Bottom-up saliency and top-down learning in the primary visual cortex of monkeys

Yin Yan (阎崟)[a,b], Li Zhaoping (李兆平)[a,b,c,1,2], and Wu Li[a,b,2]

[a]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, 100875 Beijing, China; [b]IDG/McGovern Institute for Brain Research, Beijing Normal University, 100875 Beijing, China; and [c]Department of Computer Science, University College London, WC1E 6BT London, United Kingdom

Early sensory cortex is better known for representing sensory inputs but less for the effect of its responses on behavior. Here we explore the behavioral correlates of neuronal responses in primary visual cortex (V1) in a task to detect a uniquely oriented bar—the orientation singleton—in a background of uniformly oriented bars. This singleton is salient or inconspicuous when the orientation contrast between the singleton and background bars is sufficiently large or small, respectively. Using implanted microelectrodes, we measured V1 activities while monkeys were trained to quickly saccade to the singleton. A neuron's responses to the singleton within its receptive field had an early and a late component, both increased with the orientation contrast. The early component started from the outset of neuronal responses; it remained unchanged before and after training on the singleton detection. The late component started ~40 ms after the early one; it emerged and evolved with practicing the detection task. Training increased the behavioral accuracy and speed of singleton detection and increased the amount of information in the late response component about a singleton's presence or absence. Furthermore, for a given singleton, faster detection performance was associated with higher V1 responses; training increased this behavioral–neural correlate in the early V1 responses but decreased it in the late V1 responses. Therefore, V1's early responses are directly linked with behavior and represent the bottom-up saliency signals. Learning strengthens this link, likely serving as the basis for making the detection task more reflexive and less top-down driven.

perceptual learning | bottom-up saliency | top-down influence | awake monkey | primary visual cortex

Within a visual scene, an item that is distinct from the others in a simple feature pops out perceptually. One example is a singleton bar orthogonal to uniformly oriented background bars (Fig. 1A, Middle). Such a striking visual feature contrast at a visual field location automatically attracts attention or gaze to this location in a stimulus-driven, goal-independent, or bottom-up manner. This effect is referred to as the visual saliency. The degree of saliency increases with the magnitude of the feature contrast, which has been thought to be encoded by early cortical areas.

In V1 of awake and anesthetized monkeys and cats (1–4), a neuron's responses to a bar within its receptive field (RF) are suppressed by the presence of surrounding isooriented bars. Such isoorientation suppression decreases with increasing orientation contrast between the center and the surround bars. This effect emerges from early V1 responses and has been viewed as a correlate of the perceptual pop-out effect. Meanwhile, it has been argued that V1 simply signals feature discontinuities regardless of saliency. For instance, a red vertical bar among many red horizontal and green vertical bars is not salient, but V1 neurons still signal its uniqueness (5). However, previous studies did not involve overt orienting to the singletons, making no direct link between V1 responses and behavioral saliency.

Although it has been proposed that the center-surround mechanisms in V1 create a saliency map of visual scenes to guide attention involuntarily to salient locations (6), supporting evidence has been mainly indirect. In humans, an orientation singleton bar is more salient when the singleton is presented to

one eye and the background bars to the other rather than the same eye (7), suggesting that the underlying mechanisms are at an early cortical stage that represents eye-of-origin information. When orientation singletons are rendered invisible by backward masking, they still capture attention, and fMRI signals in human V1 are significantly correlated with the local orientation contrasts (8). In monkeys, orientation singleton detectability is largely spared after a lesion in V2 (9) or V4 (10). These lines of evidence point to V1's role in mediating the saliency map.

However, a recent study argues that the superior colliculus (SC) is responsible for the initial saliency signals of the orientation contrast and that the saliency signals in V1 start only in the late response component (11). Meanwhile, late V1 responses have been reported to correlate with higher-order perceptual pop-out, such as a convex oddball defined by shading in a background of concave balls (12), a contour formed by collinear bars in a background of randomly oriented bars (13), and a foreground texture within a background texture (14). These delayed and higher-order V1 responses are strongly experience and task dependent: learning to detect the camouflaged contour only enhances the late V1 responses (15); this enhancement becomes much weaker when the contour is task irrelevant (13), and the contour-related signals are even abolished under anesthesia (16).

