Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency

Tai Sing Lee[1,2], Cindy F. Yang[1], Richard D. Romero[1,2] and David Mumford[3]

[1] Center for the Neural Basis of Cognition and [2] Computer Science Department Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, USA

[3] Division of Applied Mathematics, 182 George Street, Brown University, Providence, Rhode Island 02912, USA

Correspondence should be addressed to T.S.L. (tai@cnbc.cmu.edu)

We report here that shape-from-shading stimuli evoked a long-latency contextual pop-out response in V1 and V2 neurons of macaque monkeys, particularly after the monkeys had used the stimuli in a behavioral task. The magnitudes of the pop-out responses were correlated to the monkeys' behavioral performance, suggesting that these signals are neural correlates of perceptual pop-out saliency. The signals changed with the animal's behavioral adaptation to stimulus contingencies, indicating that perceptual saliency is also a function of experience and behavioral relevance. The evidence that higher-order stimulus attributes and task experience can influence early visual processing supports the notion that perceptual computation is an interactive and plastic process involving multiple cortical areas.

The computation of object saliency is important for directing attention and for guiding eye movements during analysis of a visual scene. This computation is mediated by both bottom-up autonomous processes[1–2] and top-down attentional selection, which is a function of perception, experience and task demands[3]. As bottom-up and top-down processes are necessarily intertwined, it is difficult to cleanly separate their contribution to neural activity in visual cortex. Earlier single-unit studies in awake and anesthetized monkeys have implicated the primary visual cortex in mediating the bottom-up pop-out saliency computation of oriented bars and contrast gratings[4–6]. Recent studies using texture-contrast stimuli have suggested that top-down feedback may also be involved[7–10]. In these studies, however, perceptual saliency and orientation contrast were correlated, making it difficult to distinguish the relative contribution of feedforward or local mechanisms from that of inter-cortical feedback. Here we studied the influence of two top-down factors—higher-order perceptual inference and behavioral experience—in shaping the saliency computation in early visual cortex.

We used a set of stimuli including shape-from-shading images that have been used to demonstrate that parallel pop-out can occur with 'high-level' perceptual constructs[11–13]. When viewing such a stimulus (**Fig. 1a**), one readily perceives a convex oddball popping out from a background of concave distractors. Two-dimensional contrast patterns, such as the white above (WA) and white below (WB) stimuli (**Fig. 1b**), do not pop out as readily. The degree of perceptual pop-out seems to depend on the three-dimensional (3D) interpretation of the elements. The 3D shape inference is influenced by a single global interpretation of lighting direction[11]. If, instead, we interpret the scene (**Fig. 1a**) as being lit from below, the oddball can be seen as concave, with the distractors in the surround appearing convex. In both the lighting-

from-above and lighting-from-below scenarios, the pop-out percept is immediate. By contrast, the interpretation of lighting direction is ambiguous for stimuli being lit from the side, resulting in a pop-out percept that is much less compelling (**Fig. 1b**).

We have coded stimuli according to the way they were generated (shading on a 3D Lambertian sphere with lighting from different directions): lighting from above (LA), below (LB), left (LL) and right (LR) (**Fig. 1b**). Two 2D contrast elements, white above (WA) and white below (WB), were used as controls. Hence, there were six stimulus sets used in the experiments. The 3D elements (LA, LB, LL, LR) were defined by shading, not stereoscopically. Four conditions were used in association with each element type: singleton, oddball, uniform and hole (**Fig. 1b**). Human psychological studies[11–13] have shown that among these stimuli, perceptual pop-out saliency is strongest for the LA and LB stimuli, weaker for the LL and LR stimuli and weakest for the WA and WB stimuli, even though WA and WB have the strongest luminance contrast and should presumably elicit the strongest stimulus-driven response in V1. These observations point to the existence of neural responses that are correlated with perceptual pop-out saliency that can be dissociated from the strength of stimulus contrast. Higher-order perceptual areas probably participate in the computation of shape from shading, and these stimuli provide a way of testing whether or not the computation of perceptual saliency involves interaction between early visual areas and higher-order extrastriate areas.

RESULTS

A series of neuronal recording and behavioral testing experiments was conducted on two rhesus monkeys in nine stages. The data presented came from the analysis of 410 units in V1 and 138 units in V2. These included both well-isolated single units and multi-
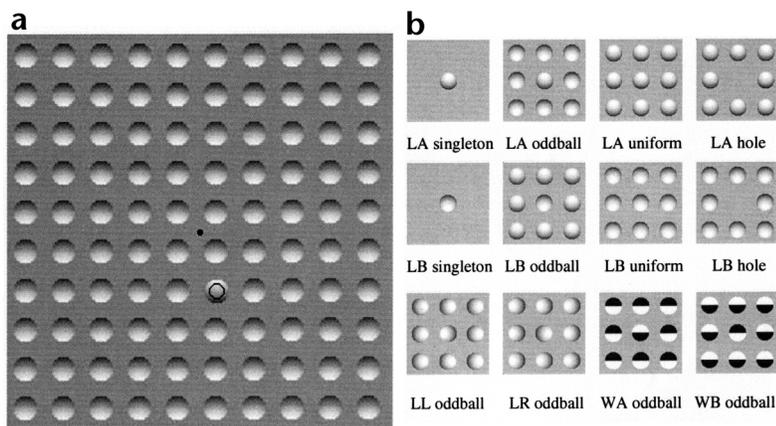
Fig. 1. Stimulus sets used in the experiment. (**a**) A typical stimulus display was composed of 10×10 stimulus elements. Each element was $1°$ visual angle in diameter. The diameter of the classical receptive field (RF) of a typical cell at the eccentricities tested ranged from $0.4°$ to $0.8°$. Displayed is the LA (lighting-from-above) oddball condition, with the LA oddball placed on top of the cell's receptive field (open circle). The solid dot indicates the fixation spot. (**b**) In stages 1, 3 and 5, each stimulus set had four conditions: singleton, oddball, uniform and hole. Shown are the iconic diagrams of all the conditions for the LA set and the LB set, as well as the oddball conditions for the other four sets. The center element in the iconic diagram covered the receptive field of the neuron in the experiment. The surround stimulus elements were placed outside the RF of the neuron. The comparison was between the oddball condition and the uniform condition; the singleton and the hole conditions were controls. The singletons measured the neuronal response to direct stimulation of the RF alone; the holes measured the response to direct stimulation of the extra-RF surround only.

ple units. Neural recording at all recording stages was done while monkeys were engaged in a fixation task (see Methods).

The stimulus was presented on the computer screen for 350 ms in each trial. During recording, the test stimuli were not relevant to the monkey's behaviors. The fixation task was chosen to allow direct comparison between the neural responses at the different stages, before and after behavioral training. It also removed the confounding enhancement effect known to be associated with saccadic eye movement toward the receptive fields of the neurons in early visual areas[14].

Only V1 cells that (i) showed a marked response to at least one of the shape-from-shading stimuli, and (ii) produced a negligible response to all hole stimuli were included in our analysis. In the experiment, the probe stimulus was displayed to the receptive field of the cell with identical stimuli or different stimuli in the surround (**Fig. 1a**). Receptive fields of the cells tested in this study were located at between $1.5°$ and $4.0°$ eccentricity in the lower visual field. The size of the V1 receptive fields ranged from $0.4°$ to $0.8°$ in diameter, and each stimulus element was $1°$ in diameter. The center-to-center distance between adjacent elements was $1.75°$.

