

Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons

Susheel Kumar¹, Peter Kaposvari^{1,2}, and Rufin Vogels¹

Abstract

■ Animals and humans learn statistical regularities that are embedded in sequences of stimuli. The neural mechanisms of such statistical learning are still poorly understood. Previous work in macaque inferior temporal (IT) cortex demonstrated suppressed spiking activity to visual images of a sequence in which the stimulus order was defined by transitional probabilities (labeled as “standard” sequence), compared with a sequence in which the stimulus order was random (“random” sequence). Here, we asked whether IT neurons encode the images of the standard sequence more accurately compared with images of the random sequence. Previous human fMRI studies in different sensory modalities also found a suppressed response to expected relative to unexpected stimuli but obtained various results regarding the effect of expectation on encoding,

with one study reporting an improved classification accuracy of expected stimuli despite the reduced activation level. We employed a linear classifier to decode image identity from the spiking responses of the recorded IT neurons. We found a greater decoding accuracy for images of the standard compared with the random sequence during the early part of the stimulus presentation, but further analyses suggested that this reflected the sustained, stimulus-selective activity from the previous stimulus of the sequence, which is typical for IT neurons. However, the peak decoding accuracy was lower for the standard compared with the random sequence, in line with the reduced response to the former compared with the latter images. These data suggest that macaque IT neurons represent less accurately predictable compared with unpredictable images. ■

INTRODUCTION

Animals are sensitive to temporal regularities in their visual environment. Behavioral studies have shown that mere exposure to sequences of visual stimuli is sufficient to learn statistical regularities embedded in these sequences (for a review, see Turk-Browne, 2012). Such extraction of statistical regularities by the animals is often referred to as “statistical learning” (Turk-Browne, 2012; Saffran, Aslin, & Newport, 1996). Spiking activity recordings in passively fixating monkeys showed that inferior temporal (IT) cortical neurons carry statistical learning signals after the animals were exposed to visual image sequences in which the statistical regularities were based on transitional probabilities (Kaposvari, Kumar, & Vogels, 2016). The stimulus set of that study consisted of three groups of five images each, defining three quintets of images. The order of the five stimuli within each quintet was fixed, but the quintets were presented repeatedly in a random order without any interruption. Thus, only transitional probabilities defined quintets of images. Postexposure recordings in IT showed an enhanced response to deviant stimuli that violated the exposed sequence. This response enhancement for unpredicted compared with predicted stimuli was also seen in single IT neurons after exposure to doublets (Ramachandran,

Meyer, & Olson, 2016; Meyer & Olson, 2011) or triplets (Meyer, Ramachandran, & Olson, 2014) that were followed by a reward and an intersequence interval.

By comparing IT responses with sequences with and without statistical regularities, Kaposvari et al. (2016) observed a response suppression for stimuli of a sequence with regularities (labeled “standard sequence”) compared with a “neutral” sequence in which images were presented in random order during exposure (labeled “random sequence”). This response difference between sequences with and without statistical regularity was caused neither by image familiarity/frequency nor by repetition suppression. Here, we ask whether the smaller response to the standard compared with the random sequence has any repercussion on the representation of the images of the sequences. In particular, we assessed how well one could decode the individual visual images from the population activity of the neurons recorded by Kaposvari et al. (2016). Intuitively, one would predict a lower classification accuracy for the standard compared with the random sequence stimuli, because the response was smaller for the former sequence. Contrary to this intuitive prediction, Kok, Jehee, and de Lange (2012) found in human V1, using fMRI multivoxel pattern analysis (MVPA), an increased classification accuracy for two grating orientations when these were expected compared with when they were unexpected. This enhanced encoding of expected grating orientations was present

¹KU Leuven, Belgium, ²University of Szeged, Hungary

despite a decreased activation for the expected gratings. These data were interpreted to suggest that expectation sharpens stimulus representations. A more recent human fMRI study (Blank & Davis, 2016) using speech stimuli reported also a decreased activation in the posterior STS for expected words compared with unexpected words but, contrary to Kok et al., reported a decreased decoding of the expected words. Interestingly, these opposite findings were both explained in terms of predictive coding theories (Feldman & Friston, 2010; Friston, 2005). The Kok et al. (2012) human fMRI data predict that the classification accuracy of IT neurons would be greater for the stimuli in the standard sequence relative to the random sequence stimuli, because the latter are unpredictable whereas the former are predictable. The Blank and Davis study makes the opposite prediction. In the present work, we examined these predictions by decoding image identity of the two sequences from the neural responses, using linear support vector machine (Cortes & Vapnik, 1995) classifiers, and assessed the time course of the classification of image identity during stimulus presentation for the two sequence types separately.

METHODS

The data were collected in Experiment 2 of Kaposvari et al. (2016). A detailed description of apparatus, recording, and experimental procedures can be found in that study. Here, we briefly describe the experimental paradigm.

Subjects and Recording Location

Data were collected from two male rhesus monkeys (H and O; *Macaca mulatta*). All animal care and experimental protocols complied with national and European guidelines and were approved by the KU Leuven ethics committee for animal experiments. Multiunit activity (MUA) recordings were performed in the ventral bank of the rostral STS of the right hemisphere.

Fixation Task

The animals were required to fixate within a 2° square fixation window, centered around a small fixation target (red color; 0.13°), to obtain a juice reward. Juice rewards were given with decreasing intervals as long as the monkeys maintained fixation, encouraging long fixation. Importantly, the timing of the juice delivery and the presentation of the stimulus sequences were uncorrelated. The fixation target was located at the center of the display, superimposed on the center of mass of the stimuli.

