

# Statistical learning of visual transitions in monkey inferotemporal cortex

Travis Meyer<sup>a,1</sup> and Carl R. Olson<sup>a,b</sup>

<sup>a</sup>Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213; and <sup>b</sup>Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA 15260

Edited by Robert H. Wurtz, National Institutes of Health, Bethesda, MD, and approved October 13, 2011 (received for review August 5, 2011)

One of the most fundamental functions of the brain is to predict upcoming events on the basis of the recent past. A closely related function is to signal when a prediction has been violated. The identity of the brain regions that mediate these functions is not known. We set out to determine whether they are implemented at the level of single neurons in the visual system. We gave monkeys prolonged exposure to pairs of images presented in fixed sequence so that each leading image became a strong predictor for the corresponding trailing image. We then monitored the responses of neurons in the inferotemporal cortex to image sequences that obeyed or violated the transitional rules imposed during training. Inferotemporal neurons exhibited a transitional surprise effect, responding much more strongly to unpredicted transitions than to predicted transitions. Thus, neurons even in the visual system make experience-based predictions and react when they fail.

macaque | vision | plasticity

The inferotemporal cortex (ITC), the terminus of the ventral stream of visual areas (1), plays a critical role in object vision (2, 3). ITC neurons respond with individual patterns of selectivity to complex images (4). Training monkeys to discriminate between images (5–7), categorize them (8, 9), or form associations between them (10–12) induces functional changes among neurons in the ITC which have the effect of strengthening the representation of image attributes relevant to task performance. Even passive viewing causes changes in neuronal visual responsiveness. Repeated viewing of a single image leads to a weakening of responses to it (13, 14). Repeated viewing of two images close together in time leads to pair coding: a tendency for neurons responsive to one image also to respond to the other (15–17). The effects of passive viewing, because they do not depend on task demands, fall into the category of unsupervised statistical learning.

An important form of unsupervised learning not previously studied at the level of single neurons concerns transitional statistics. The learning of transitional statistics has been the focus of much behavioral study in humans because it is thought to underlie the development during infancy of the ability to perceive event boundaries including word boundaries in speech (18, 19). Human infants passively exposed to a stimulus stream in which certain visual images always follow certain others automatically register the transitional rules as evidenced by their orienting preferentially to a test stream containing novel transitions (20). The adult human brain is sensitive to transitional probabilities, as evidenced by its generating strong responses to improbable transitions at the level of scalp potential and blood oxygenation measures (21–27). Monkeys, like human infants, have been reported to learn transitional probabilities and to orient preferentially to improbable transitions (28). No effort has been made as yet to characterize the underlying neuronal mechanisms (29). We hypothesized that neurons in the ITC would acquire sensitivity to the transitional statistics of visual displays over the course of prolonged training and would manifest this sensitivity in subsequent tests by responding strongly to events violating the transitional rules imposed during training.

## Results

We began by exposing two monkeys repeatedly to six pairs of images. Each pair consisted of a leading image and a trailing image ( $A_m B_n$ ,  $m = n$ ) as shown in Fig. 1A. On each trial, while the monkey looked at the center of the screen, the two images appeared in immediate succession at fixation for half a second each. The monkey was rewarded at the end of the trial subject to the sole requirement of having maintained fixation. Over the course of training, each pair was presented more than 800 times, always in the same order (Fig. 1A, blue counts). The training runs were distributed across 3 mo in monkey 1 and 1 mo in monkey 2. Following training, we began to collect data from single ITC neurons in microelectrode recording sessions. The essential question was whether trailing image  $B_n$  would elicit an enhanced response when it violated prediction by appearing after a leading image other than its training partner ( $A_m$ ,  $m \neq n$ ). To answer this question, we used a procedure identical to the training procedure except that the trained sequences and all possible untrained sequences were presented in interleaved trials. An untrained sequence consisted of a leading image and a trailing image that belonged to different training pairs. To minimize any attenuation of the effects of training, we adopted a design in which each of the six trained sequences was presented eight times in a full run, whereas each of the 30 untrained sequences was presented only once (Fig. 1A, black counts). Using this procedure, we collected data from 81 visually responsive ITC neurons (46 in monkey 1 and 35 in monkey 2).

In some neurons, the trailing image clearly elicited a stronger response when it was unpredicted (Fig. 1C) than when it was predicted (Fig. 1B). To determine whether this pattern was consistent across the population, we constructed plots representing the mean population firing rate as a function of time in trials in which the trailing image was either predicted or unpredicted (Fig. 2A). The population obviously responded much more strongly when the trailing image was unpredicted (red curve) than when it was predicted (blue curve). This effect was statistically significant at the population level (paired  $t$  test,  $n = 81$ ,  $P = 1.5e-11$ ;  $P < 0.0005$  in each monkey) and achieved significance in 33 of 81 neurons considered individually (ANOVA with prediction status and image identity as factors,  $\alpha = 0.05$ ). It also was significant at the level of the local field potential (LFP) (paired  $t$  test,  $n = 71$ ,  $P = 2.0e-9$ ;  $P < 0.001$  in each monkey). We refer to the stronger response in trials involving an unpredicted transition as a transitional surprise effect.