The receptive fields of V2 neurons were generally larger than those of V1 neurons, ranging from $0.8°$ to $2.0°$ in diameter. About 30% of the encountered V1 neurons were discarded because they did not respond to any of the shape-from-shading stimuli. On the other hand, all the V2 neurons encountered were included in the analysis because they typically responded to these stimuli (though many of them also responded to the hole stimuli).

Naive state
In the first stage, we recorded from 30 V1 units from monkey A and 55 V1 units from monkey B. The singleton condition elicit-

ed the strongest response relative to the other conditions. That is, when stimulus elements were added to the surround of the cells' classical receptive fields, the responses were suppressed from the very beginning, indicative of the immediate nature of the competitive lateral inhibition from the surround (**Fig. 2a** and **b**). Moreover, the initial responses (40–100 ms after stimulus onset) of V1 neurons to stimuli WA and WB were about 1.5–2 times stronger than were their responses to the shape-from-shading stimuli, underscoring the cells' default sensitivity to stimulus contrast. These observations were true for all the neural data recorded from V1 across the various stages for both monkeys.

The enhancement of a neuron's response to the oddball condition relative to the uniform condition is called the pop-out response. We quantified it using a normalized pop-out modulation ratio, given by $(R_o - R_u)/(R_o + R_u)$ where R_o and R_u are the responses of the neuron in a certain time window to the oddball and uniform conditions, respectively. The modulation ratio reported was computed in the time window 120–320 ms after stimulus onset (for comparisons of modulations across different time windows, see **Supplementary Note** (**Figs. 7–12**) online). The mean pop-out modulation ratios of monkey A's V1 neurons in stage 1 showed marginally significant ($P < 0.05$, t-test) pop-out responses to the LA and LB stimuli, but not to the other stimuli (**Fig. 2c**). Neurons in monkey B showed essentially the same pattern: a marginally significant ($P < 0.05$) pop-out response to LB and not to the other stimuli (**Fig. 3a–c**). It seemed that when the monkeys were naive to the behavioral relevance of the stimuli, their V1 neurons did not exhibit significant sensitivity to the 3D shape attributes in the stimuli.

Behavioral relevance
To assess the impact of behavior on perceptual processing, we trained the monkeys to perform an oddball detection task in stage 2 (see Methods). They learned the task rapidly but were each trained for 15 sessions (1,200 trials per session) to attain proficiency. We separated the stimuli into three pairs in descending order of perceptual saliency—LA and LB, LL and LR, WA and WB—and compared the animals' performance among these groups in a pair-wise fashion. Consistent with observations in humans[11–13], monkey A's reaction time was significantly shorter for the LA and LB oddballs than it was for the LL and LR oddballs ($P < 0.007$), which in turn was significantly shorter than for the WA and WB oddballs ($P < 0.0003$) (**Fig. 2g**). Monkey A's performance accuracies (percentage correct) for the 3D stimuli (LA and LB, LL and LR) were significantly better than they were for the 2D stimuli (WA and WB) ($P < 0.01$ for all 3D versus 2D comparisons) (**Fig. 2h**). The performance of monkey B was similar. Accuracy was much better for LB than for LL and LR ($P < 10^{-5}$) and was somewhat better for LL and LR than for WA and WB ($P < 10^{-5}$) (**Fig. 3h**). Reaction time was shorter for LB than for all the other stimuli ($P < 10^{-5}$) (**Fig. 3g**). The main difference between the two monkeys was that whereas monkey A performed equally well for both LA and LB, monkey B showed asymmetrical performance for the two stimuli (strongly preferring LB at the expense of LA).

Fig. 2. Monkey A's V1 neural responses and behavioral performance in stages 1–5. (**a**, **b**) Temporal evolution of the normalized population average response of 30 V1 units to the LA set (**a**) and to the WA set (**b**) in stage 1. Each unit's response was smoothed by a running average within a 15 ms window. Then the responses were averaged across the population. A very small difference (pop-out response) was seen between the population average response to the oddball condition and that to the uniform condition in the LA set. No pop-out response was seen in the WA set. Note that in these plots, as well as in other temporal response plots presented, the firing rate at a given time was normalized against the maximum instantaneous response of the population to any stimulus of the tested sets at each stage, which typically was the response to either the WA or WB stimuli. The neural response to the shape-from-shading stimuli can thus be gauged relative to the neural response to the WA and WB contrast stimuli. (**c**) Mean pop-out modulation ratios of 30 units for all six stimulus sets in stage 1. Pop-out enhancements were significant for LA (*P* = 0.011) and LB (*P* = 0.007), but not for the other stimuli. The error bars represent the standard errors of the means (s.e.m.). (**d**, **e**) Temporal evolution of the normalized population average response of 45 V1 units to the LA set (**d**) and WA set (**e**) in stage 3. Significant pop-out response was seen in LA (as well as in LB, LL and LR) starting at 100 ms after stimulus onset. No pop-