Experimental Design

The stimuli consisted of two groups of 15 stimuli each (Figure 1C). Each group consisted of modified Snodgrass and Vanderwart images of animals and objects, taken

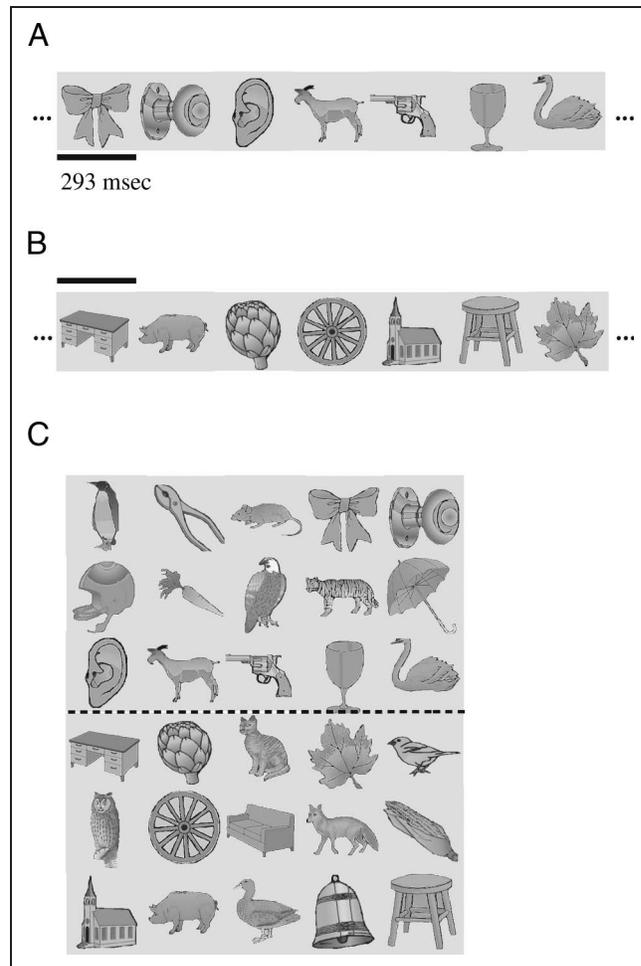


Figure 1. Stimuli and sequences. (A) The “standard sequence” consisted of a continuous presentation, without ISI, of three quintets. The five stimuli of a quintet were shown in a fixed order, but the order of the three quintets was random. Stimulus duration was 293 msec. (B) In the “random sequence,” 15 stimuli were shown in random order. Stimulus presentation parameters were identical with those in the standard sequence. (C) The two groups of 15 stimuli were employed in the study. The order of the five stimuli of each row corresponds to that in the exposed quintet. Note that the top and bottom groups of 15 stimuli were employed in the standard and random sequences, respectively, in Monkey H, whereas the opposite assignment of stimulus group to sequence type was present in Monkey O.

from the Rossion and Pourtois (2004) database. The stimuli were presented on a gray background. We resized the images so that their maximal horizontal or vertical extent was 6° and equated their mean luminance. One group of 15 images was sorted into groups of three quintets. We equated as much as possible the low-level image properties between successive presentations within and across quintets (see Kaposvari et al., 2016). During the exposure phase of the experiment, the selected three quintets were shown in random order, but with a fixed stimulus order within a quintet. We will label these sequences as “standard sequences.” An individual image was shown for 293 msec and was immediately succeeded by the next image, without any ISI or interquintet interval (Figure 1A).

The other group of stimuli consisted of 15 other animals and objects, and these were presented for 293 msec each in pseudorandom order, without any ISI (Figure 1B). These sequences will be labeled as “random.” We required that the number of stimuli in between presentations of the same stimulus should be at least four. This ensured that the average interstimulus presentation interval was equal for both the standard and random sequences. The stimulus groups presented in standard or random sequences were counterbalanced across the two animals.

We exposed the animals to both types of sequences using a block design. Each block consisted of 4050 stimuli (270 presentations per stimulus), lasting approximately 20 min, and standard and random sequence blocks were alternated in daily sessions of approximately 2 hr. The sequence type of the first block of a daily session was randomized across sessions. We kept track of the number of presentations per sequence type and, when necessary, increased temporally the number of presentations for a particular type to equate the number of presentations per sequence type. Thus, we ensured that the stimuli of the standard quintets and random sequences had equal familiarity. The exposure phase in Experiment 2 lasted 34 and 35 daily sessions in Monkeys H and O, respectively.

After the exposure phase, we recorded MUA for blocks of standard and random sequences. We searched for responsive MUA using either standard or random sequences, alternating between MUA sites. This avoided biasing responses toward one or the other sequence type. The data were obtained in two phases, described briefly below (for more details, see Kaposvari et al., 2016). In the first phase, the MUA sites were tested with blocks containing only standard or random sequences. We performed 16 and 17 daily recording sessions in Monkeys H and O, respectively, in the first phase. We tested each responsive MUA site with two blocks of each sequence type and the two block types alternated. The sequence type of the first block was randomized across sites. Each block contained 70 presentations per stimulus. We kept track of the number of presentations per sequence type, in particular, during the search periods in which a single sequence type was presented, and when necessary, compensated by presenting the less frequent stimuli more in later sessions.

In the second phase, the MUA sites were tested with blocks of random sequences and blocks that were composed of standard sequences with one fourth of the quintets containing a deviant, that is, a stimulus that did not belong to that quintet. The quintets with deviants had one image of another quintet inserted. As in the first phase, a responsive MUA was searched using either standard sequences with quintets without a deviant or random sequences, and this was done with an equal frequency. When testing the MUA sites, we alternated blocks of standard and deviant quintets and blocks of random sequences. Because of the inclusion of deviants, we increased the duration of the blocks containing quin-

tets with deviants by presenting each stimulus 480 times per block. The number of presentations per stimulus was 140 for the random block. The data for standard stimuli used in all the analyses of this study came from quintets without deviants. MUA to both types of sequences in the second phase was measured in 17 and 19 daily recording sessions in Monkeys H and O, respectively.