To resolve whether saliency of simple feature singletons is explicitly related to early V1 signals, we carried out simultaneous recordings of V1 and behavior responses while monkeys were

Significance

A visual item in sharp contrast with its neighboring items in a simple feature, such as color or orientation, automatically captures attention. Although cortical area V1 signals such simple feature contrasts, it is unclear whether these signals are utilized for saliency effects in behavior and whether they are enhanced by training. By training monkeys to detect a short bar in various degrees of orientation contrast with uniformly oriented background bars, we show that the feature contrast signals in V1, which start from the beginning of neuronal responses, are correlated with the detection performance and are therefore perceptual saliency signals. Training makes the early saliency signals more behaviorally correlated, consistent with more reflexive and less effortful task performance after training.
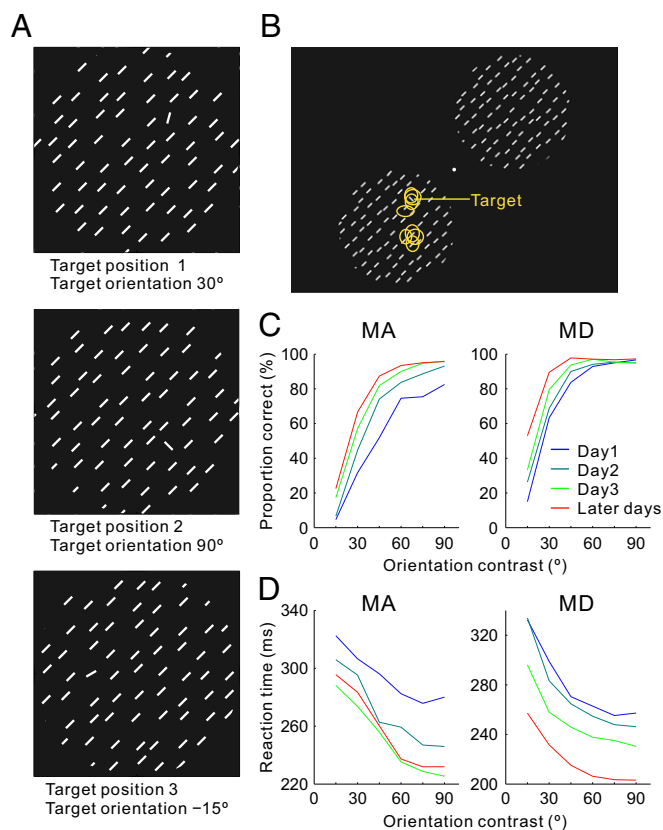
NEUROSCIENCE

Fig. 1. Experimental design and behavioral performance. (*A*) Sample singleton stimulus patterns within a visual field quadrant. The singleton bar was randomly set at one of the three possible locations and 1 of 12 orientations (only 3 shown here) relative to the background orientation. (*B*) Illustration of the two stimulus patterns in opposite quadrants with the singleton covered by one of the two RF clusters (illustrated with ellipses). Mean (*C*) detection accuracies and (*D*) reaction times as a function of the singleton's orientation contrast for monkeys MA and MD for the specified training days.

trained to detect an orientation singleton of various orientation contrasts. We also monitored whether learning to detect the singleton affected the link between behavior and V1 responses in both the early and late components.

Results

Two monkeys, MA and MD, each implanted with two microelectrode arrays in V1, learned to saccade to an orientation singleton (the target, sample stimuli in Fig. 1*A*) in a background of uniformly oriented bars. The background orientation was fixed throughout (135° for MA and 45° for MD; relative to horizontal); the singleton bar tilted randomly from the background bars by −75° to +90° in multiples of 15°, defining 12 target orientations that gave rise to seven orientation contrasts from 0° to 90°.

In each trial, after the monkey kept fixation for 1.1–1.6 s at the central fixation point (Fig. 1*B*), two patterns appeared in two opposite visual field quadrants, one containing the singleton and the other only the background bars. Each pattern was 8° (for MA) or 6° (for MD) in diameter, divided into 0.5° × 0.5° compartments, one for each bar (0.25° × 0.05°). The singleton was randomly placed in one of the six predefined compartments, three in each quadrant. The monkey was required to saccade to the singleton within 800 ms or, if the orientation contrast was 0° (catch trial without a target), to keep fixating within this 800 ms. A total of 72 target conditions (12 orientations × 6 locations) were randomly mixed within a block of trials. Each monkey practiced 30–40 blocks of trials per day.

In 50% of the trials, the target was in the pattern overlying the two clusters of the RFs recorded by the two electrode arrays

(Fig. 1*B*; eccentricity of the pattern center was 5° for MA or 3.6° for MD). The three possible singleton locations were roughly the vertices of an equilateral triangle, and two of them were within the two RF clusters. The two stimulus patterns, one with and one without the singleton, were placed symmetrically about the fixation point. In the other 50% of the trials, the visual stimuli were made simply by rotating the two patterns 180° around the fixation point. A pair of conditions linked by such a rotation are defined as the mirror conditions of each other.