out response was seen for WA or WB. (**f**) Scatter plot of modulation ratios of the neuronal population for LA and LB pop-out responses in stage 3. The population means of the modulation ratios for LA and LB were both significantly positive. (**g**, **h**) Behavioral performance of monkey A in detecting the different oddballs from 15 testing sessions at the end of stage 2. In general, stimuli with higher pop-out modulations were associated with shorter reaction times and higher accuracy (percentage correct). (**i**) Mean pop-out modulation ratios of 45 units for all six stimulus sets in stage 3. Pop-out enhancements were highly significant for stimuli LA, LB, LL and LR (*P* = 10^{-6}, 10^{-6}, 0.0045, 10^{-4}) but not for WA and WB. (**j**, **k**) Consequence of the LB-biased training. Reaction time and percentage correct in the behavioral performance of monkey B, measured in five sessions at the beginning of stage 6, reflected an improvement in LB oddball detection. (**l**) Mean population modulation ratios of 56 neurons of monkey A at stage 5 subsequent to LB-biased training. A strong asymmetry in pop-out modulation was seen for LB over the other stimuli, in parallel to the change in behavior.
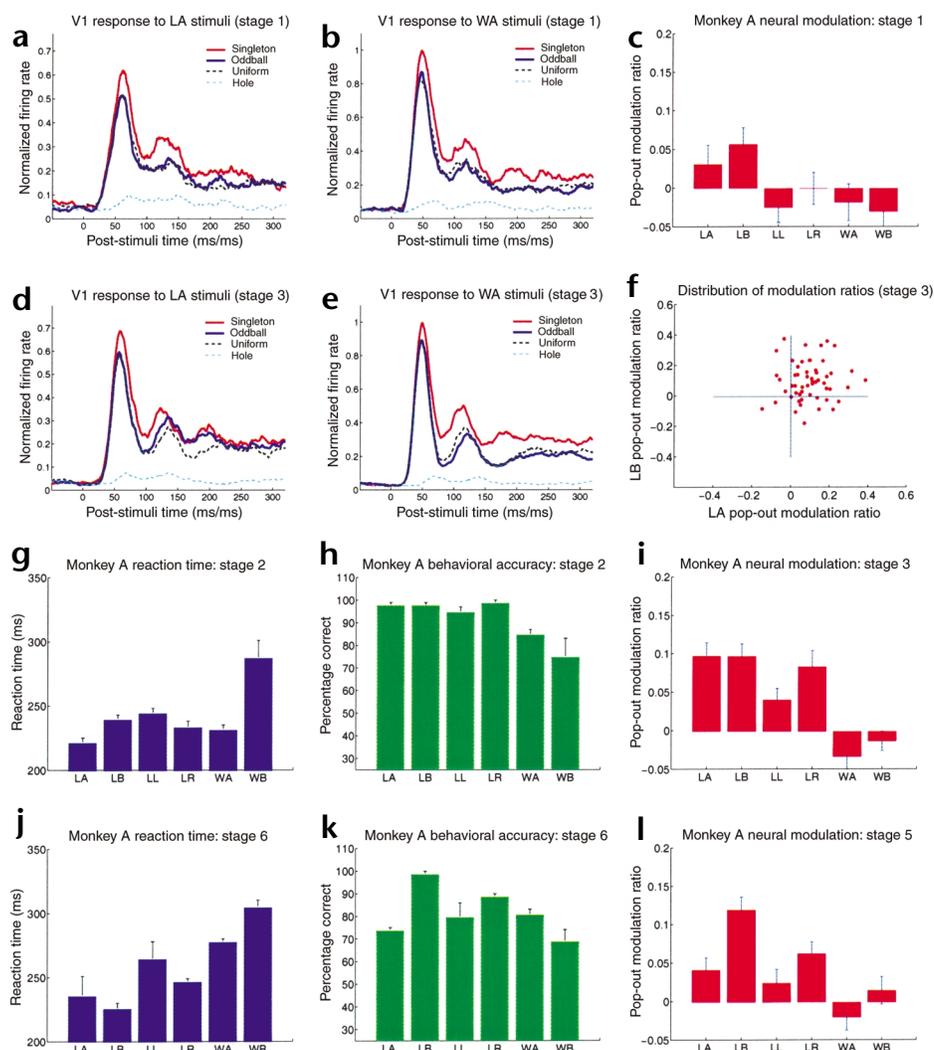
We recorded from 45 V1 units from monkey A and 47 V1 units from monkey B in stage 3 after the behavioral training, again using the fixation task paradigm. Relative to stage 1, there was a significant increase in V1 pop-out responses for both monkeys to most of the shape-from-shading stimuli. The pop-out responses were significant in both monkeys (*P* < 0.005 monkey A, *P* < 0.01 monkey B) starting at about 100 ms after stimulus onset (**Figs. 2d** and **3d**). There were no significant pop-out responses for WA and WB (**Figs. 2e** and **3e**). The increases in population mean modulation were significant for all the shape-from-shading stimuli relative to stage 1 (monkey A: *P* < 0.03 for LA, LL, LR, *P* = 0.07 for LB; monkey B: *P* < 0.001 for all) but not for WB and WA (**Figs. 2i** and **3i**). The data indicate that experience with using the stimuli activated, or enhanced, the neural pop-out effect in

V1 neurons for these higher-order stimuli. Overall, the monkeys' behavioral performance in stage 2 and their V1 neural pop-out responses in stage 3 were stronger for 3D stimuli than they were for 2D stimuli.

Adaptability to changes

There were, however, individual differences in the behavioral performance between the two monkeys, evidenced by a parallel difference between their neural pop-out modulations. Specifically, whereas monkey A's behavioral performances to LA and LB stimuli were roughly the same, monkey B had a much stronger preference for LB over LA—an asymmetry that was mirrored in their pop-out responses. Perhaps monkey B had developed a habit of looking for the LB pop-out target in a field
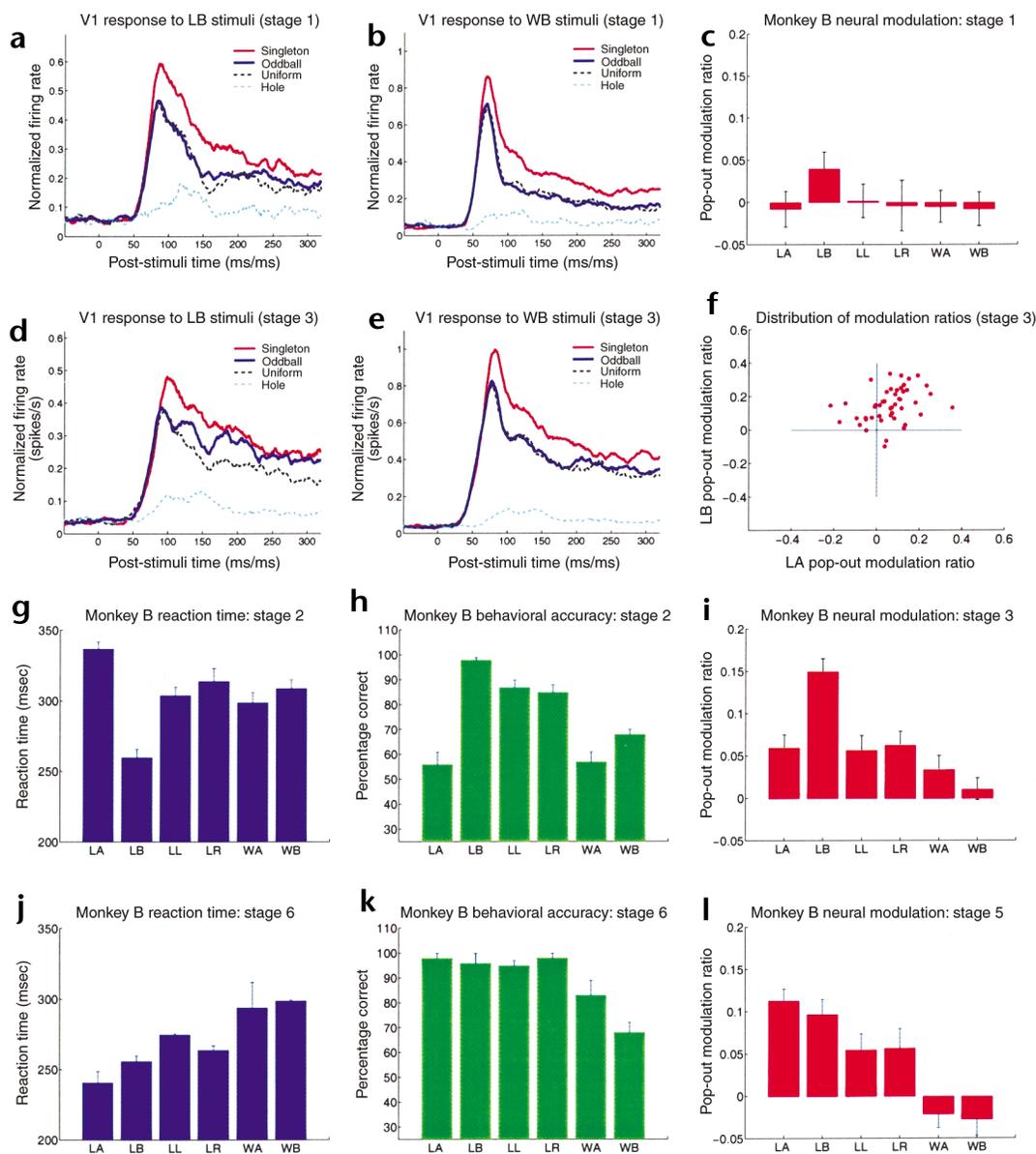
Fig. 3. Monkey B's V1 neural responses and behavioral performance in stages 1–5. Figure parts are as in **Fig. 2** except that the responses to the LB set and to the WB set are shown here as examples, because the LB oddball evoked the strongest pop-out response in monkey B. Fifty-five V1 units were recorded in stage 1, 47 in stage 3, and 47 in stage 5. The only stimulus that evoked significant pop-out responses in stage 1 was the LB oddball. The LB oddball also evoked the strongest significant pop-out response in stage 3 ($P < 10^{-7}$), followed by LA, LL and LR ($P < 0.01$, 10^{-4} and 10^{-4}, respectively). Pop-out was weak for WA and insignificant for WB. The positive shifts in modulation between stages 1 and 3 were significant for all the shape-from-shading stimuli ($P < 10^{-2}$ in all cases).

