

Statistical Learning of Serial Visual Transitions by Neurons in Monkey Inferotemporal Cortex

Travis Meyer,¹ Suchitra Ramachandran,^{1,2} and Carl R. Olson^{1,2,3}

¹Center for the Neural Basis of Cognition and ²Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, and

³Department of Neuroscience, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

If monkeys repeatedly, over the course of weeks, view displays in which two images appear in fixed sequence, then neurons of inferotemporal cortex (ITC) come to exhibit prediction suppression. The response to the trailing image is weaker if it follows the leading image with which it was paired during training than if it follows some other leading image. Prediction suppression is a plausible neural mechanism for statistical learning of visual transitions such as has been demonstrated in behavioral studies of human infants and adults. However, in the human studies, subjects are exposed to continuous sequences in which the same image can be both predicted and predicting and statistical dependency can exist between nonadjacent items. The aim of the present study was to investigate whether prediction suppression in ITC develops under such circumstances. To resolve this issue, we exposed monkeys repeatedly to triplets of images presented in fixed order. The results indicate that prediction suppression can be induced by training not only with pairs of images but also with longer sequences.

Key words: inferotemporal; monkey; prediction suppression; statistical learning; vision

Introduction

Human infants and adults are able to learn rapidly through passive experience the statistical relations governing the transitions from one element to the next in a structured stream of visual stimuli (Fiser and Aslin, 2002; Kirkham et al., 2002; Turk-Browne et al., 2005, 2008; Howard et al., 2008; Kim et al., 2009; Bulf et al., 2011) or auditory stimuli (Saffran et al., 1996; Gómez, 2002; Creel et al., 2004; Newport and Aslin, 2004; Onnis et al., 2005; Gebhart et al., 2009; Pelucchi et al., 2009; Romberg and Saffran, 2010). The neuronal mechanisms underlying this capacity are not yet well understood (Summerfield and Egner, 2009; Meyer and Olson, 2011; Wacongne et al., 2012; Gavornik and Bear, 2014).

Inferotemporal cortex (ITC) is a plausible candidate as the site for the learning of visual transitional statistics. ITC is the terminus of the ventral stream of visual areas. As such, it plays a critical role in object vision. Furthermore, neurons in ITC exhibit statistical learning. Repeated viewing of a single image leads to familiarity suppression: the experienced image elicits comparatively weak responses (Freedman et al., 2006; Mruczek and Sheinberg, 2007; Meyer and Olson, 2014). Repeated viewing of two images close together in time leads to pair coding: neurons responsive to

one image tend to respond to the other (Miyashita, 1988; Erickson and Desimone, 1999; Li and DiCarlo, 2008). Finally, and critically, repeated viewing of two images in fixed sequence, so that the leading image becomes a strong predictor for the trailing image, leads to prediction suppression: the trailing image, when presented in the trained context, elicits only a weak response (Meyer and Olson, 2011).

Prediction suppression is a plausible mechanism for sensitivity to transitional statistics at the behavioral level. However, there is a difference between the circumstances under which prediction suppression has been demonstrated—presentation of two images in sequence—and circumstances under which statistical learning is studied in humans—presentation of long strings of images. Long sequences possess two distinctive properties. First, each image can play a dual role, not only confirming or violating a prediction conveyed by a preceding image but also conveying a prediction about a subsequent image. Second, each image can condition the probability not only of the immediately succeeding image but also of later images. The aim of the present study was to determine whether prediction suppression is induced by training with sequences possessing these properties.

Materials and Methods

Subjects. We studied two adult rhesus macaque monkeys (monkey 1, male, laboratory designation Tu, and monkey 2, female, laboratory designation Ec). All experimental procedures were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee and were in compliance with the guidelines set forth in the USPHS Guide for the Care and Use of Laboratory Animals.

Images. All stimuli were digitized images of background-free objects. When presented at fixation 32 cm from the monkey's eyes, each image subtended 4° of visual angle along whichever axis, vertical or horizontal, was longer. Eighteen images were used for training each monkey.

Received March 26, 2014; revised June 2, 2014; accepted June 7, 2014.

Author contributions: T.M., S.R., and C.R.O. designed research; T.M., S.R., and C.R.O. performed research; T.M., S.R., and C.R.O. analyzed data; T.M., S.R., and C.R.O. wrote the paper.