Before collecting data in the detection task, the monkeys went through a procedural learning process, first by a couple of thousand trials with striking luminance differences between the target and background and then by several hundred trials with salient orientation singletons (>60° contrast). After this procedure training, which was completed within a single day, we started collecting data during the formal detection training (MA/MD 11/10 d).

The accuracy and speed of singleton detection increased with the orientation contrast (Fig. 1 *C* and *D*). Pooling the nonzero orientation contrasts for each day, we observed significant practice effects: detection accuracy improved with training days (Pearson correlation $r = 0.84$, $P = 0.0013$ for MA; $r = 0.94$, $P = 4.2 \times 10^{-5}$ for MD); the reaction time decreased concurrently (MA, $r = -0.74$, $P = 0.0099$; MD, $r = -0.88$, $P = 6.9 \times 10^{-4}$). The most pronounced improvements occurred in the first 3 d.

V1 Responses to the Orientation Singleton With and Without Task Experience. A neuron's responses to the singleton within its RF are influenced by two stimulus factors: the orientation matching factor—the similarity between the target's orientation and the neuron's preferred orientation—and the orientation contrast factor—the angular deviation of the target from the background. To examine interactions between these two factors and to isolate the effects of orientation contrast, we separated the V1 sites whose RFs covered the target location (*Materials and Methods*) into three groups according to their preferred orientations.

As a pretraining control, we first examined V1 responses when the monkeys performed a fixation task (Fig. 2 *A–C*). Fig. 2*A* shows the group of V1 sites that preferred orientations similar to that of the background bars. As the singleton's orientation deviated away from the background and thus away from these neurons' preferred orientations (blue to red in Fig. 2*A*), neuronal responses decreased even though the orientation contrast increased. This indicates a predominance of the orientation matching factor in influencing the neural responses, masking the effects of the orientation contrast.

Fig. 2*C* shows another group of V1 sites that preferred orientations roughly orthogonal to the background bars. Here the two influencing factors were also intermingled but in a different manner: as the singleton's orientation deviated progressively from that of the background bars, the orientation contrast and the orientation match (between the singleton's orientation and the neurons' preferred orientations) increased concurrently, leading to a progressive increase of neuronal responses.

In the above two groups of neurons, when the orientation contrast was varied by varying the target's orientation, their response changes could be well explained by neurons' orientation tuning alone. It is thus unclear whether the orientation contrast had any influences on V1 responses.

The remaining group of V1 sites came from pooling two subgroups that preferred orientations near +45° and −45° from that of the background bars (Fig. 2*B*). Hence, the two subgroups preferred orthogonal orientations: near horizontal and near vertical. When averaged together, their population responses were largely invariant to the orientations of gratings and thus unaffected by the orientation matching factor (*SI Appendix*, Fig. S1*B*). Nevertheless, to the singleton stimuli, the averaged responses increased significantly with the orientation contrast, starting from the early responses (Fig. 2*B*).