Data Analysis

In total, 119 MUA sites (58 and 61 for Monkeys H and O, respectively) were recorded in the first phase of recordings, and 62 MUA sites (32 and 30 for Monkeys H and O, respectively) were recorded in the second phase. For all analyses, we included only those stimulus presentations during which the monkey was fixating at the fixation target. Response to a stimulus in standard and random sequences was considered only when the stimulus was preceded by at least five images and succeeded by at least two images during a fixation period. As was previously shown by Kaposvari et al. (2016) that there was no significant difference among the responses to the last three stimuli of a standard quintet, we chose the last three stimuli (nine stimuli in total from three quintets) for further analyses and for comparison with the stimuli of the random sequence. We will label these nine standard sequence stimuli as “selected standard” stimuli. The sites containing responses to a minimum of 10 presentations of each of the stimuli were selected. This selection criterion yielded 116 MUA sites (55 and 61 for Monkeys H and O, respectively) of the first phase and all 62 sites of the second testing phase. To compute population peristimulus time histograms, we averaged first across responses to a stimulus for a site, and then, these mean responses were averaged across sites. The number of observations in the statistical tests corresponded to the number of MUA sites.

Decoding Analysis Methods

The population decoding analyses were performed in MATLAB (The MathWorks, Natick, MA) using Neural Decoding Toolbox (Meyers, 2013). We employed a support vector machine classifier with a linear kernel. A fivefold cross-validation scheme was used, where 80% of the stimulus presentations were used for training and the remaining 20% were used for testing. Reported classification accuracies were for the independent test data only. We used an “all-pair” multiclass classification scheme (Meyers, 2013). The neural responses were z-score normalized with the mean and standard deviations of the training data.

We created pseudopopulation response vectors using a selected set of n MUA sites. For each stimulus label and MUA site, the responses for 10 stimulus presentations were randomly assigned to 10 pseudopopulation vectors (length = n sites). The 10 presentations were randomly picked from all the presentations of a stimulus.

For each MUA site, a random selection of eight pseudopopulation response vectors per stimulus label constituted the training data set, whereas the remaining two vectors per label defined the test set. The nine stimulus labels for the standard sequence corresponded to the nine “selected standard” stimuli (third, fourth, and fifth stimuli of each of the three standard quintets), whereas 9 from the 15 stimuli of the random sequences were selected at random to train and test the classifiers. Training and testing were done using fivefold cross-validation.

In the main analysis, we randomly sampled $n = 108$ MUA sites (approximately 60%) from the 178 MUA sites (116 from the first phase and 62 from the second phase). This random selection of 108 sites was done 1000 times, and for each of these 1000 samples of sites, we performed the same population decoding of the nine standard and nine random stimuli, as described above. The data used for training and testing consisted of firing rates in 20-msec bins sampled at 20-msec intervals starting at stimulus onset sampled until 107 msec after stimulus offset. For each of the 1000 selection of sites, the decoding procedure was run 20 times, each time creating new pseudopopulation vectors and performing the fivefold cross-validation. For each of the 1000 site selections, the classification scores were averaged across the 20 resamples. Because we classified both random and standard sequence stimuli for each of the 1000 site selections, we could compute for each selection the difference between the mean classification scores for the two sequences. Statistical significance of the difference in mean classification scores for the two sequences was based on the percentile of a zero difference in the distribution of the 1000 difference scores (percentiles of <0.025 and >0.975 considered significant).

In addition, we performed the following control analyses. First, we decoded stimulus labels using the responses of the full population of 178 sites, using the nine selected stimuli of the standard sequence and a single random selection of nine stimuli of the random sequences. In this analysis, 100 instead of 20 resamplings were performed, and classification scores are the averages of the 100 resamplings. Second, we performed the same classification analyses with 100 resamplings using the data of each individual animal. Third, to check whether the classification accuracies depended on the selection of the nine random sequence stimuli, we repeated the decoding procedure 50 times, each time randomly selecting 9 from the 15 random sequence stimuli. In each of the 50 procedures, the number of pseudopopulation vector resamples was 20, and we employed the full population ($n = 178$ sites). Fourth, in the second phase of the recordings, the stimuli of the standard sequences were presented more often than the random sequence stimuli, which might have produced a difference in familiarity between the stimuli of the two sequence types. Because the number of presentations per stimulus was very high, it is unlikely that the effects we report result from this difference in

stimulus frequency. Nonetheless, we performed all the above decoding analyses using only the data from the 116 MUA sites of the first phase, in which stimulus frequency of standard and random sequence stimuli was equated. The results (data not shown) were qualitatively identical to those of the data of both phases combined.

RESULTS

We exposed monkeys to two types of sequences of visual images. One type of sequence, the standard sequence, consisted of three quintets of images. The order of presentation of the five images was fixed in each quintet, but the three quintets were presented in random order. The animals can predict the next stimulus of the standard sequence based on the previous stimulus (except for the first stimulus of a quintet). In the second type of sequence, the random sequence, 15 other images were presented in (pseudo)random order so that the presentation of a particular stimulus could not be predicted from the previous ones. Kaposvari et al. (2016) showed that IT neurons responded with higher firing rates to the stimuli of the random compared with those of the standard sequence. This “expectation suppression” effect is illustrated in Figure 2A for 108 randomly selected MUA sites in IT of the two monkeys (the same number of MUA sites was employed for the classification analysis below). The response for the standard sequence is the mean of the responses to the last three stimuli of each quintet (nine stimuli in total). These stimuli were selected because Kaposvari et al. (2016) found that the response was higher for the first stimulus of a quintet than for the last three stimuli of a quintet, likely reflecting the low transition probability (1/3) associated with the first stimulus of a quintet. Furthermore, in one animal, the second stimulus of a quintet also produced a higher response compared with the three later stimuli of the quintet. The response to the random sequence was averaged across nine randomly selected stimuli of that sequence. Note that the stimuli that were presented in the random and standard sequences were counterbalanced across the two monkeys. As reported by Kaposvari et al. (2016) for all recorded MUA sites ($n = 178$) of this experiment, this random selection of MUA sites showed a stronger transient response to the random compared with the standard sequence stimuli. In the present work, we asked whether neurons downstream from IT can decode with a greater accuracy the standard compared with the random sequence stimuli.