To assess whether the transitional surprise effect was associated with a more robust representation of image identity, we ranked the trailing images for each neuron from the least

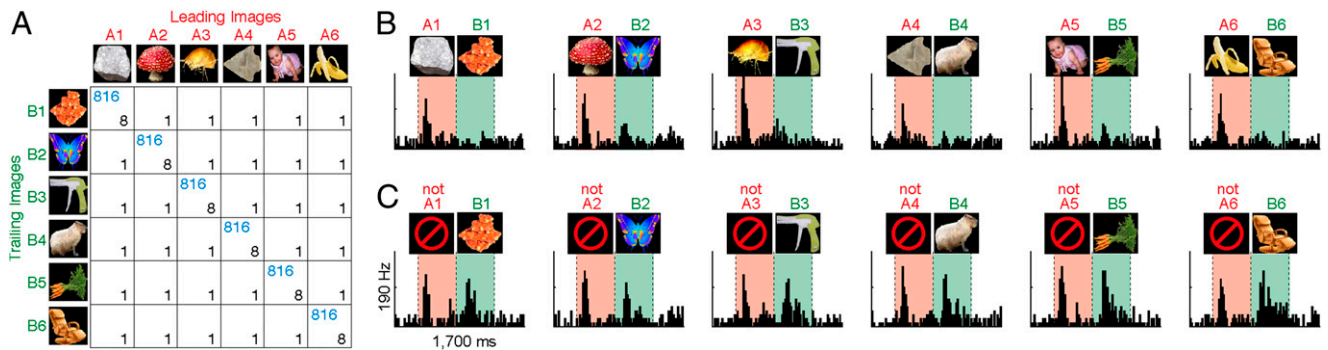
Author contributions: T.M. and C.R.O. designed research; T.M. performed research; T.M. and C.R.O. analyzed data; and T.M. and C.R.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. E-mail: tmeyer@cncb.cmu.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112895108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112895108/-DCSupplemental).

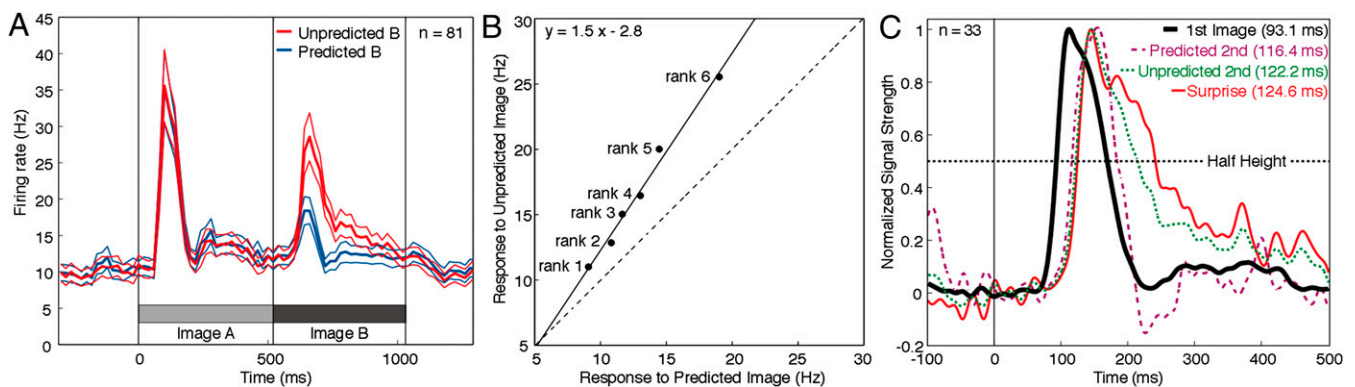


**Fig. 1.** Pairs of images were presented in sequence during training and neuronal recording. (A) Six leading images (A<sub>1</sub>–A<sub>6</sub>) and six trailing images (B<sub>1</sub>–B<sub>6</sub>). Numbers indicate how many exposures to each pair occurred during training (blue counts) and subsequently during each recording run (black counts). On each trial, during both training and testing, a leading image and a trailing image were presented in succession at screen center with the following measured timing: leading image (503 ms), gap (18 ms), and trailing image (507 ms). (B) Data from one neuron. Each histogram represents mean firing rate as a function of time across eight trials in which the leading image and trailing image were paired as they had been during training (A<sub>m</sub>B<sub>n</sub>, m = n). The leading and trailing images were visible during periods demarcated by dashed lines and shaded with red and green, respectively. (C) Data from the same neuron. Each histogram represents firing on five trials in which the trailing image followed the five leading images not paired with it during training (A<sub>m</sub>B<sub>n</sub>, m ≠ n).

effective (rank 1) to the most effective (rank 6) using a measure that gave equal weight to predicted and unpredicted trials. Then we computed the average firing rate, across all neurons, elicited by images of rank 1 through 6 when they were predicted and when they were unpredicted. On fitting a line to the six points representing unpredicted as a function of predicted firing rate (Fig. 2B), we found that the slope (1.5) was significantly ( $P < 1.0e-7$ ) greater than 1. The intercept was not significantly different from 0. One consequence of the increase in response gain was that the spread of firing rates across images was greater when they were unpredicted than when they were predicted. This effect could be expected to increase useable information about image identity. To determine whether it did so, we carried out a signal detection analysis on data from each neuron. We found that the neuronal firing rate carried significantly more information about which image had been presented if the image was unpredicted than if it was predicted (Figs. S1 and S2).