The analyses above indicate that saliency by orientation contrast is encoded by V1 neurons from their initial responses, even
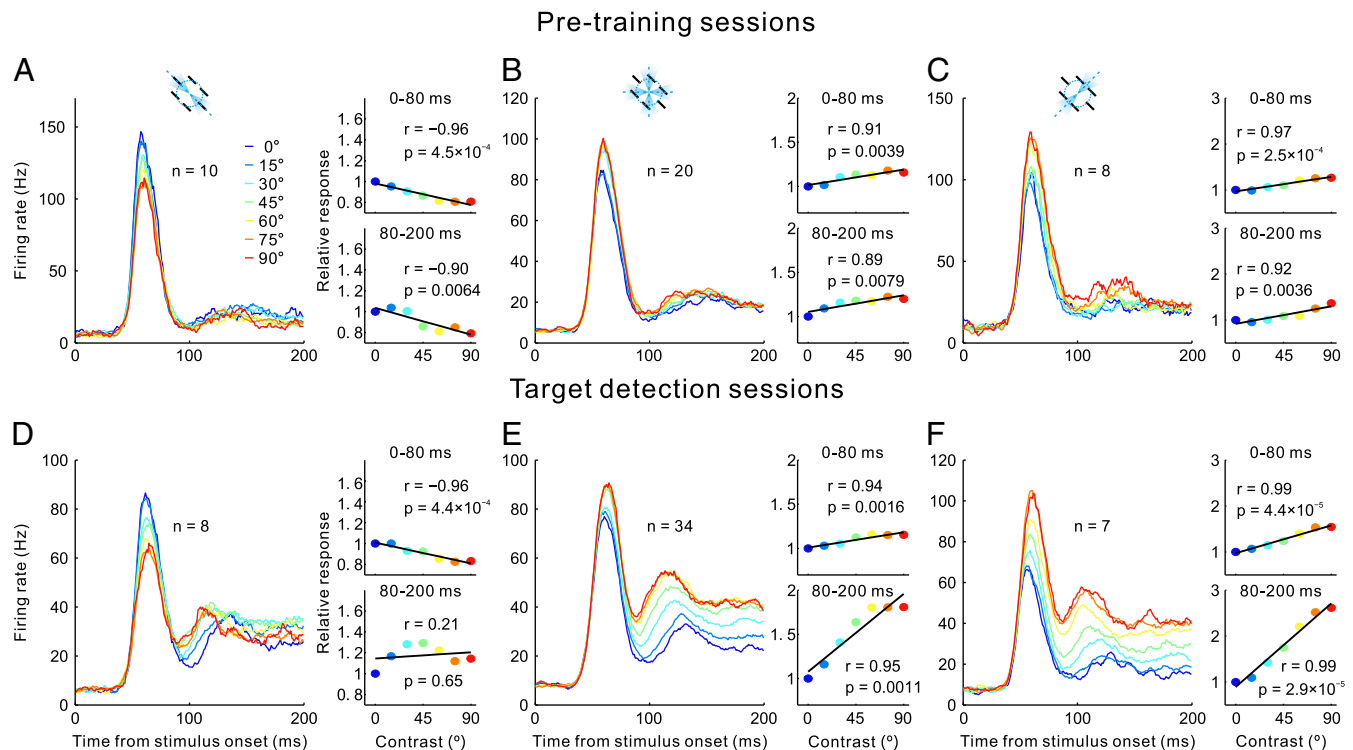
Fig. 2. Stimulus-driven and task-dependent components of V1 responses. (*A–C*) Averaged PSTHs of neuronal responses from both monkeys in the pretraining fixation task. Only V1 sites with RFs covering the target location (*Materials and Methods*) contributed to the average. They were separated into three groups according to the deviations of their preferred orientations from the background orientation: within 22.5° (*A*; $n = 10$ sites, 5/5 from MA/MD), between 22.5° and 67.5° (*B*; $n = 20$, 15/5 from MA/MD), or larger than 67.5° (*C*; $n = 8$, 6/2 from MA/MD). The top *Inset* schematizes the relationship between a contributing RF (dashed cyan circle), its preferred orientation (within the cyan sectors), and the six nearest background bars. Each PSTH corresponding to an orientation contrast (0° to 90° in multiples of 15°, blue to red) is the average of n PSTHs; each in turn is the average of the corresponding PSTHs from a single electrode across the pretraining days (5 and 3 d for MA and MD, respectively). The preferred orientation associated with an electrode was the average of the preferred orientations measured across the days. Each orientation contrast involves pooling two target orientations, clockwise and anticlockwise tilted from the background. Two small panels on the right show the relative mean neuronal responses within 0–80 and 80–200 ms, as a function of the orientation contrast. Linear regression results are shown in each panel. (*D–F*) Same as *A–C* but during the singleton detection task across the training days [11 and 10 d for MA and MD, respectively; (*D*) $n = 8$, 4/4 from MA/MD; (*E*) $n = 34$, 19/15 from MA/MD; and (*F*) $n = 7$, 5/2 from MA/MD].

though the monkeys are naive to, and not performing, the detection task, in accordance with previous findings (1, 2).

Next, we examined V1 responses during the singleton detection training, which started 5 d (MA) or 1 d (MD) after collecting the pretraining fixation data.

In all of the three groups of neurons, the initial responses (<80 ms) were little affected (compare Fig. 2 *A–C* with Fig. 2 *D–F*). However, the late responses (80–200 ms) were dramatically elevated and became strongly correlated with the orientation contrasts. Particularly, in the neurons preferring orientations near that of the background bars, the negative correlation seen before the detection training was inverted (Fig. 2*A* vs. Fig. 2*D*). Hence, practicing the detection task greatly amplifies the orientation contrast signals in the late but not the early V1 responses.

The above analyses treated data collected from the same electrode in different days as coming from the same sample (a single V1 site). Noticing a substantial change in the recorded neuronal responses between days (*SI Appendix*, Fig. S2 *A–C*), we repeated the analyses but treated recordings from the same electrode in different days as independent samples (different V1 sites). Qualitatively similar results were obtained when the data from the two animals were pooled (*SI Appendix*, Fig. S2 *D–F*) or separated (*SI Appendix*, Fig. S2 *G–L*).

Fig. 2 shows the results from V1 sites covering the target location. For V1 sites whose RFs were located on the background bars, their pretraining responses—both early and late components—were independent of the orientation contrasts of the singleton outside the RFs (*SI Appendix*, Fig. S3 *A–C*); however, practicing singleton detection induced a negative correlation between the

orientation contrast and the late, but not the early, responses (*SI Appendix*, Fig. S3 *D–F*). This further enhanced the representation of the target–background contrast in V1 (*SI Appendix*, Fig. S4).