In the first decoding analysis, we classified the label of nine stimuli (the last three of each quintet) of the standard sequence and of nine randomly selected stimuli of the random sequence. In this analysis, we employed only 108 sites and not all 178 sites, because this allowed performing randomization-based statistical tests of condition effects. We pooled the MUA sites across animals and drew at random 108 sites from the 178 recorded MUA

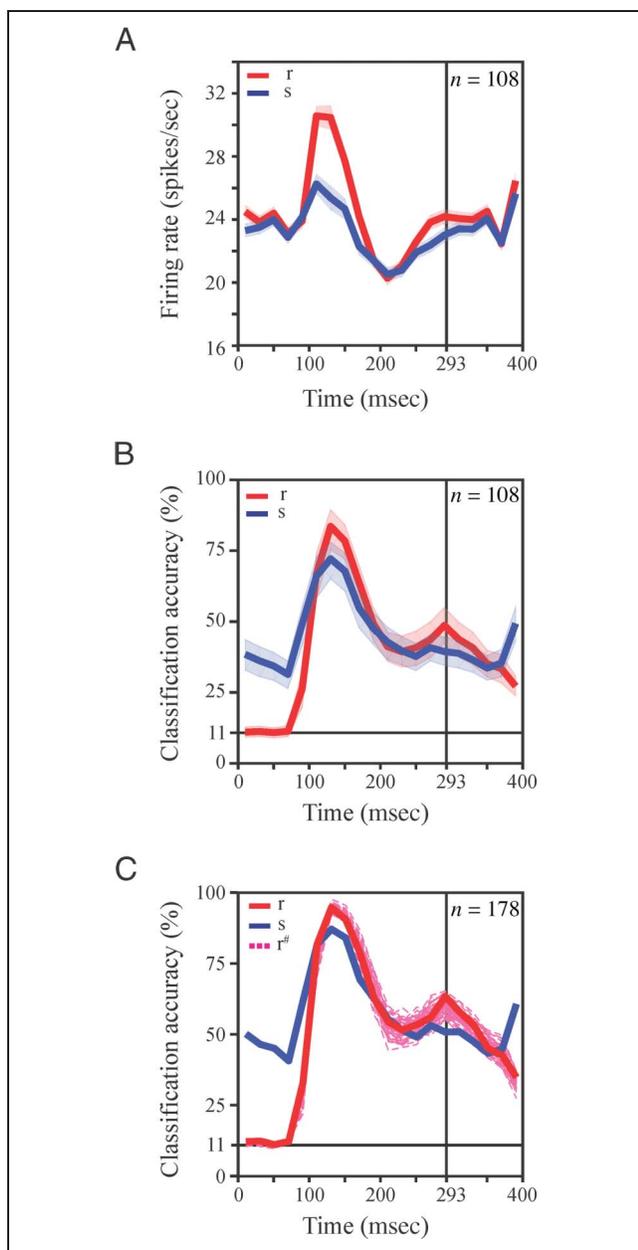


Figure 2. Responses and classification accuracy for standard and random sequence stimuli. (A) Mean firing rate to random (r) and standard (s) sequence stimuli from 108 sites that were randomly chosen from the 178 MUA sites of both animals. The shaded bands indicate standard error of the mean, computed following the procedure by Loftus and Masson (1994). Time 0 corresponds to stimulus onset, and stimulus offset is indicated by the vertical line. (B) Classification accuracy for random and standard stimulus sequences. For each of 1000 decoding runs, responses of 108 sites were randomly chosen from 178 sites. The lines indicate the mean across the 1000 runs, whereas the shaded bands indicate 95% confidence intervals based on percentiles of the distribution of the 1000 classification scores. The horizontal line corresponds to chance level performance. Same conventions as in A. (C) Mean classification accuracy for random (r) and standard (s) sequence stimuli when decoding was performed using all MUA 178 sites (100 resamplings). The dashed “r#” lines correspond to the mean classification accuracies (20 resamplings) obtained when each time sampling 9 from the 15 random sequence stimuli, showing the variability due to stimulus differences. Same conventions as in B. Bin width in all panels was 20 msec, and no smoothing is present.

sites. This process was performed 1000 times to get bootstrapped data. We classified the images for each of the 1000 randomly drawn samples. The time course of average classification accuracy for the images of the two sequences is shown in Figure 2B. The bands indicate 95% confidence intervals based on the 1000 bootstrapped data. The classification accuracy during stimulus presentation was well above chance level (11.1%) and peaked for both sequences at 130 msec after stimulus onset. The time of the peak classification accuracy coincided with the time of the peak firing rate (compare Figure 2A and B). The classification accuracy for the standard sequence stimuli was well above chance throughout the stimulus presentation, even before response onset (at about 90 msec). This is because, in Figure 2B, the classifier was trained and tested at the same time bin and the stimulus sequence of a quintet was fixed (and thus stimulus labels were correlated within a quintet). The classification accuracy for the random sequence stimuli started at chance and then rose above chance after 90 msec, matching the time course of the response. Importantly, the peak classification accuracy was significantly higher for the random compared with standard sequence (randomization test [see Methods]: $p = .003$; analysis bin: 120–140 msec). In fact, except for the initial period before and immediately after response onset, the classification scores for the standard sequence were below or equal to that of the random sequence, suggesting poorer stimulus identification for predictable compared with unpredictable stimuli.

