitum access to food pellets. Housing conditions and experimental procedures were approved by the KU Leuven Animal Ethics Committee.

### Stimuli

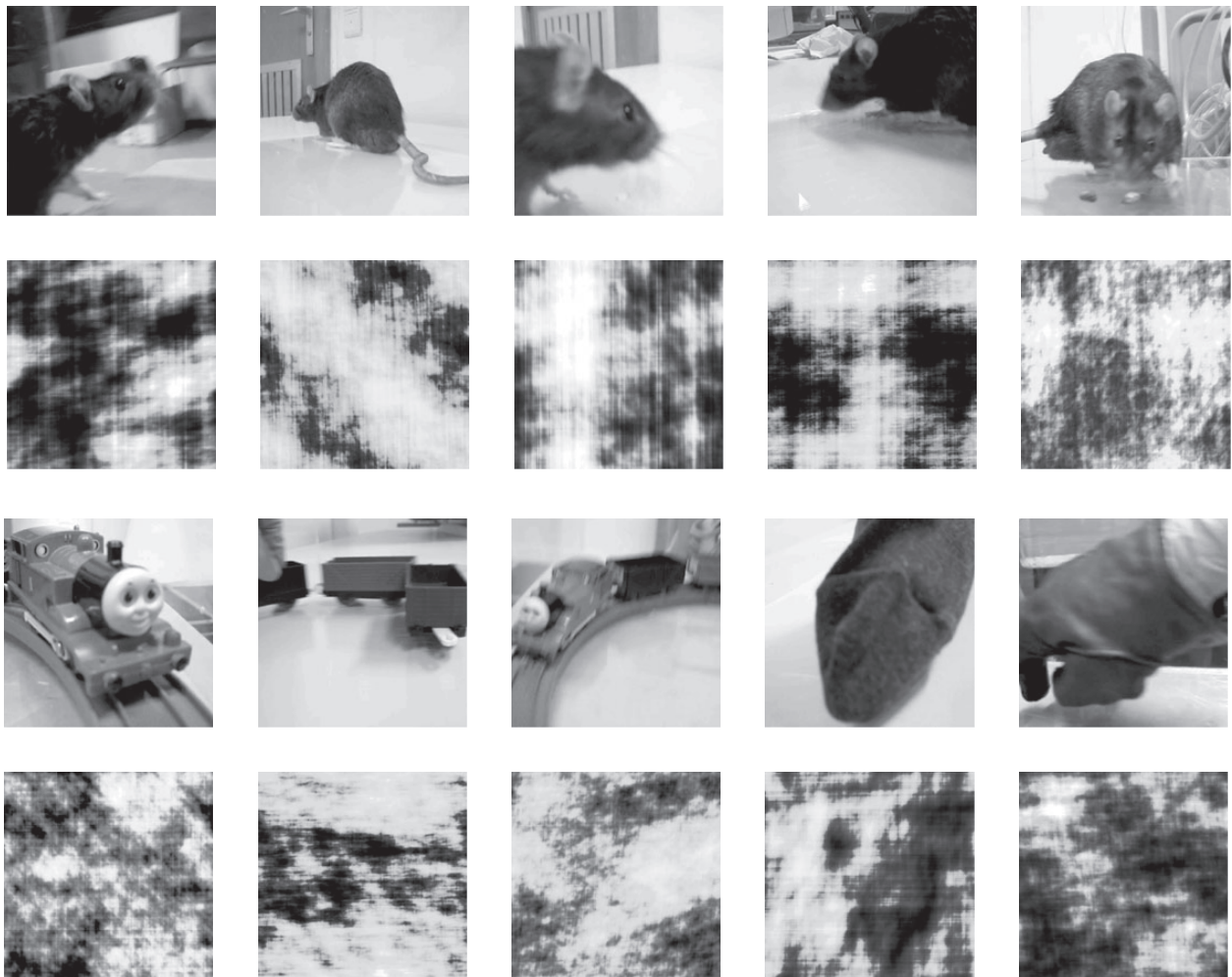
The set of stimuli used in this experiment corresponds to the training set described in Vinken et al. (2014). The set consisted of 20 movies: 10 natural movies and the phase-scrambled versions of these movies. The natural movies had a duration of 5 s and were recorded at 30 Hz (thus including 150 frames) and sized  $384 \times 384$  pixels. Five of them contained a rat, while the other 5 contained a moving object. For each rat movie, a nonrat

movie was chosen from our own database of 537 five-second movies in order to match relatively well on pixel intensities, contrast, and changes in pixel intensities (Vinken et al. 2014). The rat movies showed moving rats of the same strain as the subjects. Three of the paired nonrat movies contained a train, one a gloved hand moving in and out of the screen, and one a moving stuffed sock. For each movie a phase-scrambled version was created according to the procedure described previously, which allows for a better framewise match according to statistics such as average pixel intensity, contrast, changes in pixel intensity across successive frames, as well as spatial power spectrum compared with standard methods (Vinken et al. 2014). See Figure 1 for snapshots of each movie. The original movies and their scrambled versions were created at a size of  $384 \times 384$  pixels (to reduce memory load), but in the electrophysiological experiment the movies were shown at a size of  $768 \times 768$  pixels.

### Electrophysiological Recordings

As described at full length by Vermaercke et al. (2014), the rats were head-fixed and placed in front of a 24" LCD screen ( $1280 \times 768$  at 60 Hz), which was  $\gamma$  corrected to obtain a linear transfer function between pixel intensity values and luminance.

The animal's nose pointed at the left edge at an angle of  $40^\circ$  and a closest eye-to-screen distance of 20.5 cm. The movies were always presented at the full height of the screen (768 pixels) and positioned on the horizontal axis according to the estimated receptive field location (see receptive field mapping below). This resulted in a stimulus width ranging from 50 to 74 visual degrees depending on the position (as the eye-to-stimulus distance varies according to position). During the experiments every fifth (movie experiment) or tenth (receptive field estimation) stimulus presentation a water reward was given. Recordings were performed with a Biela Microdrive and single high-impedance electrodes (FHC, Bowdoin, ME; ordered with impedance 5–10 M $\Omega$ ) in areas V1, LI, and TO. Spike detection was done using custom written code in Matlab (The MathWorks, Inc., Natick, MA), with the spike detection threshold set to detect spikes with a peak-to-peak amplitude of 4 times the standard deviation (SD) of the noise. Single units (SUs) were isolated based on cluster analysis of the properties of the recorded waveforms (the first  $n$  principal components, where  $n$  was optimized to the situation) using KlustaKwik 1.6, followed by a manual check in SpikeSort 3D 2.5.1. Spike waveforms that could not be separated into SUs were pooled into one multiunit (MU) cluster per recording site (each spike waveform was only used once, so there is no overlap



**Figure 1.** Representative snapshots of the movies used in the experiments. First row depicts the original natural rat movies, with the corresponding scrambled versions represented on the second row. The third row depicts the natural stimuli belonging to the nonrat category, each matched to the rat movie displayed in the same column (see Materials and Methods, Stimuli). From left to right: 3 movies of a toy train, one with a stuffed sock, and one with a gloved hand, with the corresponding scrambled versions represented on the fourth row. The full movies are available at <http://ppw.kuleuven.be/home/english/research/lbp/downloads/ratMovies> (Last accessed 19 April 2016).

between SU and MU). On average the peak-to-peak amplitude of the mean spike waveform was 12.4, 12.9, and 11.3 times the SD of the noise for V1, LI, and TO units respectively. For all except 2 (one in V1 and one in TO) of the neurons included for analysis, this signal-to-noise-ratio was higher than the criterion value of 5 used by Issa and DiCarlo (2012). For MU clusters these values were 4.9, 4.7, and 4.6 for V1, LI, and TO, respectively (note that these values are limited in the lower end by the spike detection threshold of 4 times the SD of the noise).

### Receptive Field Mapping

Boundaries of the 5 aforementioned different areas were estimated based on changes in retinotopy as described previously (Vermaercke et al. 2014). A rough estimate of a site's population receptive field could be obtained manually by using continuously changing shapes or drifting gratings that could be moved across the screen. A quantitative estimate of receptive field size and location was achieved by flashing a hash symbol at 15 locations (3 rows  $\times$  5 columns) on the screen. Movies were translated along the horizontal screen axis in order to best cover the receptive field. We chose to record from V1, LI, and TO, and not the 2 additional intermediate areas LM and LL, because the elevation of the receptive fields encountered in V1, LI, and TO tends to be very similar (see Fig. 2C in Vermaercke et al. 2014). In contrast, the receptive fields encountered in LM and in particular LL show a very different elevation, which would make it difficult to compare results between the different areas (the receptive fields of the neuronal populations would then cover different parts of the movies).