Support from National Institutes of Health (NIH) R01 EY018620, NIH P50 MH084053, NIH K08 MH080329, and Pennsylvania Department of Health's Commonwealth Universal Research Enhancement Program. Technical support from NIH P30 EY08098 and P41RR03631. We thank Karen McCracken for technical assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Travis Meyer, PhD, Center for the Neural Basis of Cognition, 115 Mellon Institute, 4400 Fifth Avenue, Pittsburgh, PA 15213. E-mail: tmeyer@cnbc.cmu.edu.

DOI:10.1523/JNEUROSCI.1215-14.2014

Copyright © 2014 the authors 0270-6474/14/349332-06\$15.00/0

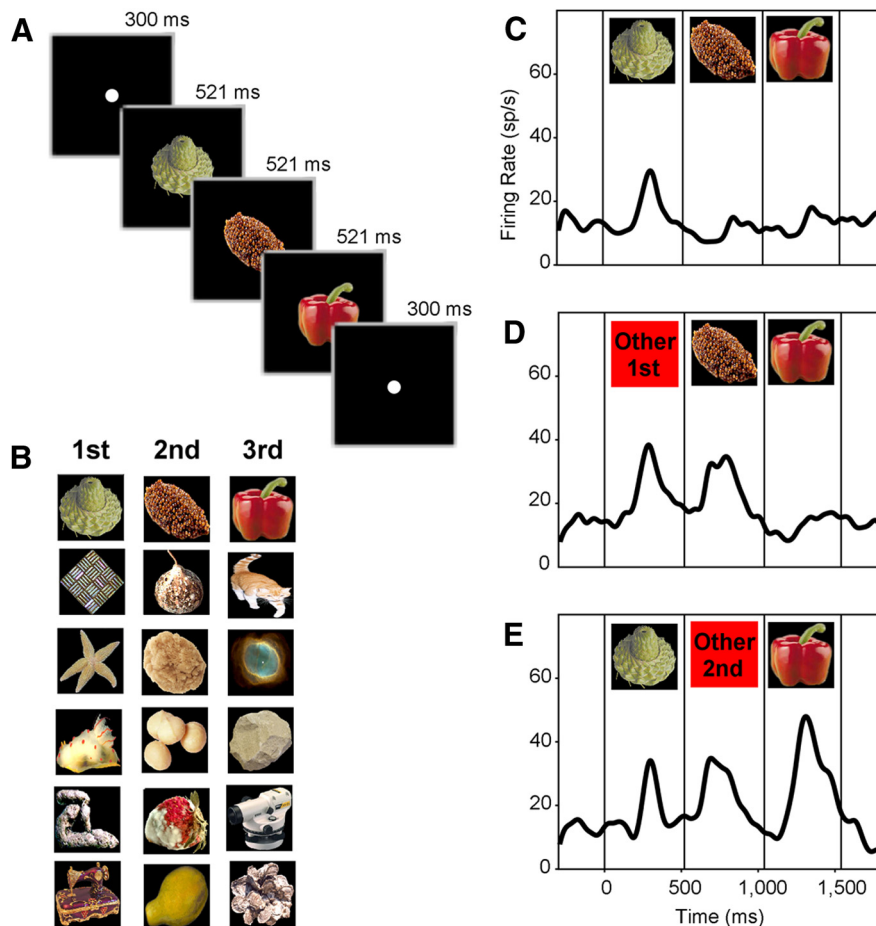


Figure 1. *A*, Timing of events in each trial conducted during training and testing. *B*, Six triplets used in training monkey 1. The triplets used in training monkey 2 were formed from different images. *C–E*, Mean firing rate of a typical neuron in response to (*C*) trained sequences, (*D*) untrained sequences in which the first stimulus had been replaced with the first item from another training triplet, and (*E*) untrained sequences in which the second stimulus had been replaced with the second item from another triplet.

The image sets used for monkeys 1 and 2 contained no items in common.

Training. Each monkey received repeated exposure to six triplets of images. The images in each triplet were always presented in the same sequence. The succession of events in each trial was as follows: fixation spot (300 ms), first image at screen center (503 ms), an 18 ms delay, second image at screen center (503 ms), an 18 ms delay, third image at screen center (503 ms), an 18 ms delay, fixation spot (300 ms), and reward delivery. A trial was aborted without reward if the monkey broke central fixation at any time. On each training day, the monkey completed one or more runs. During a run, each triplet was presented 10 times for a total of 60 trials. The sequence of trials within a run was random with the exception that during each block of six successfully completed trials each triplet must be presented once. Monkey 1 viewed each triplet 1090 times over the course of 32 d. Monkey 2 viewed each triplet 830 times over the course of 40 d.