In a subsequent analysis, we trained and tested the classifier at different, nonoverlapping time bins. This allowed us to assess whether and when the responses to the previous stimulus are carried over to the response pattern for the next stimulus. For the standard sequence, training the classifier at bins before 80 msec did not produce significant decoding above chance (based on 95% confidence intervals) when the classifier was tested at bins later than 140 msec (Figure 3A). This lack of generalization of the classifier from early (before response onset) bins to later bins in the standard sequences does not result from a nonstationary stimulus code during stimulus presentation, because there was above-chance classification between 100 and 380 msec when trained and tested bins differed (Figure 3A). For the random sequence stimuli (Figure 3B), generalization of decoding from trained to tested bins was present between 100 and 400 msec. This generalization for the random and standard sequences after stimulus offset (i.e., after 293 msec) and for the standard sequence before stimulus onset fits previous findings that the stimulus-selective response of IT neurons outlasts the stimulus duration for about 160 msec for stimuli presented without interstimulus time interval (De Baene, Premereur, & Vogels, 2007; Keyser, Xiao, Földiák, & Perrett, 2001). The difference between the classification scores for all the trained–tested pairs is shown in Figure 3C. Regions with significant bins (randomization

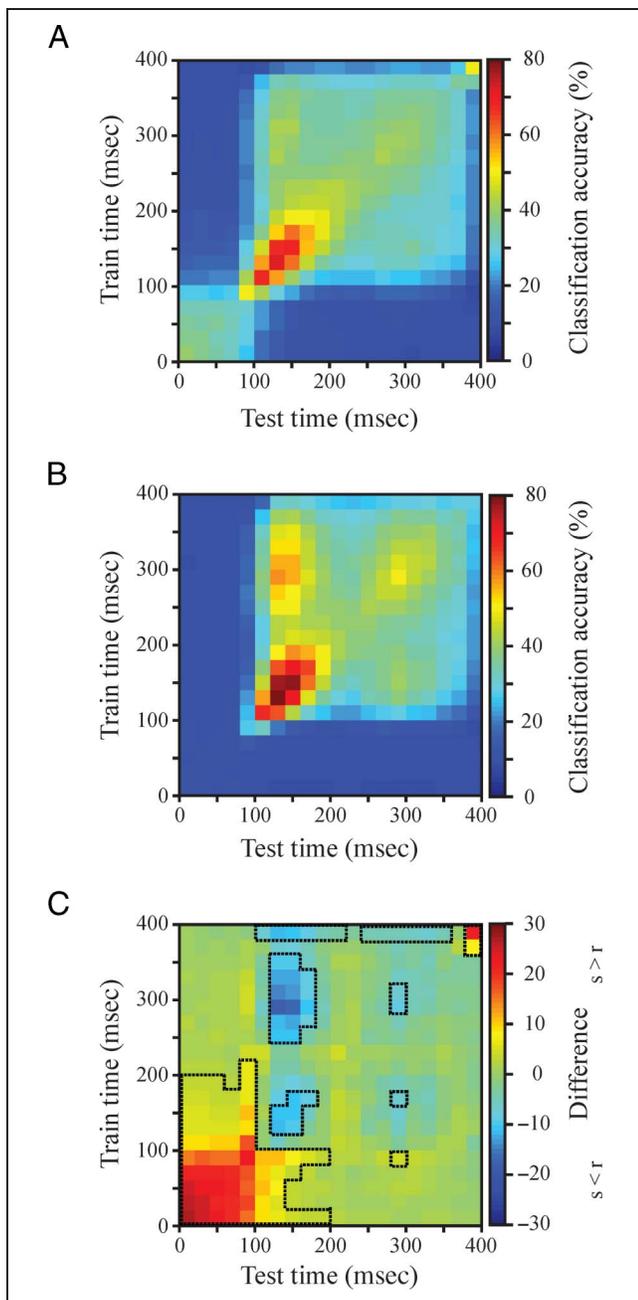


Figure 3. Classification accuracy as a function of the training and testing time bin, obtained when employing 108 randomly sampled MUA sites. Each row corresponds to the time bin (width = 20 msec) that was employed to train the classifier. The columns correspond to the time bins that were employed to test the classifier that was trained using the binned data indicated by a row. The main diagonal corresponds to the classification accuracies obtained when training and testing bins were identical. The classification scores are indicated by color (see legend to the right of each panel). The classification scores are the mean of 1000 decoding runs, using each time a novel random sample of 108 sites from the 178 available sites. (A) Mean classification accuracies for standard sequence stimuli. (B) Mean classification accuracies for random sequence stimuli. (C) Difference between classification accuracies for the standard and random stimulus sequences (standard [s] – random [r]). The boxes with stippled lines indicate the time bins where the difference was significant (randomization test; $p < .05$).

test; two-tailed $p < .05$) are indicated by stippled black outlines. When training or testing was performed in the 100-msec period before the response onset, the classification scores were significantly greater for the standard compared with the random sequence stimuli. There was a weakly enhanced classification for the standard sequence when training was performed between 100 and 200 msec and testing was before 100 msec or when training was before 100 msec and testing was between 100 and 200 msec. This is because of the sustained prolonged activity to the preceding stimulus (see above) and reflects decoding of the previous stimulus. Such decoding of the previous stimulus is only possible for the standard sequence, because only in that sequence the order of the stimuli (within a quintet) was fixed.

During the response phase of the current stimulus, that is, after 100 msec, the classification scores were higher for the random compared with the standard sequence, a difference reaching significance in several 20-msec-long bins (randomization test; stippled boxes in Figure 3C for trained–tested bins >100 msec). The only exception was a reversal of the difference in classification scores (greater for the standard sequence) for the late 380- to 400-msec bin. This likely reflects decoding of the next stimulus of the standard sequence where the order of the stimuli was fixed within a quintet. This interpretation is supported by the stronger decrease in generalization for the standard compared with the random sequence when the testing bin preceded the 380- to 400-msec training bin. Overall, the above analyses of the time course of decoding for the two sequences support a higher classification accuracy for random compared with standard sequence stimuli. The apparent higher classification accuracy for the standard sequence during the early part of the stimulus presentation reflects the sustained, stimulus-selective activity from the previous stimulus.