Focusing on the orientation contrast, we dissociated two distinct components in V1 responses. The first one was stimulus driven; it was present before and during detection training and started from the initial responses. We refer to this component as the bottom-up saliency signal. The second component was superposed on the late part of the first one. It appeared only after practicing the detection task. We refer to this component as the task-dependent signal, which could be a signature of top-down processing and learning-induced changes.

Decoding Target Presence Using V1 Population Responses. We next quantified—using a machine learning algorithm known as the support vector machine (SVM)—the ability of V1 population responses to detect a target's presence.

For each target condition (a given location and orientation) and its mirror condition in the opposite visual field, an SVM classifier was trained to classify these two paired conditions using the population spike counts from all of the V1 sites within the early (0–80 ms) or late (80–200 ms) time window. For each time window, the average classification accuracy increased with the target's orientation contrast (Fig. 3*A*). However, only the accuracy by the late responses improved over training days (Fig. 3*B*): the average classification accuracy on the last day was significantly higher than that on the first day (MA, $P = 2.9 \times 10^{-4}$; MD, $P = 4.3 \times 10^{-5}$, one-tailed Wilcoxon signed-rank test). Such a learning effect was absent in the early responses (MA, $P = 0.77$;
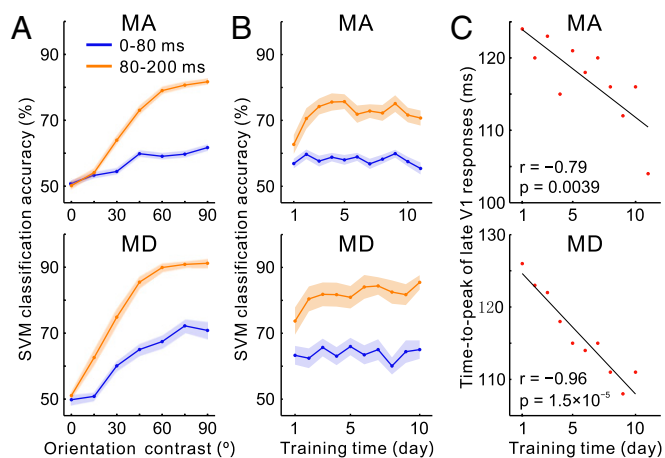
Yan et al.

www.manaraa.com

Fig. 3. Information about the target conveyed by V1 population responses. (*A*) Average SVM classification accuracies versus orientation contrast. A separate SVM classifier was trained for each target condition on each day, using the spike counts within 0–80 or 80–200 ms from all of the recorded V1 sites. The number of contributing V1 sites varied between 46 and 62 for MA and 38 and 57 for MD across days. The classification accuracies of these classifiers were averaged across days for each orientation contrast, time window, and monkey. Shading indicates ±SEM. (*B*) Average classification accuracies versus training days. The analysis was the same as in *A*, except that averaging across days was replaced by averaging across the nonzero orientation contrasts. (*C*) Time to peak of the late V1 responses versus training days. For each of the V1 sites with RF covering a target location, we obtained an average PSTH across the non-zero contrast conditions. Averaging across these V1 sites on each day, we measured the time to peak as the time when the population-averaged PSTH reached maximum during 80–200 ms (*SI Appendix*, Fig. S5). Black lines are linear fits to the red dots, with the statistics indicated in each plot.

MD, *P* = 0.39). The improvement in SVM decoding performance was most apparent in the first couple of training days (Fig. 3*B*), mirroring the behavioral progress (Fig. 1*C*).

Also correlated with the behavior was the time required for the late responses to reach maximum. This measure, defined as the time to peak, became shorter with orientation contrast (Fig. 2 *D–F*) and with training (Fig. 3*C* and *SI Appendix*, Fig. S5), mirroring the decrease of behavioral reaction times (Fig. 1*D*).

Links Between Neuronal Responses and Behavioral Performance. If the orientation contrast information in V1 is read out for behavior, trial-by-trial fluctuations of V1 responses and of behavioral performance should be correlated. Such a correlation was observed (*SI Appendix*, Fig. S6): higher V1 responses, in both the early and late components, tended to produce shorter reaction times. To better examine this correlation, we separated the trials in a given target condition on each day into two pools, the better-trial pool and the worse-trial pool.