Testing. During neuronal data collection, the monkeys performed a task identical to the one used during training with the sole exception that images were presented not only in trained sequences but also in sequences created by substitution of an item occupying the first or second position in another trained triplet. This was the smallest set of sequences required to ensure that the prediction status of an image was fully counterbalanced against other factors likely to influence neuronal firing. These factors included the identity of the preceding image, the identity of the current image and general effects carried over from earlier in the trial, such as adaptation or an off-response. Other potentially informative sequences, for example, presentation of the images as singletons, were omitted to minimize the danger that exposure to untrained sequences

would attenuate the training effect. There was a trend toward attenuation of the effect over the course of the recording sessions, but the trend did not achieve significance and the effect remained robust even during late sessions.

Recording. The electrode was introduced through a vertical guide tube into left (monkey 1) or right (monkey 2) ITC. Recording sites, identified by extrapolation from MRI-visible fiducial markers within the chamber, occupied the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus at levels anterior to the interaural plane by 16–19 mm in monkey 1 and 13–16 mm in monkey 2.

Database. We recorded from 52 sites (27 and 25 in monkeys 1 and 2). Low-pass filtered traces from these sites formed the LFP database. Neurons characterized during a complete test run numbered 112 (67 from monkey 1 and 45 from monkey 2). We classified a neuron as visually responsive if the postimage-onset firing rate (50–550 ms) differed significantly (paired *t* test, $\alpha = 0.05$) from the pre-image-onset firing rate (–300 to +50 ms). The neuronal database consisted of 75 visually responsive neurons (39 from monkey 1 and 36 from monkey 2).

Statistical analysis. To demarcate periods during the trial when the population firing rate was different for untrained sequences and trained sequences, we compared the instantaneous difference signal (untrained firing rate minus trained firing rate) to an instantaneous statistical threshold based on a Monte Carlo analysis. We applied this procedure independently to untrained sequences containing a misfit first image and those containing a misfit second image. For each neuron, we considered all 30 trained-sequence trials and all 30 untrained-sequence trials. Working with data at 1 ms resolution, we converted the discrete

spike events in each trial to a spike-density function by convolution with a 10 ms Gaussian kernel. Then, over 1000 iterations, we labeled 30 randomly selected trials from each neuron as “pseudo-trained,” labeled the 30 remaining trials from each neuron as “pseudo-untrained” and computed, for each 1 ms bin, the mean across all neurons of the signed difference between pseudo-trained and pseudo-untrained trials. Upon completion of the iterative procedure, we computed the SD of the 1000 values in each 1 ms bin. We defined +2.58 and –2.58 SDs as the upper and lower confidence limits at that point in time. For the observed signal to cross either of these limits implied a likelihood of $p < 0.01$ that it would have occurred through random shuffling. We applied an identical procedure, including smoothing with a 10 ms Gaussian kernel, to the LFP data.

Results

We exposed monkeys during a training period extending over several weeks to triplets of images presented in fixed back-to-back sequence for half a second each (Fig. 1*A*). The monkeys were rewarded at the end of each trial if they had maintained central fixation throughout the display. Each monkey viewed each of the six triplets >800 times during training (Fig. 1*B*). In ensuing microelectrode recording sessions, we measured neuronal responses elicited in anterior ITC not only by the trained triplets but also by untrained triplets created through substitution in a trained sequence of one element from another sequence. For each trained triplet, we created five untrained variants by replacing the