of LA distractors, resulting in a facilitation of response to the LB stimuli at the expense of the LA stimuli. This suggests that the pop-out response patterns, and hence the perceptual saliency of the stimuli, may be a function of individual cognitive strategy or behavioral experience.

To test this hypothesis, we carried out a biased training experiment to modify the monkeys' behavior by manipulating the frequency of occurrence of the different oddball stimuli in stage 4. In 30 training sessions (1,200 trials per session), monkey A practiced solely on LB oddball detection. A preference for LB was developed in the monkey's behavior as measured in five sessions at the beginning of stage 6. This was accompanied by a parallel

change in the pop-out modulation patterns as measured in stage 5: the pop-out response became significantly stronger for LB oddball over all the other stimuli ($P < 0.01$).

We subjected monkey B to the opposite biased training to offset the original asymmetry in favor of LB. Thirty training sessions with only LA oddball stimuli were carried out, interleaved with 15 sessions in which oddballs of all stimulus types were presented with equal frequency. Combined, the presentation frequency of LA oddball relative to any other type of oddball was 12:1. We found that the change in stimulus contingencies did remove and even reverse the asymmetry in monkey B's behavioral performance. V1 neural pop-out responses followed suit as

Fig. 4. Stage 7: spatial extent and durability of the effect. (**a–d**) There were four conditions for each of the six stimulus sets tested in stage 7. Shown here are the oddball ON, oddball NEAR, oddball FAR and oddball absent (uniform) conditions of the LA set. In the actual display, the full screen was covered by distractors as in **Fig. 1a**. The solid black dot is the fixation spot. The open black circle indicates the spatial extent of the receptive field of the tested neuron. These stimuli are grouped in the LA set because the RF is being probed by an LA stimulus element. (**e**) Population mean modulation ratios of the responses to the three oddball conditions computed against the response to the uniform condition for 30 V1 units of monkey A. Significant pop-out response was seen only when the oddball was exactly on the receptive field. The absence of pop-out response in both the NEAR and FAR conditions suggests that the effect was localized spatially in V1 at the location of the oddball. (**f**) Population mean neural pop-out modulations were correlated between stages 5 and 7 in monkey A. Each data point was the pop-out modulation for one of the six stimulus sets. The correlation of modulation before and after a long recess suggests that the stimulus-specific pop-out modulation was stable over time, or encoded in memory. (**g**) Population mean modulation ratios of three oddball conditions computed against the uniform condition for 25 V1 units in monkey B. The oddball ON condition elicited the strongest pop-out response in the V1 units. Slight and marginally significant enhancements of some of the shape-from-shading stimuli. (**h**) Similar to (**f**); population neural pop-out modulations were significantly correlated between stages 5 and 7 in monkey B.
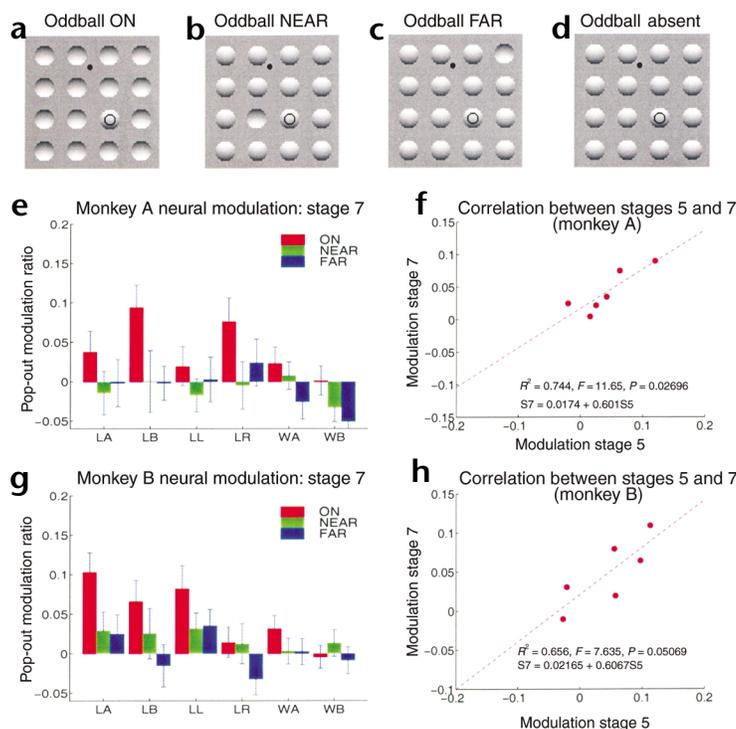

well (**Fig. 3j–l**). We thus showed that the pop-out modulations in V1 for the different stimuli could be manipulated with changes in stimulus contingencies.

Spatial extent and durability

Is this enhancement effect confined to the location of the oddball stimulus, or does it extend to nearby stimuli as well? To answer this question, we carried out another recording experiment in stage 7 after a long recess (stage 6), during which mon-