### Presentation of Movies

In the main experiment, rats were passively viewing the 10 five-second natural movies and the phase-scrambled versions of these movies. These were presented in random order intermitted by a 2-s blank screen, with 10 repetitions per movie. The pixel

intensity value of the blank screen and the part of the screen not covered by the movies was set equal to the average pixel intensity of all movies.

### Data Analysis

We maintained 2 criteria to include units for analysis: units needed to be isolated for the full 10 presentations of each movie and have an average net response of  $>2$  Hz for at least one movie.

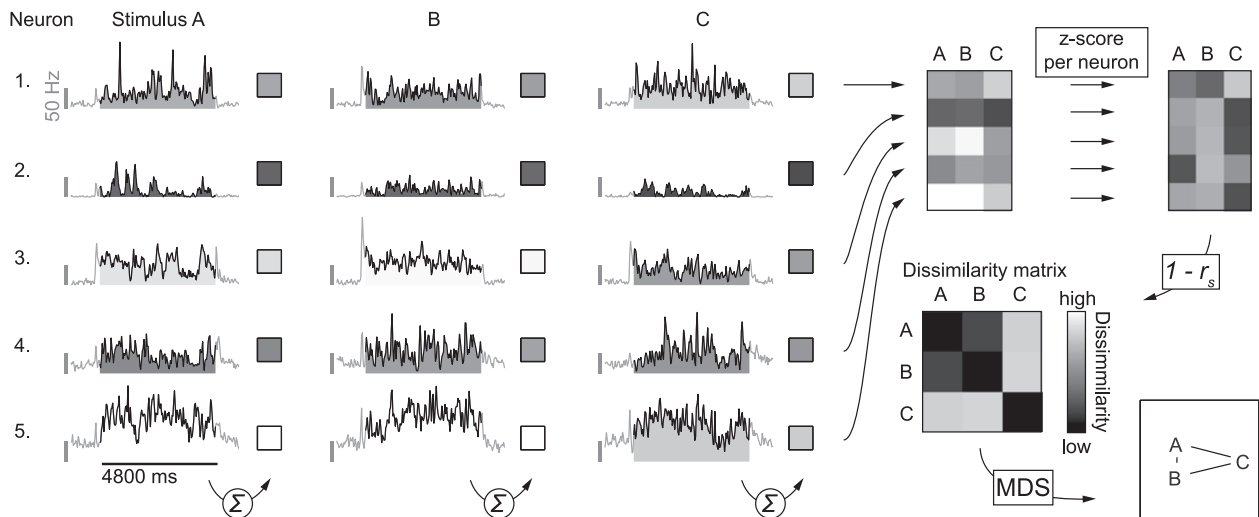
### Preprocessing

Before all analyses, peristimulus time histograms (PSTHs) with a bin width of 1 ms were made for each trial across the  $[-1999, 6000]$ -ms interval (with stimulus onset at 0 ms). To estimate the response onset latency the PSTHs were averaged across trials and stimuli and smoothed with a Gaussian kernel (3 ms full width at half maximum). Response onset latency was defined per unit as the first time point after stimulus onset where the smoothed PSTH exceeded a threshold of the baseline activity (calculated from the  $[-999, 0]$ -ms window) plus 3 times the baseline activity SD. For all further analyses, only a 4800-ms time window after response onset was used, with the first 200 ms cutoff to ignore the onset peak mainly for fitting the motion energy model. For consistency the same 4800-ms window was used for all other analyses, even though the inclusion of the window does not affect the results in any significant way.

### Sparseness and Reliability

Response sparseness for a certain neuron to a certain movie was quantified using the index defined by Vinje and Gallant (2000):

$$S = \left( \frac{1 - (1/n) \left( \frac{\sum_i r_i}{\sum_i r_i^2} \right)^2}{1 - (1/n)} \right),$$



**Figure 2.** Schematic illustration of how dissimilarity matrices were calculated. For the analysis of the population representation we started with the responses of SUs, averaged across repeated stimulus presentations and summed across the 201 to 5000 ms window calculated from response onset (as indicated by the summation symbol). PSTH's in this example figure illustrate responses averaged across stimulus presentations of 5 SUs to 3 movies (A–C), all arbitrarily selected for the purpose of illustrating the methods. This was done for each stimulus (e. g. A–C, in this example) to get raw response vectors. These raw responses were further standardized per neuron, by calculating the Z-scores across stimuli. Next the correlation matrix was calculated from these normalized response data by pairwise correlation of the stimulus response vectors. For the stimulus dissimilarity matrix each value in this correlation matrix was subtracted from one, resulting in values between 0 and 2, where 0 indicates the lowest dissimilarity (i.e., an identical population response pattern) and 2 indicates the highest dissimilarity (i.e., a highly different population response pattern). To visualize the stimulus space as represented by the population of neurons, we performed multidimensional scaling on the dissimilarity matrix and present the stimuli using the first 2 dimensions. Stimuli plotted closer together (A,B in this example) have a more similar population response pattern than stimuli plotted further apart (A,C, and B,C in this example).

where  $S$  is the sparseness index for a neuron with average (across trials) response  $r_i$  to frame  $i$  of a stimulus with  $n$  frames. Onset of the first bin is the estimated response latency plus 200 ms (see pre-processing) and the bin-width is  $\sim 33.3$  ms, which corresponds with the frame rate. This sparseness index can vary between 0 and 1, with values close to 0 indicating a dense response, and values close to 1 indicating a sparse response. Response reliability for the time course of the response of a certain neuron to a certain movie was estimated using the Spearman–Brown correction as follows

$$\frac{nr_{xx'}}{1 + (n - 1)r_{xx'}}$$

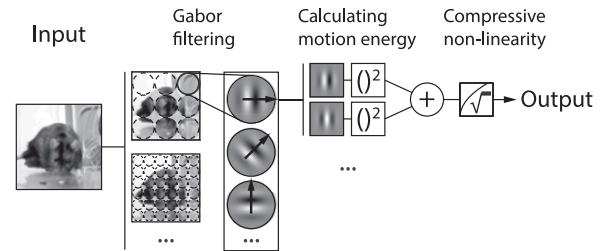
with the average correlation across time between 2 trials  $r_{xx'}$  (across all combinations) and number of trials  $n$ . As before, the bin width to calculate the reliability was set to correspond with the frame rate of 30 Hz.

### Population Representation

For each population of neurons (i.e., in V1, LI, or TO) pair-wise stimulus dissimilarities were calculated based on the correlation distance using the average responses to full movies. First, firing rates were averaged across trials and per stimulus across the entire 4800-ms interval, resulting in 20 responses per neuron. Second, responses of each neuron were transformed to Z-scores (across stimuli). Third, for each stimulus a response vector was created containing the transformed responses of each neuron to that particular stimulus. Finally, dissimilarity between a pair of stimuli is defined as  $1 - r$  (Pearson correlation) between the response vectors of the 2 stimuli in question. Figure 2 illustrates how we used this method to create dissimilarity matrices.

### Spatiotemporal Motion Energy Model

To simulate the relative response of V1-like cells we calculated the output of a spatiotemporal motion energy model (Adelson and Bergen 1985). Specifically, the spatiotemporal receptive field of a modeled V1 neuron is based on a 3-dimensional Gabor filter, with a certain frequency, orientation, and location relative to the stimulus. The output of the filter (calculated through linear multiplication with the stimulus) is then squared, and summed with the output of the quadrature pair to that filter which is  $90^\circ$  out of phase. This squared and summed output gives a physiologically plausible measure of motion energy. The square root of this measure is our modeled response of a complex V1 cell (Nishimoto and Gallant 2011). See Figure 3 for a schematic representation of this process. A wide range of Gabor filters spanning different frequencies, orientations, and locations is then used to model our set of V1 cells. Thus, we end up with a modeled V1 complex cell for each spatiotemporal frequency, orientation, and spatial location included in the model. The output of each modeled cell is then standardized by calculating the Z-score across all movie frames. The set of Gabor filters spanned 8 different directions, 6 different spatial frequencies, and 6 different temporal frequencies. The spatial frequencies were log spaced between 0.04 and 0.15 cycles per degree and the temporal frequencies between 0 and 15 Hz, based on the optimal responses of rat V1 neurons reported by Girman et al. (1999). Each filter occurred at different spatial locations. Grid spacing was identical to what is reported by Nishimoto and Gallant (2011) and depended on spatial frequency: filters were separated by 2.2 SDs of the Gaussian envelope, with one SD set to half a cycle of the sine wave. Next, the output of these filters was used as predictors in a regularized linear regression model with an early stopping