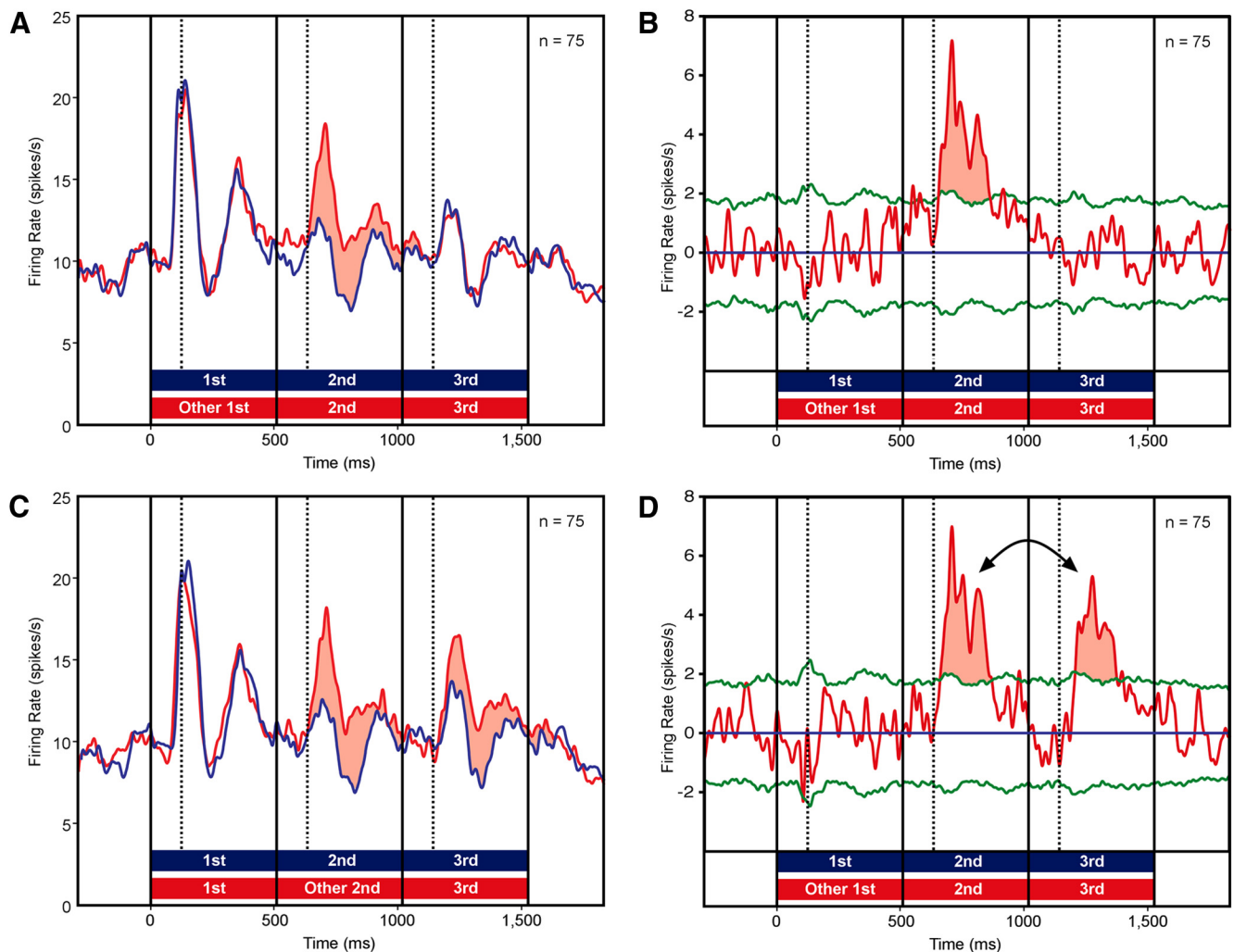


Figure 2. *A*, Population firing rate elicited by trained sequences (blue curve) and by untrained sequences with a misfit first image (red curve). Red fill indicates the period during which the response to the untrained sequence was greater than the response to the trained sequence. *B*, The difference between the two population firing rates. Green curves represent confidence limits ($p < 0.01$) based on a Monte Carlo shuffling test. Red fill indicates the period during which the response to the untrained sequence was significantly greater than the response to the trained sequence. *C, D*, These plots compare the population firing rate elicited by trained sequences (blue curve) to the population firing rate elicited by untrained sequences with a misfit second image (red curve). Conventions as in *A* and *B*. Arrows are discussed in text. The dashed lines, each marking a point in time 125 ms after stimulus onset, are included to facilitate visual estimation of effect latency. The biphasic response pattern is common (Rollenhagen and Olson, 2005) although not universal (Fig. 1) among neurons in ITC when responding to familiar images.

first image with the first image from another trained sequence and five untrained variants by replacing the trained second image with the second image from another trained sequence. During a recording session, each of the 6 trained triplets was presented five times and each of the 60 untrained triplets was presented once for a total of 90 trials. Using this procedure, we collected data from 75 visually responsive neurons (39 in monkey 1 and 36 in monkey 2).

The responses of a typical neuron are displayed in Figure 1*C–E*. Presented with a trained sequence, this neuron responded strongly to the first image but weakly to the subsequent two images (Fig. 1*C*). This outcome can be explained as a consequence of prediction suppression arising from adjacent dependencies: the second image confirmed a prediction conveyed by the first image and the third image confirmed a prediction conveyed by the second image. Presented with an untrained sequence containing a misfit first image, the neuron responded weakly to the third image (Fig. 1*D*). This outcome can also be explained in terms of adjacent dependencies: the second image violated a prediction based on the first image whereas the third image confirmed a

prediction based on the second image. Presented with an untrained sequence containing a misfit second image, the neuron responded strongly to both the second and the third image (Fig. 1*E*). This outcome likewise allows an explanation based on adjacent dependencies: the second image violated a prediction based on the first image and the third image violated a prediction based on the second image.