keys A and B were not tested behaviorally or neurophysiologically for at least one and two months, respectively. In stage 7, we tested the monkeys on the fixation task with the following four conditions for each of the six stimulus sets: oddball ON, NEAR, FAR (from the receptive field), or absent (**Fig. 4a–d**). In the conditions of this set, an LA stimulus element was always placed on the receptive field. In the NEAR condition, an LB oddball was placed 1.7° away from the receptive field stimulus (center-to-center). In the FAR condition, an LB oddball was placed 3.6° away.
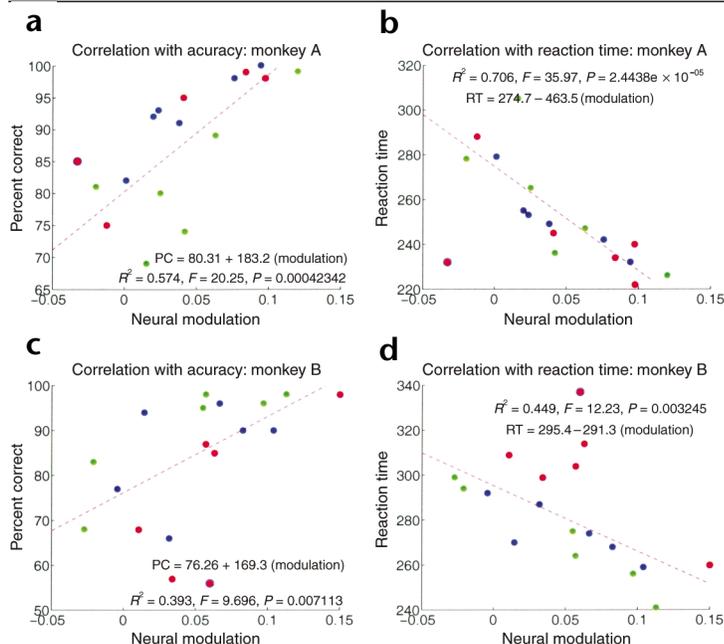
Fig. 5. Correlation between behavioral performance and V1 neural modulation across multiple stages. Behavior performance measurements (percentage correct and reaction time) in stage 2 (red), 6 (green) and 8 (blue) were paired with neural pop-out modulation data in stage 3 (red), 5 (green) and 7 (blue). Each pair of stages produced six data points, corresponding to the six stimulus types. Eighteen points are shown in each graph relating a behavioral measure with neural pop-out modulation. Reaction time and percentage correct was regressed on the pop-out modulation independently. A linear regression line, with equation and statistical significance, is shown in each plot. An outlier, which was >2.5 standard deviations away from the regression line, was discarded in each graph (red dot with blue outline). The outlier could have arisen from interference resulting from other top-down influences. R^2 indicates the portion (percentage) of the variance in the specific behavioral measure that can be explained by the neural pop-out modulation. The negative correlation between reaction time and neural modulation was highly significant for both monkeys (statistics and fitted equation are shown on graphs), as was the positive correlation between accuracy and neural modulation. The correlations remained significant when the outlier was included and when the modulations were evaluated in the following time windows: 100–180 ms, 120–250 ms and 120–320 ms (see **Supplementary Note** (**Figs. 11–12**) online for further details).
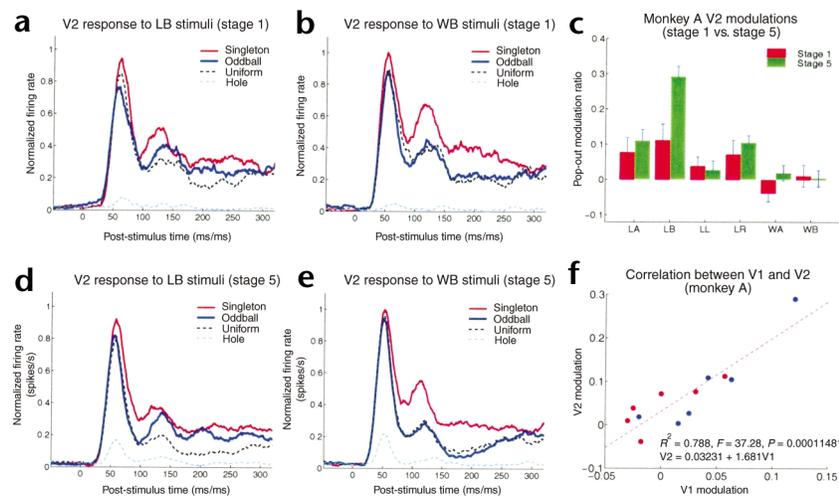
Fig. 6. V2 neural pop-out response in monkey A. Data are from 15 V2 units in stage 1 (**a**, **b**) and 22 V2 units in stage 5 (**d**, **e**). Temporal responses of the V2 population showed significant pop-out response for shape-from-shading stimuli starting at 100 ms, even in stage 1. Response to LB shown as example. No pop-out response was seen for the WA and WB stimuli before or after behaviors. Note the change in the scale of modulation axes relative to the graphs for V1 modulation. (**c**) Neural modulations in stages 1 and 5. Pronounced increase in enhancement was seen in LB subsequent to LB-biased training. (**f**) V2 modulations were correlated to V1 modulations, but were twice as strong.

The oddball targets were positioned at roughly the same eccentricity away from the fovea as that of the receptive field location.

The respective pop-out modulation ratios for the two monkeys in stage 7 consistently showed that the neural pop-out effects were still present after 1–2 months (**Fig. 4e** and **4g**). The stimulus-specificity of the pop-out modulation was stable over time, significantly correlating with that of stage 5 (**Fig. 4f** and **h**). More importantly, when the oddball appeared near the receptive fields of the neurons, no significant pop-out effect was seen in the neurons of monkey A across all stimulus sets (**Fig. 4e**). Neurons in monkey B showed a slight enhancement in some oddball NEAR and FAR conditions (**Fig. 4g**), but the magnitude of the enhancement was much smaller in these conditions than it was in the oddball ON conditions. These results show that the pop-out signals were stable over time and were spatially localized. Subsequently in stage 8, we tested the behavioral performance of both monkeys using the oddball detection task and found the behavior was again correlated with the neural responses in stage 7.

Perceptual saliency
The accurate performance of the monkeys in detecting shape-from-shading oddballs in the various behavioral testing stages where chance rate was 25% (4 target locations), indicates that the monkeys were most likely perceiving the stimuli rather than guessing randomly. To confirm that the neural pop-out signal was truly a physiological measure of subjective perceptual saliency, we performed a regression analysis between the neural pop-out modulation ratio from each recording

stage and behavioral performance in the stages immediately before or after that recording stage. We grouped the data into three pairs: stage 2 behavior + stage 3 neural response, stage 6 behavior + stage 5 response, and stage 8 behavior + stage 7 response. A significantly positive correlation between the V1 neural pop-out modulation signal and performance accuracy (**Fig. 5a** and **c**), as well as a significant negative correlation between the pop-out modulation signal and the reaction time (**Fig. 5b** and **d**) were found for both monkeys, suggesting that the neural pop-out signal could be considered a neural correlate of perceptual saliency.