**Figure 3.** Schematic representation of the motion energy model (based on Nishimoto et al. 2011), described under Materials and Methods, Data Analysis, Spatiotemporal Motion Energy Model. In short, input stimuli (movies) are run through a bank of quadrature pairs of Gabor filters, each with a certain spatiotemporal frequency and orientation and located on a grid covering the stimulus. The output of each pair is then squared and summed to give a physiologically plausible measure of motion energy. The end result is finally obtained by taking the square root to model a compressive non-linearity. This final output is calculated for each of the spatiotemporal frequencies and locations covered by the bank of Gabor filters and standardized per filter across frames. In a next step, the neural response is predicted as a linear combination of those standardized outputs.

rule (David et al. 2007; Nishimoto and Gallant 2011) fitted to the neural responses using code from the STRFlab toolkit (version 1.45, retrieved from <http://strflab.berkeley.edu/>). The model was estimated at 5 different latencies, ranging from 20 to 153 ms in steps of the duration of one frame. For each unit and latency the model was fit 10 times, each time refraining 2 movies (i.e., a natural movie and its scrambled pair) from the fitting procedure for cross-validation and using the remaining 18 movies for training the model. Reported accuracies refer always to data that were not included in the training set.

### Statistical Analysis

For statistical inference we relied on the bias-corrected accelerated bootstrap (BCa; Efron 1987) by random sampling with replacement (10 000 iterations) of the neurons/units (unless indicated otherwise) to estimate the 95% confidence interval (CI) of the statistic in question. In addition, randomization tests (10 000 iterations) are used to estimate the distribution of the test statistic in question under the null hypothesis in order to calculate P-values. In several places we report the slope of a linear regression to quantify gradual change across the 3 regions in the pathway under investigation for reasons of simplicity and interpretability, without the intention of making strong claims of linearity. However, we formally tested for a deviation of linearity by including categorical dummy variables for each region in the regression. In none of the cases where we report a slope did a categorical predictor show a significant effect, which would indicate that there would be a nonlinear component. Thus the simpler model with one linear trend is preferred.

## Results

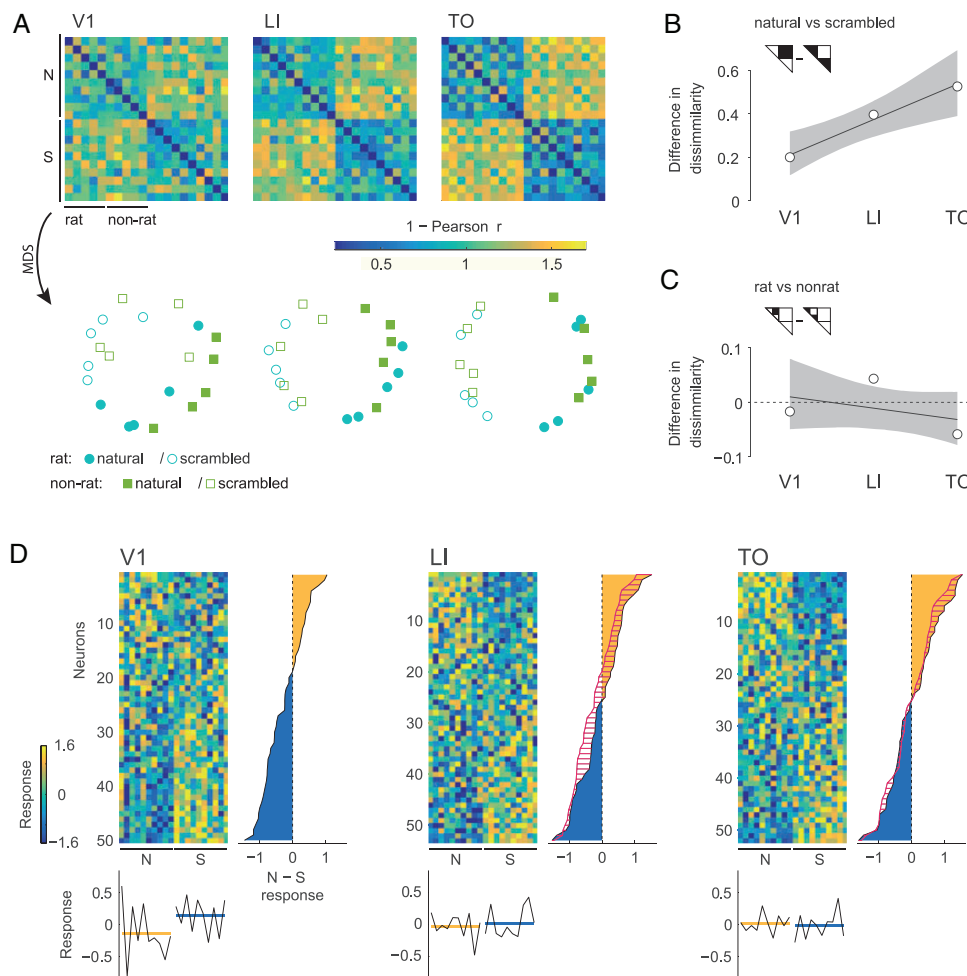
We recorded the activity of single neurons in 3 areas of awake rats, namely V1, LI, and TO (Vermaercke et al. 2014), while presenting natural movies containing a rat or not as well as scrambled movies. The recordings yielded 50 (out of 58, or 86%) responsive SUs for V1, 53 (out of 88, or 60%) for LI, and 52 (out of 84, or 62%) for TO, as well as 25 (out of 25, or 100%), 33 (out of 35, or 94%), and 26 (out of 30, or 87%) responsive MU sites for each area respectively (percentages indicate the proportion of units that passed the inclusion criteria for analysis, that is, an average net response of more than 2 Hz for at least one movie).

We tested for (1) a change in representation of stimulus type (natural versus scrambled) across areas, supporting a functional hierarchy, (2) the emergence of a categorical representation for the distinction between rat and nonrat movies, and (3) the emergence of a preference for natural movies along these areas.

### Stimulus Representations in V1, LI, and TO

To get an idea of how the stimuli are represented by the populations of neurons, we investigated the neural stimulus dissimilarities in the  $N$ -dimensional representational space defined by the average response of  $N$  neurons to the individual stimuli (Fig. 2). Two stimuli that elicit a very different response in the population of recorded neurons will result in a higher dissimilarity value. On the other hand, if a population of neurons shows the same response pattern to 2 different stimuli, the dissimilarity value will be zero. These dissimilarity values can be visualized in a

dissimilarity matrix as in Figure 4A (top row), where more yellow colors indicate higher pairwise stimulus dissimilarity. Visual inspection of the dissimilarity matrices suggests that moving from V1 to TO, a pattern emerges which can be summarized as an increased structuring by quadrants in the matrix: between stimulus type (i.e., natural or scrambled) dissimilarities increase relative to within stimulus type dissimilarities, which leads to an increased dissociation of natural versus scrambled movies. This is also illustrated by the plots on the lower row of Figure 4A, where the similarity representations are visualized in 2-dimensional space after performing nonmetric multidimensional scaling (MDS) on each dissimilarity matrix (using the function `mdscale` in Matlab, The MathWorks, Inc., Natick, MA, with the number of dimensions set to 2 and criterion set to “stress”). These plots show an increased separation between natural movies versus scrambled movies. This increased separation is also supported by further statistical analyses. The difference of



**Figure 4.** Stimulus representations based on responses averaged across movie durations. Panel A shows the stimulus dissimilarity matrices based on the correlation distance for each population of neurons recorded in V1, LI, and TO (upper row). Nonmetric MDS is then used to represent the representational space in 2 dimensions (lower row). Panel B shows the difference between dissimilarities for stimulus pairs of the opposite stimulus type (i.e., natural versus scrambled) and dissimilarities for pairs of the same type in areas V1, LI, and TO. Gray area indicates 95% confidence bounds for OLS regression, calculated by BCa. Panel C is the same as panel B, but for the difference between dissimilarities for stimulus pairs of the opposite stimulus category (i.e., rat versus nonrat) and dissimilarities for pairs of the same category. Panel D contains heatmaps (one per area) displaying the average response to each movie standardized per neuron (Z-score across stimuli). Neurons (rows) are sorted in descending order according to the values of the average standardized response to natural movies minus that to their scrambled versions. To the right of each heatmap the average of this value used for sorting is plotted per neuron, with the yellow area indicating neurons that respond more to natural images than to their scrambled version and the blue area indicating the reverse. For LI and TO, red hatching indicates how this distribution changed from V1 and LI respectively. Stimuli (columns) are sorted in the same way as in the dissimilarity matrices: 5 natural rat movies, 5 natural nonrat movies, and their scrambled versions in the same order. Column averages are displayed below each heatmap (black lines), with the average across stimulus type (yellow for natural movies, blue for their scrambled version) indicated in color.