The detection of a prediction-violating event might require processing time beyond the onset of the visual response. To

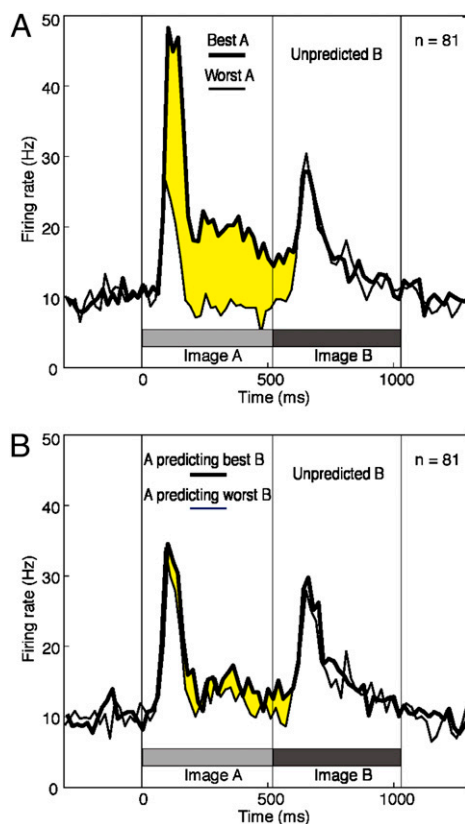
explore this possibility, we investigated the timing of stimulus-driven activity among 33 neurons significantly more responsive to unpredicted than to predicted images. There were three main findings. First, the latency of the response to the trailing image was longer than the latency of the response to the leading image (Fig. 2C). This difference probably was not an effect of training. The response to an image is delayed if it displaces another image instead of appearing against a blank background (30). Second, the trailing image elicited a slightly but significantly earlier response if it was predicted than if it was unpredicted (Fig. 2C; Wilcoxon test, time to half height,  $n = 33$ ,  $P = 0.011$ ). This difference may be related to the fact that a match image elicits a slightly earlier response than a nonmatch image in a delayed-match-to-sample task (30). Third, and most importantly, the transitional surprise signal (the difference in firing rate between unpredicted and predicted trials) developed at a time statistically indistinguishable from the time of the response to the unpredicted trailing image (Fig. 2C; Wilcoxon test, time to half height,  $n = 33$ ,  $P = 0.91$ ). From the last



**Fig. 2.** Neuronal responses to unpredicted images were enhanced in strength and selectivity from their outset. (A) The population firing rate elicited by the trailing image was greater when it followed another image's predictor (red curve) than when it followed its own predictor (blue curve). Data shown are the mean across all 81 neurons in 20-ms bins. The bounds on each curve represent  $\pm 1$  SEM. (B) Each point represents the mean across all 81 neurons of the firing rate elicited by a trailing image of a given rank. Trailing images were ranked independently for each neuron from 1 (least effective) to 6 (most effective). Responses to the trailing image when predicted and unpredicted are represented along the horizontal and vertical axes respectively. The identity line (dashed), the best-fit line (solid) and its formula (Upper Left) are shown. (C) Timing of population signal strength for the response to the leading ("1st") image, the response to the trailing ("2nd") image when predicted, the response to the trailing ("2nd") image when unpredicted, and the transitional surprise signal (response to trailing image when unpredicted minus the response when predicted). Data are based on all 33 neurons exhibiting a statistically significant transitional surprise signal. Each curve smoothed (by convolution with a Gaussian function with 10-ms SD) and normalized to bring the preresponse baseline (0–50 ms after image onset) to 0 and peak response to 1. Each time to half height is indicated in parentheses.

observation, we conclude that little or no additional processing time was required for generation of a surprise signal.

The strength of the response to the trailing image might depend on the strength of the response to the leading image (31). To assess this possibility, we classified the leading images independently for each neuron as best (the single most effective image), worst (the single least effective image), and intermediate (the four other images). Then we constructed population plots representing firing rate as a function of time in trials in which the leading image was either best or worst and the trailing images were the four paired during training with leading images in the intermediate category. Before onset of the trailing image, the firing rates under the two conditions were markedly different; however, following trailing-image onset, the difference vanished (Fig. 3A), with the firing rates becoming statistically indistinguishable (paired *t* test,  $n = 81$ ,  $P = 0.97$ ). Thus, the strength of the response to the trailing image did not depend on the strength of the response to the leading image.