Neural activity in V2
We recorded from a total of 138 V2 neurons from the two monkeys combined, to ascertain the role of cortical interaction in mediating these effects. Most V2 neurons were recorded at roughly the same eccentricities as the V1 neurons were. Several major differences and similarities in the neural pop-out responses between the two areas were seen (**Figs. 6** and **7**). First, the basic patterns of V2 responses were very similar to those of V1 responses. In both areas, the singleton stimulus elicited the strongest response, and the suppressive effect of the surround was immediate. However, V2 neurons had larger receptive fields at the same eccentricity and tended to respond more strongly to the hole stimuli than did V1 neurons. Second, the V2 neural pop-out
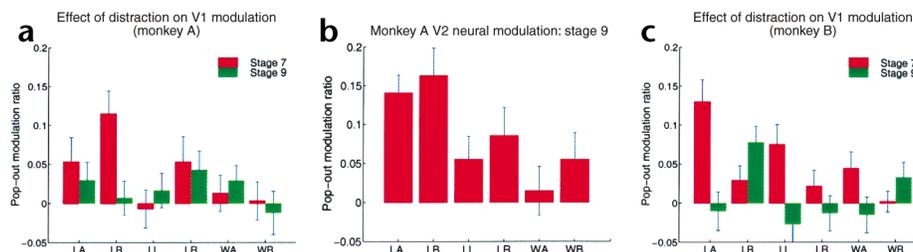
Fig. 7. V2 neural pop-out response in monkey B. Data are from 17 V2 units in stage 1 (**a**, **b**) and 24 V2 units in stage 3 (**d**, **e**). Significant pop-out response was seen in the population temporal response for LB stimuli but not for WB stimuli. (**c**) Pop-out modulation ratios for the six stimulus types in stages 1 and 3. Behavioral training in stage 2 apparently hastened the emergence of pop-out response and increased the neural modulation for some stimuli. The mean pop-out modulation ratios for 30 V2 units from stage 5 (not shown) were 0.14, 0.15, 0.16, 0.13, 0.03 and 0.008 for LA, LB, LL, LR, WA and WB, respectively. (**f**) V2 modulations were correlated with V1 modulations, but were stronger.
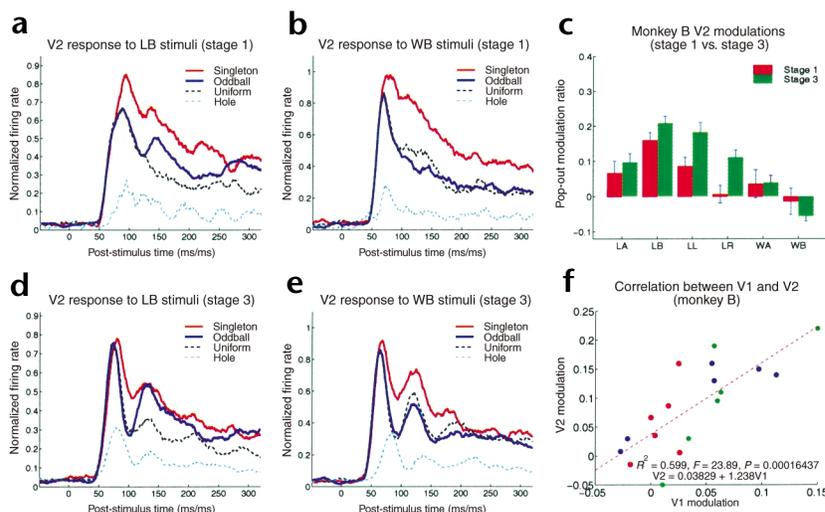
modulation in stage 1 was significantly positive for nearly all the shape-from-shading stimuli (monkey A, $P < 0.05$ all cases; monkey B, $P < 0.002$ for LA, LB, LL, not significant for LR) but was absent for the 2D contrast patterns (WA, WB) (**Figs. 6a–c** and **7a–c**). This indicates that V2 neurons were sensitive to the shape-from-shading pop-out before behavioral training. Subsequent to behavioral training, the pop-out modulation in V2 became stronger (**Fig. 6c** and **7c**). Third, the stimulus-specific patterns of pop-out modulation between V1 and V2 were correlated across the multiple stages (**Fig. 6f** and **7f**). Hence, V2's pop-out responses were also correlated with perceptual saliency. V2 neural pop-out modulation for shape-from-shading stimuli was

Fig. 8. Stage 9: effect of distraction. (**a**, **b**) Monkey A's V1 population mean modulation ratios of stage 7 (30 units) and stage 9 (35 units) were compared to illustrate the effect of distraction in stage 9. Significant attenuation of pop-out response was seen in LB ($P < 10^{-6}$) in particular. The effect of distraction thus was most evident in undermining the effect of the LB-biased training in stage 4. V2 population mean modulation ratios remained significantly positive for the shape-from-shading stimuli. (**c**) A comparison between monkey B's V1 modulation ratios of stage 7 (25 units) and stage 9 (40 units) shows that pop-out responses were eliminated for most stimuli except LB when attention was forcefully drawn away from the receptive field. Note that these modulation ratios were computed within the 100–250 ms window after stimulus onset. A shorter window was used because the monkey started reacting to the target (which appeared at variable times starting at 100 ms) in the distraction task. The earliest reaction time was ~280 ms after stimulus onset.

much stronger (1.5–2.0 times) than that of V1. Significantly more (2–5 times) individual V2 neurons had positive pop-out responses to the shape-from-shading stimuli than did V1 neurons across the various stages (**Supplementary Table 1**). Fourth, the initial response (40–100 ms) of V1 neurons to shape-from-shading stimuli was considerably weaker than it was to the 2D contrast patterns, but the initial responses of V2 neurons to the two types of stimuli were about equal. This indicates that V2 neurons may have greater feed-forward sensitivity to 3D shapes, compensating for the weaker luminance contrast of the shape-from-shading stimuli. Finally, the latency of the pop-out responses in V2 in stages 3 and 5 was estimated to be ~100 ms after stimulus onset, similar to that of the V1 pop-out responses (**Fig. 2d** versus **Fig. 6d**, and **Fig. 3d** versus **Fig. 7d**). This suggests that V1 and V2 might be tightly coupled in this computation, and that the processes are likely to be interactive and concurrent in nature.

Plasticity and attention

The persistence of the stimulus-specific pop-out modulation patterns over months suggests that the behavioral relevance of the different stimuli was encoded in memory to influence neural processing at V1. To assess the possible role of attention in mediating the extrastriate cortical feedback, we conducted a 'divided attention' experiment in stage 9. The monkeys were trained to first fixate on a red dot during stimulus presentation. Another small, faint red dot (0.1° diameter) would flash briefly (3–4 refresh frames, or 54–72 ms) in one of the three designated locations far away from the receptive field at a random interval (between 100–300 ms after stimulus onset, at 50 ms intervals). These designated locations were constant across all sessions and were in the upper visual field (4° visual angle away from the fovea), whereas the receptive fields were in the lower visual field. The monkeys had to pay attention to the three designated locations to perform successfully on the task, which was to make a saccade to the dot within 300 ms. The task was considered attentionally demanding, as the monkeys performed at only 50% correct. We recorded the responses of 35 V1 units and 30 V2 units from monkey A, and those of 40 V1 units from monkey B, to the oddball and the uniform conditions of the six sets of stimuli. We found that the pop-out responses in V1 became insignificant for some stimuli when attention was forcefully drawn away from the recep-

tive field location (**Fig. 8a** and **c**). Most strongly affected were the stimuli for which the monkeys were trained to develop a preference in stage 4. LB pop-out for monkey B, which was its original favorite, seemed to have been spared. V2 was less affected, as the pop-out responses for LA, LB and LR remained relatively strong and significant (**Fig. 8b**). The data indicate that the availability of attentional resources, or the lack of distraction, is important for the emergence of this higher-order perceptual pop-out saliency signal at the level of V1, but not in V2.

Eye-movement artifacts

The pop-out effects could potentially have arisen as a result of micro-saccades and/or eye tremors occurring during fixation[15]. When a strong pop-out target appeared, the monkeys might have made small eye movements toward the target, moving the stimulation across the receptive field and thereby enhancing the response. This effect would also have been proportional to the perceptual saliency of the target. We analyzed all the eye movement data and ruled out eye tremors as a significant cause of the pop-out effect (see **Supplementary Note** (**Figs. 1–6**) online).