average between stimulus type and average within stimulus type dissimilarities increases per area (Fig. 4B; ordinary least squares, OLS, slope 0.16, 95% CI [0.06, 0.26],  $P = 0.003$ ), with a value of 0.20 (95% CI [0.11, 0.32],  $P < 0.001$ ) for V1 neurons, 0.39 (95% CI [0.25, 0.54],  $P < 0.001$ ) for LI neurons, and 0.53 (95% CI [0.36, 0.70],  $P < 0.001$ ) for TO neurons. Thus, the distinction between natural and scrambled movies becomes more dominant in the neural representation when we move up in the cortical hierarchy.

In order to relate single cell responses to this population effect we plotted the standardized (per neuron) responses of each neuron to each stimulus that were used to create the dissimilarity matrices (Fig. 4D). Here we see that the curve showing average natural minus scrambled responses per neuron (to the right of each heatmap) is generally shifted to the right for LI compared with V1. This means that the distribution of a natural versus scrambled comparison shifts in favor of natural stimuli from V1 to LI causing more neurons to respond more to natural than to scrambled stimuli. For TO however, this curve has moved to the right nearly only for neurons responding more to natural stimuli. Thus, in TO the proportion of neurons responding more to natural stimuli is not necessarily different compared LI, but the natural/scrambled difference is higher for those that do respond stronger to natural movies.

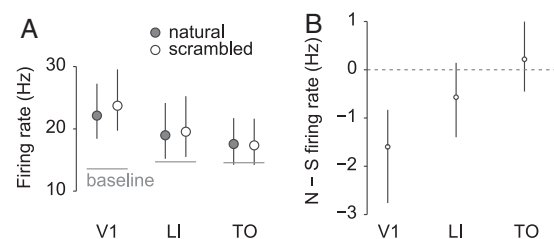
Is this increased sensitivity for natural versus scrambled movies accompanied by an increase in category selectivity, that is, a differentiation between movies that depict a rat versus movies without rat? This would result in a similar “structuring by quadrants” as described in the previous paragraph, but now within the left upper quadrant of the dissimilarity matrix. Visual inspection does not suggest that this pattern exists in the representational space for the populations of neurons recorded in either V1, LI, or TO. The MDS plots suggest an overlap in representations for rat versus nonrat movies without a clear separation. We performed further statistical analysis where we compared the average within stimulus category dissimilarities with average between stimulus category dissimilarities. The results do not show an emerging trend that would support an increased separation between representations of rat movies and those of nonrat movies (OLS slope  $-0.02$ , 95% CI  $[-0.06, 0.02]$ ,  $P = 0.432$ ). Looking at each area separately, the difference in dissimilarity is  $-0.02$  (95% CI  $[-0.08, 0.06]$ ,  $P = 0.696$ ) for V1 neurons,  $0.04$  (95% CI  $[-0.03, 0.13]$ ,  $P = 0.269$ ) for LI neurons, and  $-0.06$  (95% CI  $[-0.10, -0.01]$ ,  $P = 0.148$ ) for TO neurons. Positive values signify higher within category similarity than between category similarity, which is what one would expect in the case of a categorical representation. None of the areas shows such a categorical representation. To further strengthen these findings, we performed a potentially more sensitive population decoding analysis using support vector machines as a linear classifier. In agreement with the other analyses discussed above, the results of the linear classifier reveal no evidence for a categorical representation (see [Supplementary Material](#), Population Decoding Analysis).

As described previously ([Vermaercke et al. 2014](#)), moving from V1 to LI and to TO is characterized by systematic changes in retinotopic location and size of the receptive field. Furthermore, we might sample neurons with different receptive field properties in the area of the stimulus covered by our recorded samples of neurons across areas. However, a control analysis using average local stimulus statistics for each neuron’s receptive field shows that this confound cannot account for the increased natural/scrambled distinction from V1 to TO. In addition, there is no evidence that an emergence of a categorical rat/nonrat distinction

could be hidden by such a confound (see [Supplementary Material](#), Receptive Field Confound).

### Response Statistics for Natural and Scrambled Movies: Mean, Sparseness, and Reliability

Next, we looked at the average firing rates. Figure 5A shows the average firing rate for natural and scrambled movies (first averaged per neuron across movies for statistical analysis) for each area. Overall there is a statistically nonsignificant decrease in firing rate (OLS slope  $-2.7$ , 95% CI  $[-5.7, 0.1]$ ,  $P = 0.083$ ) moving from V1 to TO, while the average baseline firing rate does not seem to vary. The difference in response to natural movies versus scrambled versions is negative for V1 (see Fig. 5B) and this difference disappears toward the other areas, or, quantitatively, decreases significantly (OLS slope 0.9, 95% CI [0.4, 1.5],  $P = 0.003$ ). The average difference is  $-1.6$  for V1 (95% CI  $[-2.8, -0.8]$ ),  $-0.6$  for LI (95% CI  $[-1.4, 0.2]$ ), and  $0.2$  for TO (95% CI  $[-0.4, 1]$ ). Similar results are obtained when the difference in firing rate is first divided by the average firing rate per neuron (OLS slope 0.04, 95% CI [0.01, 0.08],  $P = 0.020$ ), with an average difference of  $-0.06$  for V1 (95% CI  $[-0.10, -0.02]$ ),  $0.01$  for LI (95% CI  $[-0.03, 0.06]$ ), and  $0.02$  for TO (95% CI  $[-0.02, 0.09]$ ). Control analyses show that these differences in firing rate between natural movies and their scrambled versions cannot be explained by differences in location and size of receptive fields (see [Supplementary Material](#), Receptive Field Confound). The reason why V1 neurons would prefer scrambled movies is further explored in a later section. For inference per neuron, BCa 95% CIs on the average difference in response to natural movies versus their scrambled version were calculated for each unit by means of random sampling of natural/scrambled stimulus pairs. We decided a unit prefers natural stimuli when this 95% CI excludes zero. This criterion indicates that 4% (95% CI [0, 14], based on the binomial distribution) of the units in V1 prefer natural stimuli, 23% (95% CI [12, 36]) in LI, and 27% (95% CI [16, 41]) in TO. Scrambled stimuli are preferred by 32% (95% CI [20, 47]) of the units in V1, 19% (95% CI [9, 32]) in LI, and 29% (95% CI [17, 43]) in TO. We conclude that, while there is an increase in the percentage of units consistently responding more to natural stimuli from V1 to TO, no clear change is evident for the percentage of units consistently responding more to scrambled stimuli. Applying the same criterion, rat stimuli are preferred by 18% (95% CI [9, 31]) of the units in V1, 6% (95% CI [1, 16]) in LI, and 13% (95% CI [6, 26]) in TO. Nonrat stimuli finally are preferred by 2% (95% CI [0, 11]) of the units in V1, 11% (95% CI [4, 23]) in LI, and 0% (95% CI [0, 7]) in TO. In general, it seems that a higher percentage of units tend to consistently respond more to