To determine whether the activity of the neuronal population as a whole conformed to this pattern, we performed a paired *t* test ($n = 75$) on the firing rate 50–500 ms following presentation of each image. Introduction of a misfit first image enhanced the response to the second image ($p = 1.5 \times 10^{-6}$) but not the third image ($p = 0.93$). Introduction of a misfit second image enhanced the responses to both the second image ($p = 3.2 \times 10^{-6}$) and the third image ($p = 6.4 \times 10^{-5}$). These results were present and significant ($\alpha = 0.05$) in each monkey considered individually. To examine the time course of the effect, we constructed curves representing mean population firing rate as a function of time during the trial under all three conditions. When the first image was a misfit, the response to the second image was enhanced

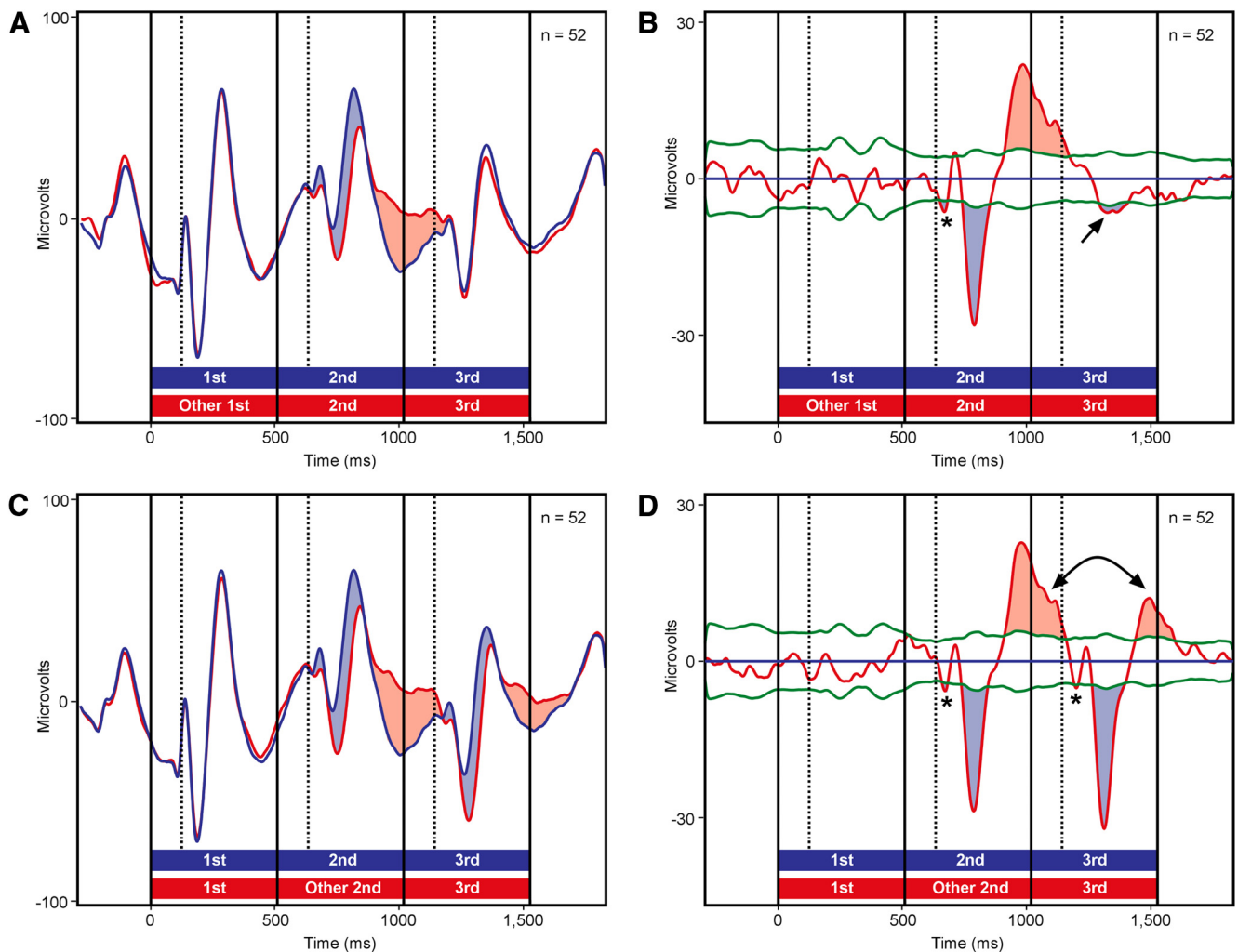


Figure 3. *A*, Mean LFP response elicited by trained sequences (blue curve) and by untrained sequences with a misfit first image (red curve). Blue (red) fill indicates periods during which the response to the untrained sequence was more negative (positive) than the response to the trained sequence. *B*, The difference between the two LFP responses. Blue (red) fill indicates periods during which the response to the untrained sequence was significantly more negative (positive) than the response to the trained sequence. *C*, *D*, These plots compare the mean LFP response elicited by trained sequences (blue curve) to the mean LFP response elicited by untrained sequences with a misfit second image (red curve). Asterisks and arrows indicate events discussed in the main text. Other conventions as in Figure 2.

relative to trained-sequence baseline (Fig. 2*A*, red fill). When the second image was a misfit, the responses to the second and third images were visibly enhanced (Fig. 2*C*, red fill). To analyze the timing of the effect, we computed the instantaneous difference in firing rate between untrained and trained sequences (Fig. 2*B*, *D*, red curves) and determined when it crossed a confidence limit ($p = 0.01$) established by a Monte Carlo procedure (Fig. 2*B*, *D*, green curves). On average, across the three instances in which an image was unpredicted by the item immediately preceding it, enhancement became significant 131 ms after image onset. Further incidental observations conform to prior report based on two-item sequences: responses became weaker and occurred at longer latency as the sequence progressed, the response to an image was unaffected by the strength of the response to the preceding image, and the response to an unpredicted image was scaled up multiplicatively from the response elicited by the same image when predicted (Meyer and Olson, 2011). We do not yet know whether the mechanism of the prediction effect is “surprise enhancement” or “prediction suppression,” because neither the original experiment nor this one contained a prediction-neutral control.