We performed several control analyses to assess the generality of the above reported findings. First, we trained classifiers using the whole population of 178 MUA sites. Except for the expected overall increase in classification accuracy, the results were highly similar with those obtained by the smaller sample of sites analyzed above (Figures 2C and 4A, D, and G). Second, in the above analysis, only one (randomly selected) sample of nine stimuli from the 15 random sequence stimuli was employed. To assess whether the results obtained above were specific to that sample of nine random sequence stimuli, we re-sampled 50 times nine random sequence stimuli and performed the classification procedure for each of these 50 sampled stimulus sets. As shown in Figure 2C, the range in classification accuracy due to stimulus variability was relatively small and could not explain the enhanced encoding for the random relative to the standard sequence. Finally, we repeated the classification analyses for the data of each individual animal, showing that similar effects were present in each animal (Figure 4). The stimulus resampling

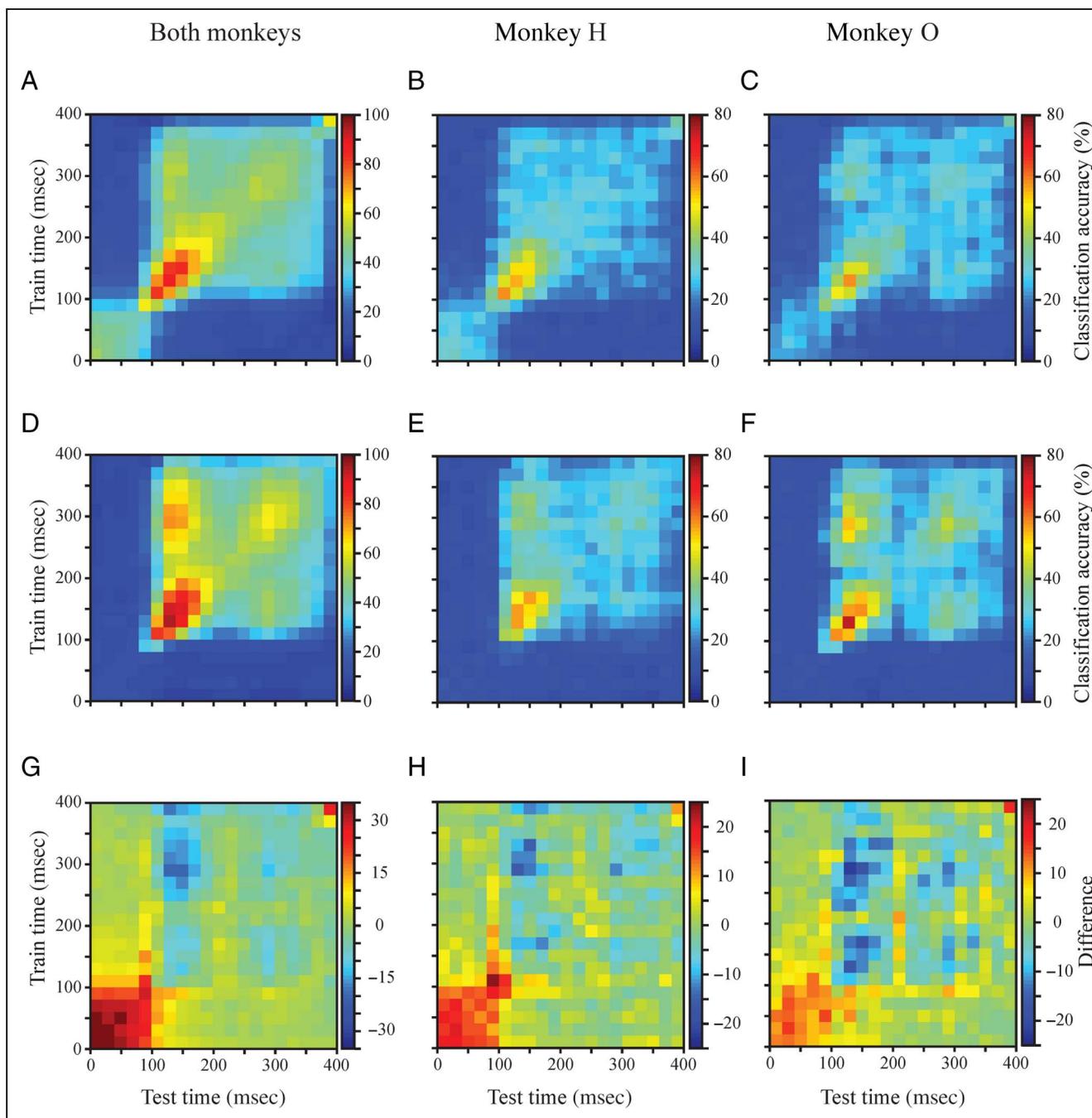


Figure 4. Classification accuracy as a function of the training and testing time bin, obtained when employing all 178 MUA sites (A, D, and G), the MUA data of Monkey H (87 MUA sites; B, E, and H), and the MUA data of Monkey O (91 sites; C, F, and I). Plot classification accuracies for (A–C) standard sequence stimuli and (D–F) random sequence stimuli, and (G–I) plot difference between standard and random sequence stimuli classification accuracies. The mean classification accuracies are averages of 100 resamplings. Same conventions as in Figure 3.

and the individual monkey analyses reassure that the effects we report here are not due to stimulus differences between the two sequence types but are related to differences between the sequences per se.