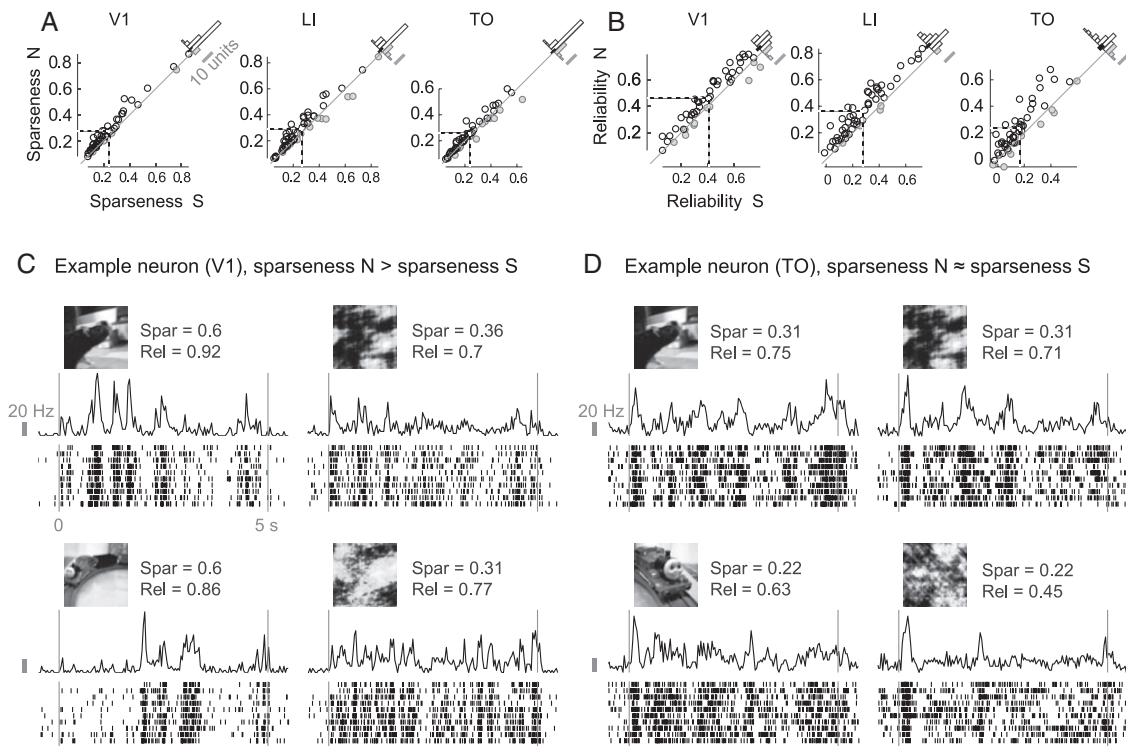
**Figure 5.** Firing rates per area. Panel A shows average (across trials, stimuli, and neurons) response strength to natural (gray markers) and scrambled (white markers) movies per area, with the average baseline firing rate indicated by a horizontal line. Error bars indicate the 95% CIs calculated by BCa. Panel B shows the average difference in response strength to natural and scrambled movies per area. Negative values indicate a stronger response to scrambled versions of the movies. Error bars indicate the 95% CIs calculated by BCa.

rat than to nonrat movies. However, since this is clearest for V1 neurons and since these percentages do not change progressively across areas, we conclude that this is most likely a result of lower level stimulus properties that V1 neurons typically respond to.

Next, we investigated the variation in responsiveness. Natural stimulation has been shown to increase the sparseness of the neural response (Vinje and Gallant 2000). We looked at the sparseness (see Materials and Methods, Data Analysis, Sparseness and Reliability) of each neuron's response to natural movies and to their scrambled counterparts. In mouse V1, sparseness has been shown to be higher for responses to natural movies than to their scrambled counterparts (Froudarakis et al. 2014). We confirm this finding for SUs in rat V1, with a difference in sparseness index of 0.035 (95% CI [0.027, 0.047],  $P < 0.001$ ); positive values mean higher response sparseness to natural movies; see Fig. 6A). Also in LI and TO we find a higher sparseness for natural movies, with a difference of 0.028 (95% CI [0.014, 0.042],  $P < 0.001$ ) and 0.015 (95% CI [0.004, 0.025],  $P = 0.011$ ), respectively. Importantly, local luminance-based stimulus sparseness calculated for each neuron's receptive field cannot explain the difference in response sparseness (see [Supplementary Material](#), Receptive Field Confound). However, if lower firing rates would tend to get higher sparseness index values and vice versa, some of these differences might be explained by the differences in firing rates shown before, in particular in area V1. Indeed differences in firing rates are negatively correlated with differences in sparseness index for each area, with a Pearson correlation of

$-0.28$  in V1 (95% BC<sub>a</sub> interval [−0.50, −0.001],  $P = 0.050$ ),  $-0.54$  in LI (95% CI [−0.69, −0.35],  $P < 0.001$ ), and  $-0.43$  in TO (95% CI [−0.67, −0.02],  $P = 0.002$ ). Thus we controlled for differences in firing rates by taking for each cortical area the 30 units with a difference in firing rates evenly distributed around zero. For these units, the average difference in sparseness indices was 0.025 in V1 (95% CI [0.017, 0.037],  $P < 0.001$ ), 0.024 in LI (95% CI [0.011, 0.037],  $P < 0.001$ ), and 0.011 in TO (95% CI [0.004, 0.019],  $P = 0.010$ ). Thus, responses to natural movies show a higher sparseness than responses to scrambled movies, even when we control for overall responsiveness.

Finally, responses to natural movies are decisively more reliable in all 3 areas (Fig. 6B). Reliability of V1 neural responses is on average 0.056 (95% CI [0.037, 0.072],  $P < 0.001$ ) higher to natural movies than to scrambled versions. For LI and TO neurons this difference is on average 0.090 (95% CI [0.074, 0.108],  $P < 0.001$ ) and 0.070 (95% CI [0.047, 0.097],  $P < 0.001$ ), respectively. In the case of reliability, we have relatively weak evidence of a possible influence of differences in response strength: the Pearson correlations between the difference in standardized reliability and the difference in standardized firing rates are 0.26 in V1 (95% CI [−0.04, 0.47],  $P = 0.070$ ), 0.12 in LI (95% CI [−0.08, 0.35],  $P = 0.385$ ), and 0.40 in TO (95% CI [−0.040, 0.64],  $P = 0.003$ ). The results for the 30 units selected to match for response strength (see paragraph above) of each cortical area are qualitatively similar to the results for the whole sample, with a difference of 0.059 in V1 (95% CI [0.034, 0.081],  $P < 0.001$ ), 0.095 in LI (95% CI [0.072, 0.120],  $P < 0.001$ ), and 0.052 in TO (95% CI [0.029, 0.077],  $P < 0.001$ ).



**Figure 6.** Sparseness and reliability of single neuron's responses. Panel A contains scatterplots of the sparseness index for natural (N) compared with the same index for scrambled (S) movies for all neurons recorded in V1 (left), LI (middle), and TO (right). Neurons with a lower index for natural movies are grayed out. Dashed lines indicate the means. Histograms of the difference between natural and scrambled stimuli are shown in the top right corner of each plot, with the 95% CI (calculated by means of BC<sub>a</sub>) of the mean indicated by a black bar. Panel B contains the same figures, but for the reliability coefficient. Panel C shows raster plots for an example neuron with relatively high response reliability of responses to 2 natural movies and their scrambled versions. For this example, response sparseness is much higher for the original stimuli compared with their scrambled version. Panel D shows raster plots for another example neuron. In this case response sparseness indices are equal for the 2 stimulus types.



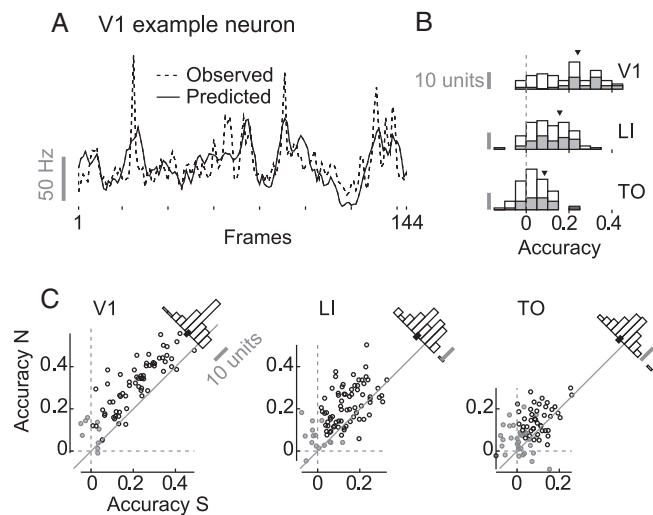
## Using a V1 Model to Explain the Preference for Scrambled Stimuli and to Predict Neural Responses

To further investigate these findings we used simulated V1 responses to see (1) whether these simulated responses can predict observed responses and, if the response to (1) is affirmative, (2) whether these simulated responses can explain the relative increase of the response for natural movies and the increased segregation between natural and scrambled movies. In the model, spatiotemporal motion energy filters (Adelson and Bergen 1985) are used as modeled V1 complex cells for a set of spatiotemporal frequencies, orientations, and spatial locations. The output of these filters can be used to estimate how strong responses in V1 would be to one stimulus relative to another one. Furthermore, the filters can be used in a model that is fitted to part of the data in order to predict independent test data (Nishimoto and Gallant 2011). Note that the V1-like filters are linearly combined to predict the neural responses, which is why the model might even capture the responses of neurons in higher visual areas to the extent that the complexity of their computations can be approximated by a linear combination of V1-like filters.