For each target condition, the worse-trial pool started with all of the miss trials. If this pool had fewer than 50% of all of the trials, then, among the remaining (hit) trials, the one with the longest reaction time was moved to this worse-trial pool till it had 50% of all of the trials for this condition. The better-trial pool contained the rest of the hit trials (also 50%). However, if the initial worse-trial pool already had more than 50% of all of the trials, then the worse- and better-trial pools simply contained the miss and hit trials, respectively. To ensure a reasonable number of trials in the better-trial pool, we excluded conditions having hit rate <15%.

After pooling all of the conditions with nonzero orientation contrasts, the mean V1 responses were higher for the better trials than the worse trials at both the beginning (Fig. 4*A*) and the end (Fig. 4*B*) of training phases. The response difference is referred to as the better–worse difference. Training increased this difference for the early stimulus-driven V1 responses but decreased it for the late task-dependent responses (Fig. 4 *C* and *D*). These changes

could be explained if training not only facilitated the relevance of the early bottom-up saliency signals to the behavioral task but also rendered the task less dependent on top-down influences.

To further examine the better–worse difference, we separated the orientation contrasts into the small (≤45°, hard) and large (≥60°, easy) conditions (Fig. 4*D*). In the initial training days, the better–worse difference in V1's early responses was statistically significant only for large orientation contrasts (Fig. 4*D*, 40–60 ms group, cyan bars). This is consistent with the idea that V1 activities represent the saliency signals for perceptual pop out and is consistent with the automaticity of such pop-out when the saliency is sufficient. Because the early V1 signals were independent of the detection task (Fig. 2) and because the visual response latencies in V1 are shorter than those in the SC (11) and other cortical areas, a parsimonious account of such a better–worse difference is a utilization of the bottom-up saliency
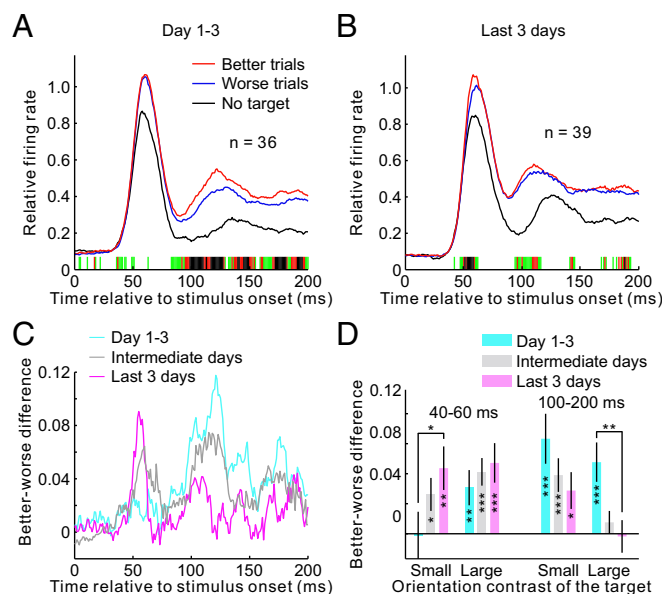

Fig. 4. Analysis of trial-by-trial correlation between V1 responses and behavior. V1 sites with RFs covering the target were pooled from both monkeys. (*A*) Population-averaged PSTHs separated for trials with better (red curve) and worse (blue) behavioral performance (see text for definitions) during the first three training days. Each of these two PSTHs was from averaging over *n* = 36 sites (23/13 from MA/MD). For each site (electrode), we first selected for each day the nonzero orientation contrast conditions that gave rise to PSTH peaks higher than that in the no-target condition (0 orientation contrast); we then pooled and averaged the PSTHs from these target-present conditions across the first 3 d. The population averaged PSTH from the no-target condition (black) is also shown for comparison. The resulting three PSTHs were rescaled so that the peak of the no-target PSTH was unity. Color bars above the *x* axis mark the time points when the averaged responses in the better trials are significantly larger than those in the worse trials (green, *P* < 0.05; red, *P* < 0.01; black, *P* < 0.001, one-tailed paired *t* test). (*B*) As in *A* but for the last three training days (*n* = 39, 24/15 from MA/MD). (*C*) The cyan and magenta curves show the better–worse differences by subtracting the blue curve from the red one in *A* and *B*, respectively; the gray curve shows the analogous better–worse difference for the intermediate training days between *A* and *B* (*n* = 53 sites, 29/24 from MA/MD). (*D*) Comparison of mean better–worse differences obtained as follows. Each curve in *C* was first remade by using only conditions with small (15°, 30°, and 45° pooled) or large (60°, 75°, and 90°) orientation contrasts. Each resulting curve was then averaged over the early (40–60 ms, corresponding to the ascending phase of the PSTH that is putatively stimulus driven) or late (100–200 ms) window. Error bars represent ±SEM. A better–worse difference significantly larger than zero is indicated on the data bar (*P* < 0.05; **P* < 0.01; ***P* < 0.001; one-tailed *t* test). The better–worse differences between the first three and last three training days were also compared (*P* < 0.05; **P* < 0.01; two-tailed *t* test).