**Fig. 3.** The response to an unpredicted trailing image is dependent neither on the strength of the response to the leading image nor on the strength of the response that the predicted trailing image would have elicited. (A) Mean firing rate computed across all 81 neurons in trials in which the leading image was the one to which the neuron responded most strongly (thick curve) or least strongly (thin curve). Consideration was restricted to trials involving the four trailing images associated with neither of the leading images. Differential activity reflecting selectivity for the best leading image (yellow shading) persisted only until the onset of the response to the trailing image. (B) Mean firing rate computed across all 81 neurons in trials in which the leading image predicted the trailing image to which the neuron responded most strongly (thick curve) or least strongly (thin curve). Consideration was restricted to trials involving the four other trailing images. Differential activity reflecting selectivity for the leading image paired during training with the best trailing image (pair coding) persisted only until the onset of the response to the trailing image.

The strength of the response to the trailing image actually presented on a given trial might depend on the strength of the response that the predicted trailing image would have elicited (32, 33). To assess this possibility, we classified the trailing images independently for each neuron as best (the single most effective image), worst (the single least effective image), and intermediate (the four other images). Then we constructed population plots representing firing rate as a function of time in trials in which either the best or worst trailing image was predicted and in which the trailing image actually presented belonged to the intermediate category. The responses to the trailing images under the two conditions (Fig. 3B) were statistically indistinguishable (paired *t* test,  $n = 81$ ,  $P = 0.80$ ). Thus, the strength of the response to the presented trailing image did not depend on the strength of the predicted response.

Repeatedly viewing a pair of images close together in time is known to lead to pair coding in the ITC, as manifested in a tendency for neurons to respond with the same strength to both members of the pair (11, 12, 15–17). It merits tangential note that we did observe pair coding, although pair coding was not the main focus of the study. The correlation across neurons between the strength of the response to a given leading image and the strength of the response to the trailing image with which it had been paired during training was positive and significant ( $n = 81$ ,  $r = 0.28$ ,  $P = 2.8 \times 10^{-10}$ ;  $P < 0.0005$  in each monkey). Pair coding is evident in the fact that the leading image paired during training with the best trailing image elicited a stronger response than the leading image paired during training with the worst trailing image (Fig. 3B, yellow shading).

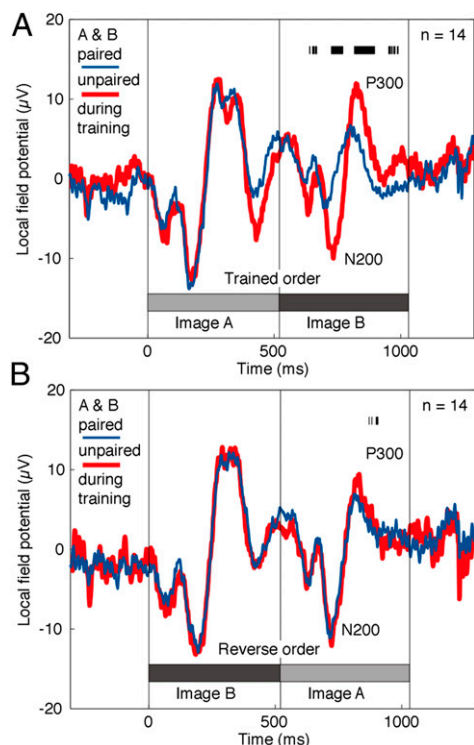
Neurons in the ITC might be truly sensitive to the transitional statistics of the training displays (with  $A_n$  unidirectionally predicting  $B_n$ ) or, alternatively, to their joint statistics (with  $A_n$  predicting  $B_n$  and vice versa). To distinguish between these possibilities, we recorded from 17 neurons at 14 sites (five neurons at five sites in monkey 1 and 12 neurons at nine sites in monkey 2) while presenting the images both in forward order as during training ( $A_n B_n$ ) and in reverse order ( $B_n A_n$ ). At the level of neuronal population activity, there was a transitional surprise effect during forward but not reverse presentation (Figs. S3 and S4). The difference between forward and reverse conditions achieved significance early in the response period (paired *t* test on transitional surprise indices for the period 50–250 ms after image onset,  $n = 17$ , forward mean = 0.13, reverse mean =  $-0.0044$ ,  $P = 0.035$ ). At the level of the LFP, the evoked response was much stronger on unpredicted trials (thick red curve) than on predicted trials (thin blue curve) under the forward condition (Fig. 4A) but not under the reverse condition (Fig. 4B). The difference between forward and reverse conditions was significant (paired *t* test on transitional surprise indices,  $n = 14$ , forward mean = 0.20, reverse mean = 0.074,  $P = 0.016$ ). Thus, the transitional surprise effect depends on transitional and not just on joint statistics.

## Discussion

ITC neurons adapt to the transitional statistics of repeatedly viewed sequential displays. This adaptation is manifest, subsequent to the training period, as a transitional surprise effect. Neurons respond with enhanced gain to images that violate the transitional rules imposed during training. The idea that local visual circuits at processing stages as early as the retina might make predictions and signal their violation is well established among visual theorists (34–42) but previously has not received support from neuronal recording studies.