DISCUSSION

fMRI studies in humans (Kok et al., 2012; Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; den Ouden, Friston, Daw,

McIntosh, & Stephan, 2008) and electrophysiological studies in nonhuman primates (Kaposvari et al., 2016; Ramachandran et al., 2016; Meyer et al., 2014; Meyer & Olson, 2011) showed a decreased activity to predictable visual stimuli in primary visual cortex and IT cortex. Despite the decreased activity to an expected grating orientation, Kok et al. (2012) found an increased encoding of orientation in human V1 using MVPA of fMRI BOLD responses. Using a statistical learning design (Kaposvari

et al., 2016), we found an increased image encoding by IT neurons for unpredictable stimuli in a random sequence compared with predictable stimuli in sequence of fixed stimuli. Although, during the initial part of the stimulus presentation, classification was higher for the predictable compared with unpredictable stimuli, this difference could be explained by the well-known prolonged sustained response of IT neurons and does not require other expectation-based mechanisms.

Although we do not have behavioral evidence that the monkeys learned the transitional probabilities of the standard sequence, we do know that their IT neurons carry predictive signals. Kaposvari et al. (2016) showed for the same MUA sites and stimulus sequences in the same animals an enhanced response to a stimulus that was presented at the wrong position inside a quintet (“deviant”) and a decreased response to the standard compared with the random sequence, two signatures of prediction-related neural responses (also see Figure 2A). Furthermore, we took in the present analysis only the three last stimuli of each quintet because, for these stimuli, predictive signals were present (Kaposvari et al., 2016). Despite this evidence for predictive signals in these animals’ IT neurons, no evidence for an increased encoding of the predictable stimuli was present. To the contrary, we found evidence for a decreased encoding of the predictable stimuli.

The reason(s) of the apparent discrepancy between Kok et al. (2012) and our macaque data is difficult to pinpoint because of the many differences between the two studies. First, we employed MUA spiking activity while they used BOLD responses, which have a much coarser spatial resolution. In fact, it is not clear which neural properties drive MVPA orientation signals in human V1 (Pratte, Sy, Swisher, & Tong, 2016; Freeman, Brouwer, Heeger, & Merriam, 2011; Alink et al., 2010). Previous studies showed that classification analyses of MUA can be employed as a proxy to classification of single-unit responses in IT (Yamins et al., 2014), although the relationship between BOLD fMRI MVPA and single-unit selectivity is not straightforward (Dubois, de Berker, & Tsao, 2015). Second, the fMRI MVPAs have information about the across-presentation correlations among the simultaneously measured activations of the voxels, although across-presentation correlations in spiking activity (“noise correlations”) were not included in our decoding analyses. However, we believe that this is unlikely to explain the discrepancy between the Kok et al. (2012) study and our results. Spiking activity correlations occur in the 100- to 1000-msec range (e.g., Arandia-Romero, Tanabe, Drugowitsch, Kohn, & Moreno-Bote, 2016; Engel et al., 2016; Bair, Zohary, & Newsome, 2001), whereas that for BOLD is in the 10-sec range (e.g., Li, Bentley, Snyder, Raichle, & Snyder, 2015; peak frequency of correlations = 0.06 Hz). Given the long time constant of the hemodynamic response function, compared with the shorter time constant of the dominant noise correlations of spik-

ing activity, we argue that, basically, fMRI MVPA is more comparable to a multivariate analysis of spike counts that ignores spike correlations, like we did, than one that includes noise correlations. Furthermore, recent theoretical work (reviewed in Kohn, Coen-Cagli, Kanitscheider, & Pouget, 2016) emphasized that details of the response covariance matrix are critical for evaluating the effect of correlations on population decoding, and such details are lost in MVPA because of the low spatial resolution of fMRI.

Third, Kok et al. (2012) compared expected and unexpected, deviant stimuli when the subjects had the same expectation of a stimulus, whereas we compared predictable stimuli with stimuli that were shown in random order and thus for which no or perhaps a weak (transitional probability = 1/15) expectation was present. Kaposvari et al. (2016) showed that unpredicted, deviant stimuli produce an enhanced “surprise” response with respect to the neutral, random sequence stimuli, whereas predictable stimuli are suppressed relative to the random sequence condition. The latter expectation suppression had also a different time course compared with the surprise response to deviants. One might conjecture that the stimulus selectivity of the enhanced response to surprising stimuli is less than for expected stimuli, thus explaining the Kok et al. (2012) finding. However, Meyer and Olson (2011) showed a higher discrimination of unpredicted, surprising stimuli compared with predicted stimuli in monkey IT, which runs counter to the Kok et al. (2012) finding in human V1 and the above conjecture. Meyer and Olson (2011) found that deviant-induced prediction effects in IT scaled with stimulus preference, whereas the Kok et al. (2012) study found that the suppressive effect was strongest in voxels that least preferred the stimulus. The latter subtractive effect leads to a sharpened representation, which may explain the increased encoding for expected stimuli observed in the Kok et al. study. Thus, potential differences between areas in multiplicative versus additive expectation effects may explain the discrepancy between the monkey IT and fMRI V1 data. Fourth, the subjects in both our study and the Meyer and Olson monkey study (Ramachandran et al., 2016; Meyer et al., 2014; Meyer & Olson, 2011) were exposed to the sequences for several weeks before recordings, whereas in the Kok et al. (2012) study, expectation was manipulated within a single session. In addition, the subjects in Kok et al. (2012) were engaged in a discrimination of the stimuli, whereas our monkeys and those in the Meyer and Olson study were passively fixating. Thus, top-down processes are likely to have been fundamentally different in the monkey studies compared with the subjects in the Kok et al. (2012) study. In fact, the effects observed by Kok et al. (2012) may be more related to feature-based attention, which is known to result in sharpened stimulus representations (Maunsell & Treue, 2006), instead of expectation per se. Whichever the reason(s) of the discrepancy between

the different findings, the Kok et al. and Meyer and Olson studies compared surprising with expected stimuli. However, we examined the effect of expectation suppression on image classification, which may well be different from the effect of surprise on classification. In fact, our results are in line with a recent human fMRI (Blank & Davis, 2016) that showed a decreased decoding in the posterior STS of expected auditory speech stimuli when these were primed by a visual presentation of the same word compared with auditory stimuli after a nonsense word, similar to our neutral, random condition. Interestingly, an enhanced classification was present when the primed speech stimuli were acoustically degraded. These findings were, as those of Kok et al. (2012), interpreted in terms of predictive coding, but alternative accounts are possible, for example, feature-based attention in the case of the strongly degraded stimuli and cross-modal (visual–auditory) repetition suppression in the case of the less degraded auditory stimuli because the visual prime and primed auditory words were identical in the Blank and Davis study.