Yan et al.

signals for the detection behavior. In the last training days, the better–worse difference in V1's late responses was statistically significant only for small orientation contrasts (Fig. 4D, 100–200 ms group, magenta bars), suggesting the importance of top-down influences on detecting less salient targets.

The analyses in Fig. 4 included miss trials, which could be taken as trials with reaction times longer than the permitted response window (800 ms). Qualitatively similar results were obtained if we included only the hit trials and evenly split them into the fast- and slow-trial pools (*SI Appendix*, Fig. S7).

Above analyses treated the data recorded from the same electrode in different days as coming from the same sample. Treating these data as independent samples produced qualitatively similar results when the data from the two animals were pooled or separated (*SI Appendix*, Fig. S8).

Discussion

Using a reaction time task, we simultaneously observed behavioral orienting to a feature singleton and neural representation of the corresponding saliency signals in V1. These signals started from V1 initial responses regardless of the detection task, consistent with previous findings without the orienting behavior (1). Our findings provide direct evidence supporting the idea that V1 creates a bottom-up saliency map through orientation-dependent contextual interactions (6, 17).

On top of the bottom-up saliency component, a distinct V1 response component emerged when the animal practiced the singleton detection task. This task-dependent component was delayed by ∼40 ms relative to the bottom-up signals. Moreover, task practice had no effect on the bottom-up component but profoundly modified the task-dependent, late, response component by shortening its time to peak and increasing the amount of information about the presence of the singleton. These learning-induced changes could result from an interplay between top-down and bottom-up processes; they might reflect a better utilization of the early neural saliency signals for target detection.

In addition, a significant correlation was observed between the fluctuations of the neural signals and the fluctuations in behavioral performance: Given a target condition, higher V1 responses were associated with faster and more accurate detection. In early V1 responses, this behavior-linked differentiation was present already in the initial training days for large orientation contrasts (Fig. 4D), consistent with the idea that attention capture by a sufficiently salient input is automatically guided by the bottom-up saliency map. For singletons with both small and large orientation contrasts, training boosted this behavior-linked differentiation in the early V1 responses but concurrently decreased it in the late task-dependent responses. This is consistent with the observations that singleton detection became more preattentive with training (18–21).

Different temporal components of V1 responses are suggested to play different roles: the late components are related to feedback modulations (12, 14, 22–24), whereas the initial bursts are largely driven by bottom-up inputs. Training to detect camouflaged global contours has been observed to refine neuronal population code in V1 by affecting only the late responses (15). The current study discovered a distinct practice effect: besides modifying the late V1 responses, training also increased the correlation between the early bottom-up saliency signals in V1 and the animals' target detection performance. Speculatively, this suggests a better utilization of the early V1 signals through training, which could be related to the faster buildup of the late, task-dependent, V1 response component (Fig. 3C and *SI Appendix*, Fig. S5).

Contextual interactions enable V1 to compute saliency signals from simple feature contrasts. For example, isoorientation suppression—suppression between nearby V1 neurons preferring similar orientations (1, 4, 25)—makes the V1 responses to a bar higher when this bar is an orientation singleton rather than one of the background bars because V1 neurons tuned to the singleton's orientation escape the suppression. The degree of this escape decreases with decreasing orientation contrast, mirroring the decreasing saliency effect behaviorally. Analogously, iso-color

suppression (26), iso-motion direction suppression (3, 27), iso-spatial frequency suppression (25), and even iso–eye-of-origin suppression (28) are all examples of the V1 phenomena referred to as the iso-feature suppression (6). This mechanism, which generates relatively higher responses to a feature singleton, is most likely responsible for attracting spatial attention exogenously to the singleton's location in visual scenes (7, 29).

Modifiable by training, the task-dependent late V1 responses could be associated with top-down influence from the fronto-parietal network (30–33). Meanwhile, the early bottom-up saliency map in V1 may provide inputs to the attentional priority map in the parietal (34) and frontal (35, 36) cortex, which combine the bottom-up and top-down factors to guide attention in a task-dependent manner. V1 also projects monosynaptically to the SC (37, 38), which directs gaze shifts through the brainstem. It is likely that this V1-to-SC circuit becomes more involved as training makes the task more reflexive and less top-down dependent. Indeed, lesioning the SC eliminates express saccades and prolongs saccadic reaction times, whereas lesioning the frontal eye field promotes express saccades (39).