The LFP responses recorded at all 52 sites (27 in monkey 1 and 25 in monkey 2) depended in a similar fashion on the prediction status of each image. The response to each unpredicted image deviated from the response to the same image in a trained triplet by coursing first more negatively (Fig. 3, blue fill) and then more positively (Fig. 3, red fill). Each effect achieved statistical significance ($\alpha = 0.01$) as indicated by its exceeding Monte Carlo-based confidence limits (Fig. 3*B*, *D*, green curves). The fact that the untrained sequences contained three images violating a prediction conveyed by the immediately preceding item allowed us to judge which features of the prediction effect were consistently present. Of particular note is the negative deflection that achieved brief significance shortly after image onset (all three instances are marked by asterisks in Fig. 3*B* and *D*). Although this event was of low amplitude, it was absolutely consistent. The average time of attainment of significance across three conditions was 121 ms. This was 10 ms earlier than the onset of the spiking effect. On the assumption that the earliest phase of the LFP is generated by bottom-up synaptic input to ITC, this observation raises the possibility that neurons afferent to ITC are sensitive to the prediction status of an image.

The effects described up to this point can be explained entirely in terms of adjacent dependencies. The response to each image was strong if it violated a prediction conveyed by the immediately preceding image and weak otherwise. Nonadjacent dependencies could, however, have exerted a superadded effect on neural activity because the first image in each triplet strongly predicted the final image. There are trends in the data suggesting that the response to the third image was indeed affected by its violating or confirming a prediction conveyed by the first image. In sequences containing a misfit first image, the LFP evoked by the third image exhibited a small but significant negative deflection (Fig. 3*B*, arrow). The effect occurred in both monkeys. It is consistent with an interpretation based on the third image violating a prediction conveyed by the first image. In sequences containing a misfit second image, the second image violated a prediction conveyed by the first image and the third image violated a prediction conveyed by the second image. Nevertheless, the prediction effect was weaker for the third than for the second image both at the level of spiking activity (Fig. 2*D*, double arrow) and at the level of the positive deflection of the LFP (Fig. 3*D*, double arrow). The difference fell short of statistical significance in the case of spiking activity ($p = 0.70$; paired t test; unpredicted minus predicted firing rate 150–500 ms after stimulus onset; $n = 75$) but did achieve significance in the case of the LFP ($p = 0.014$; paired t test on unpredicted minus predicted voltage 450–650 ms after stimulus onset; $n = 52$). The effect occurred and achieved significance ($\alpha = 0.05$) in each monkey considered individually. It is consistent with an interpretation based on the third image's confirming a prediction conveyed by the first image.