DISCUSSION

Why shape-from-shading stimuli (which presumably require processing by higher-order perceptual areas) pop out 'pre-attentively' has been a long-standing mystery. Here we found that long-latency neural signals in V1 and V2 were correlated with behavioral performance in monkeys. This correlation suggests that the long-latency signal might be a neural correlate of the subjective perceptual pop-out saliency that results from shading in the visual stimulus. The findings that V2—but not V1—neurons (i) responded vigorously in their initial phase of responses to the shape-from-shading stimuli and (ii) showed a significantly high degree of sensitivity to the shape-from-shading stimuli in the later part of their responses (even before behavioral training) together suggest that V2 may be the first cortical area that is sensitive to or provides the primitives for encoding 3D surface shape[16] (see also F. T. Qiu *et al. Soc. Neurosci. Abstr.* **26**, 593.2, 2000). V1 neurons, in contrast, only showed a marginal sensitivity to the most salient shape-from-shading pop-out stimuli before behavioral training. The monkeys became significantly more sensitive to these stimuli only after they had used the stim-

uli in their behavior. These findings suggest that the 3D shape sensitivity in V1 may be mediated by recurrent feedback connections from V2 and/or other extrastriate areas. **Supplementary Table 1** shows the numbers of V1 and V2 units recorded in each stage and the percentage of neurons individually showing statistically significant pop-out responses for each type of stimuli. The magnitude of pop-out response and the percentage of neurons showing a significant effect were markedly greater in V2 than in V1, indicating that the correlation between neural activity and subjective perception increases along the visual hierarchy. This increase is consistent with an earlier finding on the neural correlates of perception as revealed by binocular rivalry[17].

Our data also suggested that perceptual saliency was not static, but dynamic and malleable, contingent on the animal's experience and on the behavioral relevance of the stimuli. The stimulus-specific pattern of the modulation was stable over months until it was changed by new experience. Both the persistence and the adaptability of the effect indicate that the behavioral relevance of the stimuli must have been encoded in memory, exerting an influence over early visual processing. Given that the observed effect was not restricted to the retinotopic location of the target during training, we suspect that the plasticity components of the effect were distributed over multiple memory and perceptual areas or in the feed-forward/feedback connections between cortical areas, although changes in the V1 intrinsic circuitries were also possible[18–20].

What is the mechanism underlying these changes? One possibility is covert attention, which might be attracted by the salient pop-out target automatically. The attenuation of pop-out signals when attention was diverted away from the receptive field location implicated the involvement of attention, particularly for those stimuli for which the animals had developed a preference during biased training. Because of the late onset of the saliency effect, this attention was likely triggered by the input stimulus. Stimuli with stronger perceptual saliency (as determined by higher-order brain areas) would attract more attention. Previous studies have manipulated top-down spatial attention and top-down feature attention[21–23], but our study shows a potential interaction between top-down perceptual inference and attentional allocation processes and the parallel computations in the early visual areas. We propose that the input stimulus generated an initial representation in V1, which then initiated a cascade of perceptual computations across multiple extrastriate visual areas for target selection and for deduction of shape from shading and figure–ground and target selection. This higher-order perceptual and attentional processing interacts with the early visual processing to determine the perceptual saliency of the stimuli, and thereby modifies the representations across the whole visual hierarchy. The observed phenomena therefore reflect changes in both covert spatial attention and object attention in response to the input stimuli.

The idea that attention is involved in pop-out computation seems to be at odds with the conventional notion that pop-out is necessarily a pre-attentive process. This conventional idea, however, has been challenged by recent psychological studies[24–25] showing that attention may be critical for the covert detection, and even the overt perception, of pre-attentive stimulus features. Further, the interactions between bottom-up and top-down processes have been shown to be modifiable by perceptual training[25]. Our findings are consistent with these psychological observations, suggesting that there is a tight coupling between the parallel pop-out computation and the top-down perceptual and attentional processes.

What is the role of V1 in this computation? We have proposed elsewhere that V1 serves as a 'high-resolution buffer' for visual processing[9]. As only V1 neurons provide an explicit representation for precise encoding of orientation and spatial information, higher-order perceptual inference involving fine details, curvilinear geometry and spatial precision would necessarily engage V1 in their computation. The effects of such higher-order perceptual computations should therefore be reflected in the later part of V1 activity. Several recent physiological studies support this conjecture[26,27]. As feedback from the extrastriate cortex tends to be diffuse and broad in spatial extent, V1 could play an important role in localizing the pop-out target by actively sharpening the pop-out response spatially using its well known lateral inhibition mechanism.

Thus, the higher-order pop-out saliency effect seen here is probably sub-served by the same mechanisms that mediate the bottom-up pop-out effect for oriented bars[3] and the orientation contrast effect for sine wave gratings seen in anesthetized monkeys in numerous earlier studies[4–6]. The orientation contrast effect could be supported purely by intra-cortical lateral inhibition mechanisms in V1[28,29]. In awake behaving monkeys versus anesthetized ones, orientation contrast effects have been found to be enhanced in both magnitude and spatial extent, resulting in the so-called figure–ground effect[7–10,30,31]. Here we suggest that many or all of these phenomena may be interpreted as parts of the same set of bottom-up and top-down mechanisms for computing perceptual saliency. Notably, the time frame (100–150 ms) in which the target selection signal emerges in the frontal eye field during a visual search task[32] is roughly the same as the time frame for the emergence of the higher-order pop-out signals in both V1 and V2. Taken together, the present findings indicate that the representation of perceptual saliency of objects in a visual scene is distributed across multiple cortical areas and that its computation is interactive in nature, involving the concerted action of many areas in the brain[9,33–37].

METHODS

Recording technique. Recordings were made transdurally with epoxy-coated tungsten electrodes through a surgically implanted well overlying the operculum of area V1 of the awake behaving monkey[9]. A protocol covering these studies was approved by the Institutional Animal Care and Use Committee of Carnegie Mellon University, in accordance with Public Health Service guidelines for the care and use of laboratory animals. The neurons were isolated on the basis of spike heights using a window discriminator. The cells' classical receptive fields (RFs) were mapped by a small oriented bar. Based on the depth of penetration, most V1 cells studied were estimated to be cells in layers 2 and 3 of V1. V2 cells were drawn from arbitrary layers of V2. Eye position was measured using the scleral search coil technique and sampled at 200 Hz during the experimental sessions.

Fixation task. Throughout recording stages 1, 3, 5 and 7, a fixation task was performed by the monkey. In each trial, while the stimuli were presented on the screen for 350 ms each, the monkey was required to fixate on a red dot, maintaining gaze within a fixation window ranging from 0.5° to 0.65° of visual angle in diameter. When the presentation was complete, the fixation dot disappeared, a second red dot appeared at a different location, and the monkeys were required to make a saccade to it in order to receive a juice reward. The probe stimulus (the center stimulus in each of the iconic diagrams in **Fig. 1b**) was placed on the receptive field of the cell. The position of the second dot target was not correlated with the stimulus, and hence the test stimulus was irrelevant to the monkeys' behaviors in the recording sessions. Twenty-four conditions were tested in each session: six stimulus sets, each with four conditions. Each condition was repeated 12–15 times for each cell. The presentation of the conditions was randomly interleaved. The distraction task in stage 9, as described in the text, was a variant of the fixation task.