### Predicting Neural Responses

The standardized output of each V1 filter was used as a predictor in a regularized linear regression model that was fit to the neural data, resulting in a spatiotemporal receptive field estimate consisting of a linear combination of these filters. This receptive field estimate can then be used to predict the response to a new stimulus. If this predicted response captures a certain amount of variability in the response to movies that were not used for fitting the model, then the fitted receptive field can explain some of the response properties of the neuron. For this test, we also included the MU data, since their firing rate is the sum (i.e., a linear combination) of the firing rates of the SUs contained in the MU cluster which can be accounted for because the model can use a linear combination of outputs of modeled single cells.

The observed and predicted responses of a V1 example neuron on one natural movie are illustrated in Figure 7A. Histograms with prediction accuracy (i.e., Pearson correlation between observed and predicted responses) averaged across all stimuli are shown in Figure 7B. The model performs reasonably well for V1 units, with an average prediction accuracy of 0.24 (95% CI [0.21, 0.27], randomization test for difference from zero  $P < 0.001$ ). To put this number in perspective, this is lower than the prediction accuracy of 0.52 reported by Nishimoto and Gallant (2011) in monkeys. This was to be expected since the number of frames (data) we used to train the model (2736) is one order of magnitude lower (on average 27 120 in their case) and since we used natural movies and not motion enhanced movies. Prediction accuracy is lower but still significantly different from zero for LI (0.16, 95% CI [0.14, 0.18],  $P < 0.001$ ) and even lower but still significantly different from zero for TO (0.09, 95% CI [0.07, 0.10],  $P < 0.001$ ). In general, accuracy decreases across areas (OLS slope  $-0.08$ , 95% CI  $[-0.09, -0.06]$ ,  $P < 0.001$ ). Even when including reliability as a covariate accuracy decreases across areas (OLS slope  $-0.05$ , 95% CI  $[-0.06, -0.03]$ ,  $P < 0.001$ ). Dividing prediction accuracy by reliability gives us an estimate of the proportion obtained prediction accuracy out of the total possible prediction accuracy. When we do this per neuron and movie, we get an average of 0.46 (95% CI [0.41, 0.51], randomization test for difference from zero  $P < 0.001$ ) for V1, 0.34 (95% CI [0.30, 0.38],  $P < 0.001$ ) for LI, and 0.20 (95% CI [0.16, 0.24],  $P < 0.001$ ) for TO. This amounts to a decrease across areas (OLS slope  $-0.13$ , 95% CI  $[-0.16, -0.10]$ ,  $P < 0.001$ ). Thus, the aforementioned decrease in response reliability cannot explain the decrease in the performance of the model from V1 to TO.

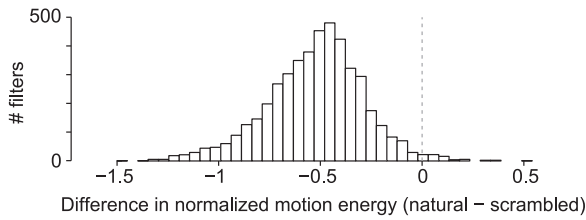


**Figure 7.** Performance of the V1-like motion energy model. Panel A shows the observed (dashed line) and predicted (full line) response of a V1 example neuron to each frame of a natural movie (note that the number of frames is 144, because the first 6 were omitted to get rid of the onset peak). Panel B contains stacked histograms for the prediction accuracy (Pearson correlation) averaged across all movies for single neurons (white) and MU clusters (gray) recorded in V1, LI, and TO. Panel C contains scatterplots of average (across movies) accuracy for natural versus scrambled movies for SUs and MUs. Histograms of the difference between natural and scrambled stimuli are shown in the top right corner of each plot, with the 95% CI (calculated by means of BCa) of the mean indicated by a black bar. Grayed out markers represent units for which the 95% CI of the average accuracy as calculated by BCa (resampling all of the 20 stimuli) does include zero.

Focusing on the performance for single neurons and multi-unit clusters, prediction accuracy averaged across all 20 stimuli (natural and scrambled) is significantly different from zero (i.e., the 95% CI—calculated by resampling the stimulus labels—excludes zero), for 85.3% of the units in V1, 82.5% of the units in LI, and 56.4% of the units in TO.

To get an estimate of receptive field size, we estimated the percentage of pixels in the area covered by the movie that modulate the neural response. Specifically, we used the median regression weights of the V1-like filters (across the 20 training sets used for cross-validation) in order to estimate the spatial receptive field. Pixels that were estimated to modulate the response with a magnitude  $< 50\%$  of that of the pixel that maximally modulated the response were excluded, to ignore the pixels that contribute relatively nothing. Based on this approach, V1 neurons were estimated to be modulated on average by 8% of the movies' pixels (95% CI [7%, 10%]), LI neurons by 14% (95% CI [12%, 17%]), and TO neurons by 16% (95% CI [13%, 20%]). This means that for LI neurons this movie frame coverage was 6% higher than for V1 neurons (95% CI [3%, 9%],  $P < 0.001$ ) and for TO neurons it was 8% higher than for V1 neurons (95% CI [5%, 12%],  $P < 0.001$ ). The difference in coverage between TO and LI neurons was 2% (95% CI  $[-2\%, 7\%$ ],  $P = 0.294$ ).

In sum, the simulated V1 model allows us to predict neural responses in each of the investigated areas, and reveals several differences between the areas which can be expected given their position in the cortical hierarchy: the model works better for V1 than for the other areas, and the estimated receptive field size is smallest in V1. Thus, the model can be used to further investigate potential differences in how these 3 neuronal populations respond to natural and scrambled movies.



**Figure 8.** Stimulus type preference in output of the V1-like motion energy model. Negative values indicate filter output for scrambled movies is on average higher than filter output for natural movies.

### Preference for Scrambled Stimuli in V1 Filters

To investigate the preference for one stimulus type over another, we calculated the output of the V1-like spatiotemporal motion energy model, thus before combining the V1 filters into a spatiotemporal receptive field estimate. For the vast majority of modeled V1 filters, the overall (across time) response to natural movies is lower than that to their scrambled versions. Figure 8 contains a histogram depicting the standardized response to scrambled movies subtracted from the response to their natural counterparts for all filters in the model (in this case 4616), with 98.6% of them preferring scrambled movies. This means that scrambled movies do seem to contain relatively more motion energy in virtually all spatiotemporal frequencies regardless of orientation and location. Other scrambling methods, such as segment/box scrambling (i.e., random repositioning of rectangular image segments), give qualitatively the same result (data not shown). This characteristic of scrambled stimuli has been reported before in the context of scrambling of still images using various methods including phase scrambling (Stojanoski and Cusack 2014) and in an fMRI experiment using phase-scrambled movies (Fraedrich et al. 2010). The higher amount of motion energy in scrambled movies can explain the neural preference for scrambled stimuli especially prominent in the V1 data.

### Prediction Accuracy on Natural Versus Scrambled Movies

Scatter plots of average prediction accuracy on natural versus that on scrambled movies for each unit are shown in Figure 7C. For all 3 cortical areas, accuracy was higher for natural compared with scrambled movies: the difference in accuracy for natural minus that for scrambled was 0.10 for V1 (95% CI [0.09, 0.12],  $P < 0.001$ ), 0.07 for LI (95% CI [0.06, 0.09],  $P < 0.001$ ), and 0.05 for TO (95% CI [0.03, 0.06],  $P < 0.001$ ). Given the earlier finding that responses to natural movies are more reliable than responses to scrambled movies, this difference in prediction accuracy might be caused by the difference in reliability. Including reliability as a covariate still resulted in an estimated higher accuracy for natural movies of 0.07 for V1 (OLS estimate, 95% CI [0.06, 0.09],  $P < 0.001$ ), 0.05 for LI (OLS estimate, 95% CI [0.03, 0.09],  $P = 0.003$ ), and 0.05 for TO (OLS estimate, 95% CI [0.03, 0.07],  $P < 0.001$ ). Thus, the effect of scrambling on prediction accuracy does not seem to be the result of differences in response reliability.