It is an interesting question whether the transitional surprise effect arises from the suppression of responses to predicted images or the enhancement of responses to unpredicted images. A suppressive mechanism would suggest a possible relation to the phenomenon of repetition suppression, whereby, if an image is presented twice in succession, the second response is reduced





**Fig. 4.** The transitional surprise effect is evident at the level of the LFP. (A) LFP activity elicited by presenting images in the trained order. The N200-to-P300 peak-to-peak amplitude of the response to the trailing image was greater when the order was unpredicted ( $A_m B_n$ ,  $m \neq n$ , thick red curve) than when it was predicted ( $A_n B_m$ ,  $m = n$ , thin blue curve). (B) LFP activity elicited by presenting images in reverse order ( $B_n A_m$ ). An unpredicted ( $m \neq n$ ) and a predicted ( $m = n$ ) transition elicited nearly identical responses. Each panel is based on the 14 sites at which testing was carried out with images presented in both orders. The tick marks near the top of each panel indicate those points in time at which a paired  $t$  test ( $n = 14$ ,  $\alpha = 0.05$ ), applied successively to each 1-ms bin in the range from 0–1,000 ms, revealed a significant difference between voltages measured under the two conditions.

in strength (43). Distinguishing between enhancement and suppression would require comparing responses elicited by predicted and unpredicted images to responses collected under an appropriate baseline condition involving no prediction and counterbalanced for other relevant factors. These factors would have to include image familiarity, because familiarization reduces response strength (13, 14), and the timing of the test sequence, because a leading image exerts forward suppression on the response to a trailing image to a degree that depends on the interval between them (44).

The transitional surprise effect might arise within the ITC or, alternatively, might be relayed from an earlier area. To resolve this issue definitively would require recording from areas at earlier stages of the ventral stream hierarchy. However, we can provide one piece of evidence suggestive of an origin in the ITC. Following onset of the trailing image, the surprise effect seems to develop first at the level of spiking activity and only later at the level of the LFP (Fig. S5). On the reasonable assumption that the earliest phase of the LFP response reflects currents generated at bottom-up synaptic terminals, this observation suggests that the surprise effect develops in the ITC and is not present in bottom-up inputs.

To characterize fully the conditions necessary for the development of transitional surprise signals is outside the scope of the present experiment. However, our results do allow placing upper and lower limits on the amount of visual experience

required. At the upper limit, intensive exposure (more than 800 repetitions of each image sequence over the course of weeks) is clearly sufficient. At the lower limit, brief exposure (eight repetitions of each image sequence within a session) does not suffice. Otherwise, we would have observed stronger responses to improbable than to probable transitions during sessions in which the image pairs were presented in reverse order.

The most obvious functional interpretation for transitional surprise signals is that they confer salience on unpredicted events. Unpredicted events (45–47) including unpredicted transitions (20) capture attention automatically. They merit processing because they tend to mark event boundaries (18, 19) and because the information they carry can guide learning. It is well established in studies of animal behavior that sensory events drive associative learning effectively only to the degree that they are surprising (48–50). Formal models of learning allow for gradations of surprise dependent on how markedly an event deviates from prediction (49). We do not know whether manipulating the degree of similarity between the presented and the predicted trailing images would modulate the strength of the transitional surprise effect.

The idea that the transitional surprise effect is related to the capture of attention by unpredicted events raises a chicken-and-egg problem. Attention to an image enhances the gain of neuronal responses to it (51, 52). Is the transitional surprise effect a cause, or is it only a consequence of the capture of attention? We favor regarding it as a cause, not a consequence, for two reasons. First, the display consisted of a single foveal image. Attentional enhancement of visual response strength in extrastriate cortex depends on the simultaneous presence of competing images elsewhere in the visual field (31, 53, 54). The strength of the response to an isolated foveal image is not affected by whether the monkey is processing it actively or merely is maintaining passive fixation (5, 55). Nor is the strength of the response enhanced by making the image surprising through the violation of expectations other than those based on learned image–image transitional rules (56, 57). Second, the transitional surprise effect was present from the outset of the visual response. If the effect depended on top-down attentional influences from areas beyond the ITC, one would expect it to develop after a delay.

The reward-prediction error signal posited in classic learning theory (32) and observed in dopamine neurons of the ventral tegmental area (33) is proportional to the value of the delivered reward minus the value of the predicted reward. By direct analogy, ITC neurons might carry a visual prediction error signal proportional to the response associated with the presented image minus the response associated with the predicted image. In this case, the response to an intermediate trailing image should have been low following prediction of the best trailing image and high following prediction of the worst trailing image. No such effect occurred (Fig. 3B). We conclude that the transitional surprise effect in the ITC, although it may contribute to learning by highlighting surprising events, is not a reward-prediction error signal in the classic sense.