Discussion

The stream of experience is far from random. Events that have just occurred carry with them dependable predictions about events that will occur next. Some predictions are based on physical principles so fundamental that they may possess a hard-wired representation in the brain (Alink et al., 2010). Other predictions must be learned, such as in music and language. The brain of an acculturated listener listening to a melody or a sentence is sensitive to how it will probably unfold. This is manifest in the fact that events violating reasonable expectation elicit strong neural responses (Dien et al., 2003; James et al., 2008; Vuust et al., 2009; Pearce et al., 2010; Kim et al., 2011). The ability of the brain to detect and respond to improbable events is thought to depend not only on mastery of complex rules that govern the event stream but also on the encoding of simple statistical relations (Aslin and Newport, 2012). Melodies and sentences exhibit tonotactic and phonotactic regularities: the value of an upcoming note or phoneme is probabilistically related to the values of at least the two preceding elements (Pearce and Wiggins, 2004; Gonzalez-Gomez and Nazzi, 2013). Human infants and adults are able to learn rapidly, during passive listening, the statistical relations between immediately adjacent items in a structured auditory stream (Saffran et al., 1996; Pelucchi et al., 2009; Romberg and Saffran, 2010). They are also able, under favorable circumstances, to learn nonadjacent dependencies between events separated by an intervening item (Gómez, 2002; Creel et al., 2004; Newport and Aslin, 2004; Onnis et al., 2005; Gebhart et al., 2009). The capacity for the learning of transitional statistics, although most studied in the auditory domain, extends to visual sequences as well (Fiser and Aslin, 2002; Kirkham et al., 2002; Turk-Browne et al., 2005, 2008; Howard et al., 2008; Kim et al., 2009; Bulf et al., 2011).

Prediction suppression, as observed in ITC, is a plausible mechanism for behavioral sensitivity to visual transitional statistics, as demonstrated in the human studies. However, for it to serve this role would require that it develop in the context of multi-item image sequences or, in other words, under conditions in which the same image can be both predicted and predicting and in which the potential for nonadjacent predictions exists. The key observation of this study is that prediction suppression does indeed occur under these conditions.

The fact that prediction suppression occurred under circumstances in which the second image was both predicted and predicting casts light on the nature of neuronal mechanisms mediating suppression. In the simplest possible model of the phenomenon, ITC neurons responsive to a given leading image induce a state of suppression among neurons responsive to the predicted trailing image. In this framework, the rate of firing of neurons representing the leading image could reasonably be expected to determine the degree of suppression of the response to the trailing image. However, it did not. Regardless of whether the second image confirmed a prediction and elicited a weak response or violated a prediction and elicited a strong response, the response to the predicted third image was the same (Fig. 2*A*, red vs blue).

The occurrence of subtle effects apparently dependent on whether the third image violated or confirmed a prediction conveyed by the first image is consistent with findings indicating that human observers learn nonadjacent dependencies (Gómez, 2002; Creel et al., 2004; Newport and Aslin, 2004; Onnis et al., 2005; Gebhart et al., 2009). However, it is surprising in light of the fact that learning of nonadjacent statistics by human observers depends on use of a training display that emphasizes relations between nonadjacent items, either by making them physically similar (Creel et al., 2004; Onnis et al., 2005; Gebhart et al., 2009) or by randomizing the identity of the intervening element (Gómez, 2002; Newport and Aslin, 2004). The use of prolonged training in our study may have allowed nonadjacent dependencies to exert an effect even without these manipulations. This interpretation fits with the observation that humans are sensitive to nonadjacent dependencies after prolonged natural exposure (Pearce and Wiggins, 2004; Gonzalez-Gomez and Nazzi, 2013).

References

- Alink A, Schwedzrik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. *J Neurosci* 30:2960–2966. [CrossRef Medline](#)
- Aslin RN, Newport EL (2012) Statistical learning: from acquiring specific items to forming general rules. *Curr Dir Psychol Sci* 21:170–176. [CrossRef Medline](#)
- Bulf H, Johnson SP, Valenza E (2011) Visual statistical learning in the newborn infant. *Cognition* 121:127–132. [CrossRef Medline](#)
- Creel SC, Newport EL, Aslin RN (2004) Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J Exp Psychol Learn Mem Cogn* 30:1119–1130. [CrossRef Medline](#)
- Dien J, Frishkoff GA, Cerbone A, Tucker DM (2003) Parametric analysis of event-related potentials in semantic comprehension: evidence for parallel brain mechanisms. *Brain Res Cogn Brain Res* 15:137–153. [CrossRef Medline](#)
- Erickson CA, Desimone R (1999) Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J Neurosci* 19:10404–10416. [Medline](#)
- Fiser J, Aslin RN (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn* 28:458–467. [CrossRef Medline](#)
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2006) Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex* 16:1631–1644. [Medline](#)