## Discussion

We investigated the neural responses to natural movies and their phase-scrambled versions in rat V1 and 2 extrastriate visual areas LI and TO, which belong to a distinct pathway reminiscent of the primate ventral visual stream (Vermaercke et al. 2014).

First, we found an increased clustering of natural versus scrambled movie representations when progressing from V1 to TO. The increased dissociation of the 2 stimulus types correlates

with a decreased overall preference for scrambled stimuli in spite of the stronger motion energy contained in scrambled movies. A closer look at single cell preferences suggests that the population effect is driven by the increase in the proportion of cells preferring natural stimuli and by an increase in strength of preference for those neurons that prefer natural stimuli.

Second, unlike what one would expect to see in an object representation pathway such as the primate ventral visual stream (Orban 2008), the population representations of the stimulus set do not culminate into a higher level categorical representation in area LI or in the most downstream area TO. Of course, we are restricted in making strong claims about this by our small stimulus set and limited amount of neurons per area. However, the neurons used for our analyses were all responsive to at least one stimulus and did show selectivity, indicating that they did encode information. Furthermore, as far as we can judge from the available primate literature, the distinction between animate and nonanimate stimuli in the monkey and human brain is very clear. For example, the matrices from monkey and human data as shown by Kriegeskorte, Mur, Ruff et al. (2008, Fig. 1) suggest that this animate/inanimate distinction in primates is at least as clear as the distinction between natural and scrambled in our rat data. Given that we can easily pick up the natural versus scrambled distinction in our rat data, we think we should be able to pick up a similarly sized effect of rat (animate) versus nonrat (inanimate). An effect of this size does not seem to be present for our stimuli in the recorded neuronal populations. Nonetheless, we cannot exclude that factors we did not control for, such as attention or rather the lack thereof, might have influenced such findings. Nor can we exclude the possibility that other areas in the rat brain would show such category selectivity. However, it is not very obvious which other areas would do so.

Finally, a V1-like model that has previously been used to model receptive field properties of neurons (Nishimoto and Gallant 2011; Talebi and Baker 2012) as well as voxels in human brain imaging (Nishimoto et al. 2011) could predict responses of V1 neurons reasonably well, especially when we take into consideration that our experiment was not optimized for fitting such a model. Similar to what is reported in humans when comparing primary visual cortex with extrastriate areas (Nishimoto et al. 2011), prediction accuracy was reduced in LI and even more so in TO, suggesting that such a model was progressively less able to capture response properties of areas further along this visual stream. Of course, this is only one model that will not capture all possible tuning properties of V1 neurons, therefore caution should be taken in drawing strong conclusions from this piece of evidence alone.

The sampling bias of upper cortical layers for the V1 recordings could explain (some of) the differences that we observe between recordings in V1 on the one hand and recordings in LI and TO on the other hand. A previous systematic comparison between responses in upper and lower layers of V1 did not show consistent differences (Vermaercke et al. 2014). Likewise, Froudarakis et al. (2014) reported no differences between responses to natural movies in V1 layer 2/3 and V1 layer 4. Moreover, all the changes across the succession of areas that we do describe continue in the same direction from LI to TO, where there is no difference in layer sampling bias. Thus, we argue that a sampling bias of upper layers in V1 does not invalidate conclusions of gradual changes across the visual processing pathway under investigation.

Together, these findings support the idea of a functional hierarchy in these areas. The data suggest that an increasing number of neurons are driven by more complex stimulus features that are

not captured by V1-like filters and destroyed by a phase-scrambling method. Indeed, a linear combination of V1-like receptive fields decreases in efficiency in predicting neural responses the further up this hierarchy. Nevertheless, the functional hierarchy does not seem to culminate in neural representations as found in primates.

### Comparison with Previous Research on Rodent Extrastriate Visual Cortex

Previous research on the functional properties of rodent extrastriate areas LI and/or TO have only used drifting gratings (Marshel et al. 2011; Vermaercke et al. 2014) or simplistic artificial stimuli (Vermaercke et al. 2014).

The current study is the first that allows an investigation of selectivity in more downstream visual cortex for the complex features present during natural stimulation. The increased response to natural movies relative to scrambled movies and the decreased performance of a V1-like energy model significantly extend the earlier findings of position invariance and simple shape representations, and support the notion of an increasingly high-level stimulus representation when progressing from V1 to TO.

Nevertheless, we could not find any evidence for a category-selective representation in rat extrastriate cortex and to the best of our knowledge there is no other neural data supporting this notion. In a recent study, rats could be trained in a 2 alternative forced choice task to discriminate rat movies from nonrat movies and could generalize to new previously unseen exemplars (Vinken et al. 2014). However, the training was difficult and took a substantial amount of time, which is consistent with the absence of a categorical representation in naïve animals.

### Comparison with Previous Experiments Using Natural Stimuli

Previous studies using stationary natural and scrambled images indicate a preference for natural images in responses of human lateral occipital complex (Grill-Spector et al. 1998) and monkey inferior temporal cortex (Vogels 1999; Rainer et al. 2002). In areas earlier in the ventral visual processing hierarchy responses have been reported to show a preference for scrambled images in V1 that disappears in extrastriate visual areas (Rainer et al. 2002). The current study included a method of scrambling which falls within the range of methods used previously. Recent studies zoomed in on this general difference between intact and scrambled images by including specific methods of scrambling and focusing upon particular characteristics of natural images. For example, Freeman et al. (2013) showed in monkeys and in humans that responses in V2 were stronger for naturalistic textures than for spectrally matched noise, while they were the same in magnitude in V1.

In the current study, the change in preference for natural and scrambled is similar to all this earlier work when the preference is expressed in relative terms: more preference for natural movies when moving away from V1. We find a stronger firing rate in response to scrambled movies compared with their original natural counterparts in V1 and this difference decreases in extrastriate area LI and ends in an equal firing rate for both stimulus types in TO. Similarly, an fMRI study in humans has reported stronger early visual cortex activity to spatiotemporally phase-scrambled movies relative to their original version (Fraedrich et al. 2010). This result is supported by the motion energy model we used to show an increased output for scrambled movies when passed through a bank of V1-like filters, as well as by previous modeling

studies using still images (Stojanoski and Cusack 2014). In the study by Freeman et al. (2013) the reported equal firing rate for natural and scrambled images in V1 might be the result of control stimuli that are more carefully matched in spectral properties than is allowed by our spatiotemporal phase scrambling of the movies. These previous reports related to phase-scrambling combined with our own modeling results indicate a parallel between our experimental results and the earlier findings in primates: namely, a preference for scrambled stimuli that disappears in extrastriate cortex (Rainer et al. 2002). Similar to the present study, Rainer et al. (2002) used a scrambling method that introduced distortions that the earlier visual system is sensitive to (Stojanoski and Cusack 2014). However, in the present study this gradual change in stimulus type preference did not culminate in the higher response to natural stimuli that is observed in monkey inferior temporal cortex (Rainer et al. 2002) and human lateral occipital complex (Grill-Spector et al. 1998). This means that inasmuch as the succession of areas where we recorded can be compared with the primate ventral visual stream, we could not find support for a preference for natural stimuli typical of these primate higher visual areas, and find a resemblance with more mid-level areas, at best. Another recent study in rodents reports a stronger response to phase-scrambled movies in mouse V1 when the animal was sitting still and not whisking (Froudarakis et al. 2014), which is consistent with our findings. On the other hand, their recordings when the animals were whisking and/or running as well as their recordings in anaesthetized animals showed an equal response to natural movies and their phase-scrambled controls. These findings suggest that brain/behavioral state interacts with the effect of phase scrambling. Our rats were awake and passively viewing the stimuli during recordings, but we did not monitor behavioral cues such as whisking, so we cannot control for this. Overall, the animals tended to be sitting very still during the recordings. We speculate that behavioral state might increase the sensitivity for natural stimulation in rodent visual cortex overall, which then overcomes the difference in motion energy in V1. Several other comparisons in our report between natural and scrambled movies were also included in the investigation of V1 by Froudarakis et al. (2014), and for those indices the results tend to be consistent between the 2 studies. More specifically, we replicate a higher sparseness and more reliable responses for natural movies.