The transitional surprise effect conceivably could result from principles of operation that allow the brain to settle to an efficient representation of the most likely current state of the environment as based on visual input. In models involving hierarchical predictive coding (34–42), visual activation feeds forward through a chain of areas leading from V1 to the ITC and beyond, with neuronal activity at successively later stages representing hypotheses about successively more global attributes of the visual stimulus. If a hypothesis represented by activity in a high-order area predicts a hypothesis represented by activity in a low-order area, rendering the latter redundant, then feedback from the high-order area induces a reduction of activity in the low-order area. This reduction may take the form of suppressing or “explaining away” the activity of neurons representing the more

local hypothesis (36, 37) or of silencing neurons that otherwise would signal a signed (38–42) or an unsigned (34) prediction error. This form of processing, although commonly considered in relation to the prediction of local attributes by simultaneously present global attributes, also can accommodate the prediction of subsequent events by antecedent ones (39). An area hierarchically superior to the ITC might respond to the leading image by generating feedback signals that reduce the visual response gain of ITC neurons representing the predicted trailing image, so that, upon its appearance 500 ms later, it elicits only a weak response.

We conclude by noting that the present observations are unprecedented. One prior study demonstrated an effect of transitional probability on neuronal auditory responses in the zebra finch but did not address whether the effect was a product of experience (58). Two studies other than ours have assessed the impact of transitional probability on neuronal activity in the ITC. Neither of these revealed evidence for surprise signals. Kalikhovich and Vogels (57) found that when the same image was presented twice in succession, the response to the second presentation was the same regardless of whether the repetition was expected or unexpected. Anderson and Sheinberg (56) found that when a cue predicted onset of a second image after a certain interval, firing was attenuated (not enhanced) if the second image appeared at an unpredicted time. In these studies, like ours, predictions were violated. Why, then, did the ITC neurons not carry surprise signals? We propose that enhanced firing is not a general response to surprise but rather a specific response to image–image transitions assigned low probability on the basis of the transitional statistics of prior experience. This result is exactly what one would expect if transitional surprise signals in the ITC were representative of neural processes underlying the incidental and implicit learning of perceptual transitional statistics as demonstrated in human studies (18, 19).

## Methods

**Training Runs.** Each monkey received extensive initial exposure to six sequential pairs of images:  $A_1$ – $B_1$  through  $A_6$ – $B_6$  (Fig. 1A). Each trial consisted of the following successive events: fixation on a central spot (300 ms), image A at screen center (503 ms), image-free gap (18 ms), image B at screen center (507 ms), fixation on a central spot (300 ms), and delivery of reward. (Durations were measured with a photosensitive diode.) During a single run, over the course of 48 trials, each pair was presented eight times. The sequencing of conditions within a run was random subject to the constraint that each pair had to be presented once in each block of six successfully completed trials. The stimuli were digitized images of natural and man-made objects extracted from their original scenes to render them maximally distinctive. On a liquid crystal display monitor at a viewing distance of 32 cm, the horizontal or vertical axis of each image, whichever was longer, subtended 4° of visual angle (80 pixels along the vertical axis or 88 pixels along the horizontal axis). Monkey 1 completed training sessions on 42 d spanning three months. Monkey 2 completed training sessions on 13 d spanning 1 mo. The number of runs per day was 3.6 on average, with a minimum of one and a maximum of 13. Each monkey completed 102 runs over the course of the training period and thus saw each sequential image pair 816 times.

**Test Runs.** During neuronal data collection, the timing of events in each trial was the same as during training. The A image could be any of the six images presented in the leading position during training ( $A_1$ – $A_6$ ). The B image could be any of the six images presented in the trailing position during training ( $B_1$ – $B_6$ ). Each of six “trained” sequences ( $A_m B_n$ ,  $m = n$ ) was presented eight times. Each of 30 “untrained” sequences ( $A_m B_n$ ,  $m \neq n$ ) was presented once. Thus, a full run consisted of 78 trials. The sequence of trials was random.

**Reverse Runs.** In a subset of neurons characterized using the standard procedure, we also monitored responses to reverse displays. The same conditions were imposed as in the standard procedure, but the order of the images in each trial was reversed. For each condition  $A_m B_n$  in a forward run, there was a corresponding condition  $B_n A_m$  in a reverse run. The training, test, and reverse runs were the only contexts in which the monkeys saw the training images.

**Recording Sites.** Two adult rhesus macaque monkeys participated in the experiments (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 United States Public Health Service Guide for the Care and Use of Laboratory Animals. In each monkey, a surgically implanted cranial implant held a post for head restraint and a vertically oriented chamber through which the electrode could be introduced via a guide tube into the ITC along tracks forming a square grid with 1-mm spacing. Recording was carried out in the left hemisphere of monkey 1 and the right hemisphere of monkey 2. The location of recording sites relative to gross morphological landmarks was determined by extrapolation from MRI-visible fiducial markers placed at known locations within the chamber. The recording sites occupied the ventral bank of the superior temporal sulcus and the inferior temporal gyrus lateral to the rhinal sulcus at levels A16–19 mm relative to the interaural plane in monkey 1 and A13–16 mm in monkey 2.

**Database.** We monitored neuronal responses at 71 recording sites (42 in monkey 1 and 29 in monkey 2). Low-pass filtered traces from these sites formed the LFP database (see *SI Results* for further details). Eighty-one visually responsive neurons encountered at these sites (46 in monkey 1 and 35 in monkey 2) formed the neuronal database. Results from the two monkeys were closely similar (Fig. S6).