Oddball detection task. During behavioral training and testing in stages 2, 6 and 8, the monkeys performed an oddball detection task. In each trial, the oddball target was randomly drawn from the six basic stimulus types and placed at one of four random locations, distributed over the four quadrants of the visual field and at 4° eccentricity away from the fovea. The oddball was embedded in a field of distractors, each of which was the reflected image of the oddball. The monkeys had to make a saccade to the oddball location to complete the trial correctly. The chance rate was therefore 25% correct. No reward was given for incorrect trials. The order of presentation of the stimuli was randomly interleaved in each block. At least five sessions were carried out in each testing stage to assess behavior. Fifty trials were tested for each condition per session. During the biased training in stage 4, the same protocol was used except that the frequency of oddball occurrence was a function of the stimulus types.

Data analysis. The *t*-test, ANOVA for two groups, was used throughout to test for significant differences between means, either at the individual neuron level or at the population level. For an individual neuron, we compared response distributions to the oddball condition with those to the uniform conditions. For neuronal populations, we used the *t*-test to evaluate the significance of (i) the mean pop-out modulation in each stage and (ii) the change in distribution between stages. In population analysis, the modulation ratio of each cell was treated as a data point, and the significance of the population distribution of modulation ratios was evaluated against either a zero-mean distribution for positive pop out, or against another distribution for shift in the distribution. $P > 0.05$ indicated that the difference between the two means was not significant. The onset time of the pop-out response was estimated by first computing a temporal profile of the pop-out modulation ratio of the neuronal population evaluated within a 15 ms running window for each stimulus set, and then determining the time at which the population modulation ratio became significantly and consistently positive (when P dropped below 0.05 in a population *t*-test).

Note: Supplementary information is available on the Nature Neuroscience website.

Competing interests statement

The authors declare that they have no competing financial interests.

1. Julesz, B. Experiments in the visual perception of texture. *Sci. Am.* **232**, 34–43 (1975).
2. Treisman, A. & Gelade, G. A feature-integration theory of attention. *Cognit. Psychol.* **12**, 97–136 (1980).
3. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
4. Knierim, J. J. & Van Essen, D. C. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* **67**, 961–980 (1992).
5. Levitt, J. B. & Lund, J. S. Contrast dependence of contextual effects in primate visual cortex. *Nature* **387**, 73–76 (1997).
6. Nothdurft, H. C., Gallant, J. L. & Van Essen, D. C. Response modulation by texture surround in primate area V1: correlates of 'pop-out' under anesthesia. *Vis. Neurosci.* **16**, 15–34 (1999).
7. Lamme, V. A. F. The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* **10**, 649–669 (1995).
8. Zipser, K., Lamme, V. A. F. & Schiller, P. H. Contextual modulation in the primary visual cortex. *J. Neurosci.* **16**, 7376–7389 (1996).
9. Lee, T. S., Mumford, D., Romero, R. & Lamme, V. A. F. The role of the primary visual cortex in higher level vision. *Vis. Res.* **38**, 2429–2454 (1998).
10. Super, H., Spekreijse, H. & Lamme, V. A. F. Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci.* **4**, 304–310 (2001).
11. Ramachandran, V. S. Perception of shape from shading. *Nature* **331**, 163–166 (1988).
12. Sun, J. & Perona, P. Early computation of shape and reflectance in the visual system. *Nature* **379**, 165–168 (1996).
13. Braun, J. Shape from shading is independent of visual attention and may be a 'texton'. *Spat. Vis.* **7**, 311–322 (1993).
14. Wurtz, R. Response of striate cortex neurons to stimuli during rapid eye movements in the monkey. *J. Neurophysiol.* **32**, 975–986 (1969).
15. Leopold, D. A. & Logothetis, N. K. Microsaccades differentially modulate neural activity in the striate and extrastriate visual cortex. *Exp. Brain Res.* **123**, 341–345 (1998).
16. Bakin, J. S., Nakayama, K. & Gilbert, C. Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J. Neurosci.* **20**, 8188–8198 (2000).
17. Logothetis, N. K. Object vision and visual awareness. *Curr. Opin. Neurobiol.* **8**, 536–44 (1998).
18. Crist, R. E., Li, W. & Gilbert, C. D. Learning to see: experience and attention in primary visual cortex. *Nat. Neurosci.* **4**, 519–525 (2001).
19. Karni, A. & Sagi, D. The time course of learning a visual skill. *Nature* **365**, 250–252 (1993).
20. Ahissar, M. & Hochstein, S. Task difficulty and the specificity of perceptual learning. *Nature* **387**, 401–406 (1997).
21. Motter, B. 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 (1993).
22. Ito, M. & Gilbert, C. D. Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron* **22**, 593–604 (1999).
23. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2 and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 22–42 (1997).
24. Joseph, J. S., Chun, M. M. & Nakayama, K. Attentional requirements in a 'preattentive' feature search task. *Nature* **387**, 805–807 (1997).
25. Ahissar, M., Laiwand, R. & Hochstein, S. Attentional demands following perceptual skill training. *Psychol. Sci.* **12**, 56–62 (2001).
26. Roelfsema, P. R., Lamme, V. A. & Spekreijse, H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* **395**, 376–381 (1998).
27. Lee, T. S. & Nguyen, M. Dynamics of subjective contour formation in the early visual cortex. *Proc. Natl. Acad. Sci. USA* **98**, 1907–1911 (2001).
28. Stemmler, M., Usher, M. & Niebur, E. Lateral interactions in the primary visual cortex: a model bridging physiology and psychophysics. *Science* **269**, 1877–1880 (1995).
29. Li, Z. Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. *Network* **10**, 187–212 (1999).
30. Hupe, J. M. *et al.* Feedback connections act on the early part of the responses in monkey visual cortex. *J. Neurophysiol.* **85**, 134–145 (2001).
31. Rossi, A. F., Desimone, R. & Ungerleider, L. G. Contextual modulation in primary visual cortex of macaques. *J. Neurosci.* **21**, 1698–1709 (2001).
32. Bichot, N. P. & Schall, J. D. Effects of similarity and history on neural mechanisms of visual selection. *Nat. Neurosci.* **2**, 549–554 (1999).
33. Deco, G. & Lee, T. S. A unified model of spatial and object attention based on inter-cortical biased competition. In *Neurocomputing* (Elsevier, Amsterdam, in press).
34. Grossberg, S. Competitive learning: from interactive activation to adaptive resonance. *Cogn. Sci.* **11**, 23–63 (1987).
35. McClelland, J. L. & Rumelhart, D. E. An interactive activation model of context effects in letter perception (Part I). *Psychol. Rev.* **88**, 375–407 (1981).
36. Pollen, D. A. On the neural correlates of visual perception. *Cereb. Cortex* **9**, 4–19 (1999).
37. Mumford, D. On the computational architecture of the neocortex II. *Biol. Cybern.* **66**, 241–251 (1992).