Computational studies have shown that the effect of noise correlations on population coding can depend on many factors, including the pattern of signal and noise correlations and their readout (Kohn et al., 2016; Averbeck, Latham, & Pouget, 2006). Given that we did not measure the response covariance matrix and are ignorant about its readout, we cannot make claims about how well (i.e., quantitatively) the brain can decode the IT population responses. Chen, Lin, Hsu, and Hung (2015) assessed the effect of noise correlation in a small population of IT neurons (up to 87 neurons) on object decoding with a linear classifier. They found that noise correlations slightly decreased decoding, which is in line with their, on average, positive signal correlations. Thus, our decoding performance might be an overestimation. Our results and conclusion hold under the assumption that the response covariance structure is not affected by predictability. In this regard, our decoding analysis should be viewed as an assessment of the effect of predictability on the overall stimulus selectivity of a population of IT neurons.

The reduced classification that we observed for the standard sequence compared with the random sequence stimuli fits the decreased response for the former compared with the latter stimuli. This effect of expectation suppression on classification is similar to the effect of repetition suppression on classification. Although fundamentally different neural mechanisms very likely underlie expectation suppression and repetition suppression (Vogels, 2016), the net functional effect could be the same: both resulting in decreased responses and classification accuracy (Kaliukhovich, De Baene, & Vogels, 2013). From a metabolic perspective, it makes sense that an object recognition system devotes less energy to processing a stimulus that is the same as a recently presented one (repetition suppression) or predicted by a preceding one (expectation suppression).

Acknowledgments

We thank P. Kayenbergh, G. Meulemans, I. Puttemans, Christophe Ulens, M. De Paep, W. Depuydt, and S. Verstraeten for technical assistance. This work was supported by Fonds voor Wetenschappelijk Onderzoek Vlaanderen (G.0582.12N and G.00007.12-Odyseus), Interuniversitaire Attractiepool and Programma Financiering (PF 10/008), and the European Community's Seventh Framework Programme FP7/2007-2013 under Grant agreement number PITN-GA-2008-290011 (ABC).

Reprint requests should be sent to Rufin Vogels, Neurofysiologie, Campus Gasthuisberg, Herestraat, Leuven, Belgium 3000, or via e-mail: rufin.vogels@kuleuven.be.

REFERENCES

- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, *30*, 2960–2966.
- Arandia-Romero, I., Tanabe, S., Drugowitsch, J., Kohn, A., & Moreno-Bote, R. (2016). Multiplicative and additive modulation of neuronal tuning with population activity affects encoded information. *Neuron*, *89*, 1305–1316.
- Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*, 358–366.
- Bair, W., Zohary, E., & Newsome, W. T. (2001). Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *Journal of Neuroscience*, *21*, 1676–1697.
- Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biology*, *14*, e1002577.
- Chen, Y. P., Lin, C. P., Hsu, Y. C., & Hung, C. P. (2015). Network anisotropy trumps noise for efficient object coding in macaque inferior temporal cortex. *Journal of Neuroscience*, *35*, 9889–9899.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.
- De Baene, W., Premereur, E., & Vogels, R. (2007). Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *Journal of Neurophysiology*, *97*, 2900–2916.
- den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2008). A dual role for prediction error in associative learning. *Cerebral Cortex*, *19*, 1175–1185.
- Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *Journal of Neuroscience*, *35*, 2791–2802.
- Engel, T. A., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., & Boahen, K. (2016). Selective modulation of cortical state during spatial attention. *Science*, *354*, 1140–1144.
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215.
- Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *Journal of Neuroscience*, *31*, 4792–4804.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *360*, 815–836.
- Kaliukhovich, D. A., De Baene, W., & Vogels, R. (2013). Effect of adaptation on object representation accuracy in macaque inferior temporal cortex. *Journal of Cognitive Neuroscience*, *25*, 777–789.
- Kaposvari, P., Kumar, S., & Vogels, R. (2016). Statistical learning signals in macaque inferior temporal cortex. *Cerebral Cortex*. doi:10.1093/cercor/bhw374.

- Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, *13*, 90–101.
- Kohn, A., Coen-Cagli, R., Kanitscheider, I., & Pouget, A. (2016). Correlations and neuronal population information. *Annual Review of Neuroscience*, *39*, 237–256.
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*, 265–270.
- Li, J. M., Bentley, W. J., Snyder, A. Z., Raichle, M. E., & Snyder, L. H. (2015). Functional connectivity arises from a slow rhythmic mechanism. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, E2527–E2535.
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, *1*, 476–490.
- Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*, 317–322.
- Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 19401–19406.
- Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *Journal of Neuroscience*, *34*, 9332–9337.
- Meyers, E. M. (2013). The neural decoding toolbox. *Frontiers in Neuroinformatics*, *7*, 8.
- Pratte, M. S., Sy, J. L., Swisher, J. D., & Tong, F. (2016). Radial bias is not necessary for orientation decoding. *NeuroImage*, *127*, 23–33.
- Ramachandran, S., Meyer, T., & Olson, C. R. (2016). Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. *Journal of Neurophysiology*, *115*, 355–362.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Turk-Browne, N. B. (2012). Statistical learning and its consequences. *Nebraska Symposium on Motivation*, *59*, 117–146.
- Vogels, R. (2016). Sources of adaptation of inferior temporal cortical responses. *Cortex*, *80*, 185–195.
- Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, 8619–8624.