- Gavornik JP, Bear MF (2014) Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat Neurosci* 17:732–737. [CrossRef Medline](#)
- Gebhart AL, Newport EL, Aslin RN (2009) Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon Bull Rev* 16:486–490. [CrossRef Medline](#)
- Gómez RL (2002) Variability and detection of invariant structure. *Psychol Sci* 13:431–436. [CrossRef Medline](#)
- Gonzalez-Gomez N, Nazzi T (2013) Effects of prior phonotactic knowledge on infant word segmentation: the case of nonadjacent dependencies. *J Speech Lang Hear Res* 56:840–849. [CrossRef Medline](#)
- Howard JH, Howard DV, Dennis NA, Kelly AJ (2008) Implicit learning of predictive relationships in three-element visual sequences by young and old adults. *J Exp Psychol Learn Mem Cogn* 34:1139–1157. [CrossRef Medline](#)
- James CE, Britz J, Vuilleumier P, Hauert CA, Michel CM (2008) Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. *Neuroimage* 42:1597–1608. [CrossRef Medline](#)
- Kim R, Seitz A, Feenstra H, Shams L (2009) Testing assumptions of statistical learning: is it long-term and implicit? *Neurosci Lett* 461:145–149. [CrossRef Medline](#)
- Kim SG, Kim JS, Chung CK (2011) The effect of conditional probability of chord progression on brain response: an MEG study. *PLoS One* 6:e17337. [CrossRef Medline](#)
- Kirkham NZ, Slemmer JA, Johnson SP (2002) Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83:B35–B42. [CrossRef Medline](#)
- Li N, DiCarlo JJ (2008) Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* 321:1502–1507. [CrossRef Medline](#)
- Meyer T, Olson C (2014) Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. In press.
- Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc Natl Acad Sci U S A* 108:19401–19406. [CrossRef Medline](#)
- Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335:817–820. [CrossRef Medline](#)
- Mruczek RE, Sheinberg DL (2007) Context familiarity enhances target processing by inferior temporal cortex neurons. *J Neurosci* 27:8533–8545. [CrossRef Medline](#)
- Newport EL, Aslin RN (2004) Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn Psychol* 48:127–162. [CrossRef Medline](#)
- Onnis L, Monaghan P, Richmond K, Chater N (2005) Phonology impacts segmentation in online speech processing. *J Mem Lang* 53:225–237. [CrossRef](#)
- Pearce MT, Wiggins GA (2004) Improved methods for statistical modeling of monophonic music. *J New Music Res* 33:367–385.
- Pearce MT, Ruiz MH, Kapasi S, Wiggins GA, Bhattacharya J (2010) Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage* 50:302–313. [CrossRef Medline](#)
- Pelucchi B, Hay JF, Saffran JR (2009) Statistical learning in a natural language by 8-month-old infants. *Child Dev* 80:674–685. [CrossRef Medline](#)
- Rollenhagen JE, Olson CR (2005) Low-frequency oscillations arising from competitive interactions between visual stimuli in macaque inferotemporal cortex. *J Neurophysiol* 94:3368–3387. [CrossRef Medline](#)
- Romberg AR, Saffran JR (2010) Statistical learning and language acquisition. *Wiley Interdiscip Rev Cogn Sci* 1:906–914. [CrossRef Medline](#)
- Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928. [CrossRef Medline](#)
- Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13:403–409. [CrossRef Medline](#)
- Turk-Browne NB, Jungé JA, Scholl BJ (2005) The automaticity of visual statistical learning. *J Exp Psychol Gen* 134:552–564. [CrossRef Medline](#)
- Turk-Browne NB, Isola PJ, Scholl BJ, Treat TA (2008) Multidimensional visual statistical learning. *J Exp Psychol Learn Mem Cogn* 34:399–407. [CrossRef Medline](#)
- Vuust P, Ostergaard L, Pallesen KJ, Bailey C, Roepstorff A (2009) Predictive coding of music—brain responses to rhythmic incongruity. *Cortex* 45:80–92. [CrossRef Medline](#)
- Wacongne C, Changeux JP, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci* 32:3665–3678. [CrossRef Medline](#)