**Statistical Assessment of Transitional Surprise Signals.** For each neuron, we carried out an ANOVA with prediction status of the trailing image (predicted or unpredicted) and image identity (six levels) as factors and with firing rate 50–500 ms after stimulus onset as the dependent variable. The analysis was based on all five trials in which each trailing image was unpredicted and on a randomly selected subset of five trials out of the eight on which it was predicted. This selection ensured an identical number of observations for each combination of prediction status and image identity. A significant main effect of prediction status, with firing greater under the unpredicted than under the predicted condition, constituted evidence for a transitional surprise effect.

**Response Timing.** To measure the latency of the responses to the leading and trailing images and of the transitional surprise signal, we adopted an approach based on measuring time to half height. It might seem more reasonable to measure the time at which the signal becomes statistically significantly different from baseline. However, this approach confounds response strength with response timing because, with two rising responses identical except for strength, the stronger will cross statistical threshold first. For each neuron, we created four poststimulus time histograms representing the response to the leading stimulus (L); the response to the trailing stimulus in trials in which it was unpredicted (Tu); the response to the trailing stimulus in trials in which it was predicted (Tp); and the surprise signal ( $S = Tu - Tp$ ). Each histogram was binned at 1 ms and smoothed by convolution with a Gaussian function ( $\sigma = 10$  ms). We took baseline (B) as the average value in the range 0–50 ms after image onset. We took peak (P) as the maximal value in the range 60–500 ms after image onset. From these parameters, we computed the time at which the spike density function achieved half height:  $(B + P)/2$ . The time to half height for each signal was taken as the time at which the population average achieved half height. To determine whether the timing of two signals was statistically different, we applied a Wilcoxon test to the two distributions obtained by measuring time to half height in individual neurons or LFP sites.

**Quantification of LFP Responses.** The most sharply defined peaks in the response evoked by the trailing stimulus were a negative peak at roughly 200 ms and a positive peak at roughly 300 ms, which we refer to as the N200 and P300 peaks. We took as a measure of response strength the amplitude of the excursion from the N200 trough to the P300 peak. Despite some variations in waveform from site to site, a simple procedure for obtaining this measure yielded consistent results. We took the minimal voltage in a window 150–250 ms after stimulus onset as the value of the N200 trough and the maximal voltage in a window 250–350 ms after stimulus onset as the value of the P300 peak. We then took the difference between the maximum and minimum as the amplitude of the response. The peak-to-peak amplitude measured in these experiments was lower than in most other published studies (59, 60). This discrepancy is irrelevant to all statistical comparisons because they concern the relative rather than the absolute amplitude of the response.

**Transitional Surprise Index.** We characterized the prediction error effect with a scalar measure to compare its strength when images appeared in the order in which they were presented during training ( $A_m B_n$ ) with its strength when they were presented in the reverse order ( $B_n A_m$ ). The measure had to take account of the fact that the responses to the A images might differ in overall strength from the responses to the B images, either because of the arbitrary selection of images and arbitrary sampling of neurons or as a consequence of training. To factor out any such difference, we used an index normalized to response strength:  $(x - y) / (x + y)$  where  $x$  was the firing rate (or LFP amplitude) when the trailing image followed another image's training

partner and  $y$  was the firing rate (or LFP amplitude) when the trailing image followed its own training partner.

**ACKNOWLEDGMENTS.** We thank Karen McCracken for technical assistance and Tai Sing Lee, Jan Kalkus, Roma Konecky, Marvin Leathers, and Suchitra Ramachandran for their comments on the manuscript. This work was supported by Grants RO1 EY018620 and P50 MH45156 from the National Institutes of Health and by the Pennsylvania Department of Health's Commonwealth Universal Research Enhancement Program. Technical support came from National Institutes of Health Grants P30 EY08098 and P41RR03631.