### Implications for the Rodent as a Model for Object Vision

What are the implications of our findings on the idea of the rodent as a model for object vision? The rodent has become a popular model in the neuroscience community for tackling questions on the topic of higher-level and object vision (Glickfeld et al. 2014; Cooke and Bear 2015). A large and consistent body of evidence exists in the behavioral literature revealing encouraging visual capabilities in rats (Zoccolan 2015). However, while steps have been taken to show the existence of 2 anatomically and functionally distinct streams in mouse visual cortex (Andermann et al. 2011; Marshel et al. 2011; Wang et al. 2012), it remains an open question whether and to what extent the rodent ventral visual stream can be considered as homologous to that of the primate. Here we show that the proposed homolog of the rat ventral visual stream may not show certain properties to the same extent as the primate ventral visual stream, such as higher responses to natural images, and even lack defining properties like a categorical representation. This story parallels previous findings that show both typical (an increase in tolerance for stimulus position), as well as atypical (an increased response to moving stimuli)

properties of the pathway (Vermaercke et al. 2014). Together, Vermaercke et al. (2014) and our results show that we should be cautious in assuming functional similarities in visual processing between rodents and primates. Perhaps, we should reconsider the concept of a ventral visual stream tuned for object recognition in rats and mice. After all, these are nonfoveal animals, with a very low visual acuity (Prusky et al. 2000), that might rely so much on their other senses for object recognition in natural situations that they lack the functional specialization in visual cortex. The situation is complicated further by the finer differentiation of this ventral stream in primates into multiple pathways (Kravitz et al. 2013), and it is unclear which pathway(s) might be present in rodents, if any.

## Conclusion

We recorded neural responses in areas belonging to a proposed rodent homolog of the primate ventral visual stream in order to investigate 2 hallmarks of high-level representations in primates: preference for intact versus scrambled stimuli and category-selective responses. We found that our results parallel changes in response strength to natural versus scrambled stimuli from primate primary visual cortex to early extrastriate visual areas. However, unlike in primate ventral visual stream, in our results we failed to find a preference for natural stimuli in most temporal visual area TO, nor did the targeted pathway lead to category-selective representations.

## Funding

This work was supported by the Fund for Scientific Research Flanders (PhD Fellowship of K.V. and grant numbers G.0819.11, and GOA3913N) and the KU Leuven Research Council (grant number GOA/12/008). Funding to pay the Open Access publication charges for this article was provided by Fonds voor Wetenschappelijk Onderzoek (FWO) Flanders.

## Notes

*Conflict of Interest.* None declared.

## References

- Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A*. 2:284–299.
- Alemi-Neissi A, Rosselli FB, Zoccolan D. 2013. Multifunctional shape processing in rats engaged in invariant visual object recognition. *J Neurosci*. 33:5939–5956.
- Andermann ML, Kerlin AM, Roumis DK, Glickfeld LL, Reid RC. 2011. Functional specialization of mouse higher visual cortical areas. *Neuron*. 72:1025–1039.
- Brooks DI, Ng KH, Buss EW, Marshall AT, Freeman JH, Wasserman EA. 2013. Categorization of photographic images by rats using shape-based image dimensions. *J Exp Psychol Anim Behav Process*. 39:85–92.
- Cooke SF, Bear MF. 2015. Visual recognition memory: a view from V1. *Curr Opin Neurobiol*. 35:57–65.
- David SV, Mesgarani N, Shamma SA. 2007. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network*. 18:191–212.
- DiCarlo JJ, Cox DD. 2007. Untangling invariant object recognition. *Trends Cogn Sci*. 11:333–341.
- Dicarlo JJ, Zoccolan D, Rust NC. 2012. How does the brain solve visual object recognition? *Neuron*. 73:415–434.
- Efron B. 1987. Better bootstrap confidence intervals. *J Am Stat Assoc*. 82:171.
- Espinoza SG, Thomas HC. 1983. Retinotopic organization of striate and extrastriate visual cortex in the hooded rat. *Brain Res*. 272:137–144.
- Fraedrich EM, Glasauer S, Flanagan VL. 2010. Spatiotemporal phase-scrambling increases visual cortex activity. *Neuroreport*. 21:596–600.
- Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA. 2013. A functional and perceptual signature of the second visual area in primates. *Nat Neurosci*. 16:974–981.
- Froudarakis E, Berens P, Ecker AS, Cotton RJ, Sinz FH, Yatsenko D, Saggau P, Bethge M, Tolias AS. 2014. Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat Neurosci*. 17:851–857.
- Girman SV, Sauvé Y, Lund RD. 1999. Receptive field properties of single neurons in rat primary visual cortex. *J Neurophysiol*. 82:301–311.
- Glickfeld LL, Reid RC, Andermann ML. 2014. A mouse model of higher visual cortical function. *Curr Opin Neurobiol*. 24:28–33.
- Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzhak Y, Malach R. 1998. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp*. 6:316–328.
- Huberman AD, Niell CM. 2011. What can mice tell us about how vision works? *Trends Neurosci*. 34:464–473.
- Issa EB, DiCarlo JJ. 2012. Precedence of the eye region in neural processing of faces. *J Neurosci*. 32:16666–16682.
- Kampa BM, Roth MM, Göbel W, Helmchen F. 2011. Representation of visual scenes by local neuronal populations in layer 2/3 of mouse visual cortex. *Front Neural Circuits*. 5:18.
- Kiani R, Esteky H, Mirpour K, Tanaka K. 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol*. 97:4296–4309.
- Kravitz DJ, Saleem KS, Baker CI, Mishkin M. 2011. A new neural framework for visuospatial processing. *Nat Rev Neurosci*. 12:217–230.
- Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci*. 17:26–49.
- Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2:1–28.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*. 60:1126–1141.
- Marshall JH, Garrett ME, Nauhaus I, Callaway EM. 2011. Functional specialization of seven mouse visual cortical areas. *Neuron*. 72:1040–1054.
- Mishkin M, Ungerleider LG. 1982. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res*. 6:57–77.
- Niell CM. 2011. Exploring the next frontier of mouse vision. *Neuron*. 72:889–892.
- Nishimoto S, Gallant JL. 2011. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci*. 31:14551–14564.
- Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr Biol*. 21:1641–1646.
- Orban GA. 2008. Higher order visual processing in macaque extrastriate cortex. *Physiol Rev*. 88:59–89.

- Prusky GT, Harker KT, Douglas RM, Whishaw IQ. 2002. Variation in visual acuity within pigmented, and between pigmented and albino rat strains. *Behav Brain Res.* 136:339–348.
- Prusky GT, West PW, Douglas RM. 2000. Behavioral assessment of visual acuity in mice and rats. *Vision Res.* 40:2201–2209.
- Rainer G, Augath M, Trinath T, Logothetis NK. 2002. The effect of image scrambling on visual cortical BOLD activity in the anesthetized monkey. *Neuroimage.* 16:607–616.
- Rosselli FB, Alemi A, Ansuini A, Zoccolan D. 2015. Object similarity affects the perceptual strategy underlying invariant visual object recognition in rats. *Front Neural Circuits.* 9:1–22.
- Stojanoski B, Cusack R. 2014. Time to wave good-bye to phase scrambling: creating controlled scrambled images using diffeomorphic transformations. *J Vis.* 14:1–16.
- Tafazoli S, Di Filippo A, Zoccolan D. 2012. Transformation-tolerant object recognition in rats revealed by visual priming. *J Neurosci.* 32:21–34.
- Talebi V, Baker CL. 2012. Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. *J Neurosci.* 32:1560–1576.
- Vermaercke B, Gerich FJ, Ytebrouck E, Arckens L, Op de Beeck HP, Van den Bergh G. 2014. Functional specialization in rat occipital and temporal visual cortex. *J Neurophysiol.* 112:1963–1983.
- Vermaercke B, Op de Beeck HP. 2012. A multivariate approach reveals the behavioral templates underlying visual discrimination in rats. *Curr Biol.* 22:50–55.
- Vinje WE, Gallant JL. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science.* 287:1273–1276.
- Vinken K, Vermaercke B, Op de Beeck HP. 2014. Visual categorization of natural movies by rats. *J Neurosci.* 34:10645–10658.
- Vogels R. 1999. Effect of image scrambling on inferior temporal cortical responses. *Neuroreport.* 10:1811–1816.
- Wang Q, Burkhalter A. 2007. Area map of mouse visual cortex. *J Comp Neurol.* 502:339–357.
- Wang Q, Sporns O, Burkhalter A. 2012. Network analysis of cortico-cortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J Neurosci.* 32:4386–4399.
- Zoccolan D. 2015. Invariant visual object recognition and shape processing in rats. *Behav Brain Res.* 285:10–33.
- Zoccolan D, Oertelt N, DiCarlo JJ, Cox DD. 2009. A rodent model for the study of invariant visual object recognition. *Proc Natl Acad Sci USA.* 106:8748–8753.