- Ungerleider LG, Mishkin M (1982) Two cortical visual systems. *Analysis of Visual Behavior*, eds Ingle DJ, Goodale MA, Mansfield RJW (MIT Press, Cambridge, MA), pp 549–586.
- Buckley MJ, Gaffan D, Murray EA (1997) Functional double dissociation between two inferior temporal cortical areas: Perirhinal cortex versus middle temporal gyrus. *J Neurophysiol* 77:587–598.
- De Renzi E (2000) Disorders of visual recognition. *Semin Neurol* 20:479–485.
- Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856–867.
- Baker CI, Behrmann M, Olson CR (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci* 5:1210–1216.
- Jagadeesh B, Chelazzi L, Mishkin M, Desimone R (2001) Learning increases stimulus saliency in anterior inferior temporal cortex of the macaque. *J Neurophysiol* 86:290–303.
- Kobatake E, Wang G, Tanaka K (1998) Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80:324–330.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23:5235–5246.
- Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415:318–320.
- De Baene W, Ons B, Wagemans J, Vogels R (2008) Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learn Mem* 15:717–727.
- Messinger A, Squire LR, Zola SM, Albright TD (2001) Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proc Natl Acad Sci USA* 98:12239–12244.
- Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. *Nature* 354:152–155.
- 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.
- Mruczek REB, Sheinberg DL (2007) Context familiarity enhances target processing by inferior temporal cortex neurons. *J Neurosci* 27:8533–8545.
- Erickson CA, Desimone R (1999) Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J Neurosci* 19:10404–10416.
- Li N, DiCarlo JJ (2008) Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* 321:1502–1507.
- Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335:817–820.
- Perruchet P, Pacton S (2006) Implicit learning and statistical learning: One phenomenon, two approaches. *Trends Cogn Sci* 10:233–238.
- Saffran JR (2003) Musical learning and language development. *Ann N Y Acad Sci* 999:397–401.
- Kirkham NZ, Slemmer JA, Johnson SP (2002) Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 83:B35–B42.
- Abla D, Okanoya K (2009) Visual statistical learning of shape sequences: An ERP study. *Neurosci Res* 64:185–190.
- Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. *J Neurosci* 30:2960–2966.
- Bobes MA, Valdés-Sosa M, Olivares E (1994) An ERP study of expectancy violation in face perception. *Brain Cogn* 26:1–22.
- 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.
- James CE, Britz J, Vuilleumier P, Huetter C-A, Michel CM (2008) Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. *Neuroimage* 42:1597–1608.
- Qiu Y, Zhou X (2010) Perceiving the writing sequence of Chinese characters: An ERP investigation. *Neuroimage* 50:782–795.
- Ranganath C, Rainer G (2003) Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci* 4:193–202.
- Hauser MD, Newport EL, Aslin RN (2001) Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition* 78:B53–B64.
- Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13:403–409.
- Woloszyn L, Sheinberg DL (2009) Neural dynamics in inferior temporal cortex during a visual working memory task. *J Neurosci* 29:5494–5507.
- Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol* 70:909–919.
- Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, eds Black AH, Prokasy WF (Appleton-Century-Crofts, New York), pp 64–99.
- Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412:43–48.
- Carpenter GA, Grossberg S (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. *Comput Vis Graph Image Process* 37:54–115.
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836.
- Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66:241–251.
- Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis* 20:1434–1448.
- Rao RP, Ballard DH (1997) Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Comput* 9:721–763.
- Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Spratling MW (2008) Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience* 2(4):1–8.
- Spratling MW (2008) Predictive coding as a model of biased competition in visual attention. *Vision Res* 48:1391–1408.
- Spratling MW (2010) Predictive coding as a model of response properties in cortical area V1. *J Neurosci* 30:3531–3543.
- McMahon DBT, Olson CR (2007) Repetition suppression in monkey inferotemporal cortex: Relation to behavioral priming. *J Neurophysiol* 97:3532–3543.
- Perrett DI, Xiao D, Barraclough NE, Keysers C, Oram MW (2009) Seeing the future: Natural image sequences produce “anticipatory” neuronal activity and bias perceptual report. *Q J Exp Psychol (Colchester)* 62:2081–2104.
- Asplund CL, Todd JJ, Snyder AP, Gilbert CM, Marois R (2010) Surprise-induced blindness: A stimulus-driven attentional limit to conscious perception. *J Exp Psychol Hum Percept Perform* 36:1372–1381.
- Brockmole JR, Boot WR (2009) Should I stay or should I go? Attentional disengagement from visually unique and unexpected items at fixation. *J Exp Psychol Hum Percept Perform* 35:808–815.
- Howard CJ, Holcombe AO (2010) Unexpected changes in direction of motion attract attention. *Atten Percept Psychophys* 72:2087–2095.
- Kamin LJ (1969) Predictability, surprise, attention, and conditioning. *Punishment and Aversive Behavior*, eds Campbell BA, Church RM (Appleton-Century-Crofts, New York), pp 279–296.
- Pearce JM, Hall G (1980) A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87:532–552.
- Roesch MR, Calu DJ, Esber GR, Schoenbaum G (2010) All that glitters ... dissociating attention and outcome expectancy from prediction errors signals. *J Neurophysiol* 104:587–595.
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647.
- Williford T, Maunsell JHR (2006) Effects of spatial attention on contrast response functions in macaque area V4. *J Neurophysiol* 96:40–54.
- Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* 77:24–42.
- Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784.
- Lehky SR, Tanaka K (2007) Enhancement of object representations in primate perirhinal cortex during a visual working-memory task. *J Neurophysiol* 97:1298–1310.
- Anderson B, Sheinberg DL (2008) Effects of temporal context and temporal expectancy on neural activity in inferior temporal cortex. *Neuropsychologia* 46:947–957.
- Kaliukhovich DA, Vogels R (2008) Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cerebral Cortex* 21:1547–1558.
- Gill P, Woolley SMN, Fremouw T, Theunissen FE (2008) What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. *J Neurophysiol* 99:2809–2820.
- Anderson B, Mruczek REB, Kawasaki K, Sheinberg D (2008) Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cereb Cortex* 18:2540–2552.
- Kreiman G, et al. (2006) Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron* 49:433–445.