A recent study argues that the SC generates the initial, feature-agnostic, saliency signals by pooling V1 inputs and then relays these signals back to V1 (11). Surprisingly, that study did not see any orientation contrast signals in early V1 responses, in contrast to our current study and previous studies (1, 2). Lesioning monkey V1 abolishes all visually guided saccades until after 2 mo of training and recovery (40), suggesting that in normal circumstances, retina-to-SC projections are insufficient to generate orienting behavior (except perhaps in lower vertebrates; ref. 41). Future studies by selective manipulation of neural activities in monkey V1 and SC will be helpful for dissecting their causal relationship.

Materials and Methods

Animal Preparations. Two adult male monkeys (*Macaca mulatta*, MA and MD) were first trained, with the head restrained, to perform a simple fixation task. Afterward, two microelectrode arrays (two 4 × 8 in MA and two 6 × 8 in MD; Blackrock Microsystems) were implanted in V1. The neighboring electrodes were 0.4 mm apart; all electrodes were 0.5 mm long except for those in one of the two arrays implanted in MA that were 1.5 mm long (array 1 in *SI Appendix*, Fig. S4A). Surgical procedures were performed in aseptic environment under anesthesia with vital signs maintained. The current study was conducted 12 mo after the implantation for MA and 3 mo for MD. The monkeys had been used in an earlier study on global contour detection within a background of randomly oriented bars, but they were naive to the singleton detection task. All experimental procedures complied with the US National Institutes of Health Guide for the Care and Use of Laboratory Animals (42) and were approved by the Institutional Animal Care and Use Committee of Beijing Normal University.

Visual Stimuli and Experimental Settings. Visual stimuli were generated by a stimulus generator (ViSaGe MKII; Cambridge Research Systems) on a 22-inch cathode ray tube (CRT) monitor (Iiyama HM204DTA, 1,200 × 900 pixels at 100 Hz, 100-cm viewing distance). Eye positions were sampled at 30 Hz (for MA) by a self-made infrared tracking system (43), or at 500 Hz (for MD) by a commercial eye tracking system (Eyelink 1000; SR Research Ltd.).

The fixation point was 0.12° in diameter. Fixation had to be held within an invisible window of 1.2° in diameter. The stimulus bars were 25.8 cd/m² on a gray background of 8.6 cd/m². The position of each bar except the singleton was jittered randomly from its compartment center by 0–0.18° in random directions. For target-present trials (nonzero orientation contrasts), rewards were given to valid saccades (within an 800-ms response window after stimulus onset), defined as follows: Once the gaze escaped the fixation window, it had to enter a target window (3° diameter) centered on the singleton within 100 ms and stay within this window for at least 100 ms. The monkey's reaction time was the duration between the stimulus onset and the gaze reaching the target window. Reward in a catch trial (zero orientation contrast) was twice the typical size for maintaining fixation during the 800-ms response window.

Training Procedure. We taught the animal the saccade task through the following transitional steps. (*i*) We first hid all of the nontarget bars in the stimuli by setting their luminance to that of the CRT background. The animal learned to saccade to the isolated target bar after a few hundred trials.

Yan et al.

(*ii*) We progressively increased the luminance of the nontarget bars but left them noticeably dimmer than the target bar. The animals learned to saccade to the target defined by both luminance and orientation contrasts after a couple of thousand trials within the same day. (*iii*) We set the luminance of the target and background bars identical but kept their orientation contrast large (>60°). As soon as the monkey understood the saccade task simply based on orientation contrasts after several hundred trials, we started collecting behavioral and V1 data over the course of perceptual training (11 d for MA and 10 d for MD).

Electrophysiological Recording. Multiunit activities were recorded with a data acquisition system (Cerebus, Blackrock Microsystems). The raw data were high-pass filtered (fourth-order Butterworth with 250-Hz corner frequency). Spikes were detected by applying a voltage threshold with a signal-to-noise ratio of 4, and their waveforms were sampled and saved at 30 kHz.

For each electrode and on each day, the RF location and size along the horizontal axis were mapped using a vertically elongated grating patch (0.3° × 7° in size, square wave, 2 cycles per degree, 99% Michelson contrast, 3-Hz drifting frequency) placed at different horizontal locations. Neuronal responses were fitted by a Gaussian function of the horizontal position, with center x_0 and SD σ_x. Analogously, y_0 and σ_y were determined along the vertical axis using a grating patch elongated horizontally (7° × 0.3°). The RF, defined as the oval centered at (x_0, y_0) with axes (σ_x, σ_y), is said to cover the target location when this oval enclosed the center of the target. This oval area could also cover parts of the background bars, but the results were insensitive to shrinking or enlarging this area to include less or more background bars. The preferred orientation was measured with a circular grating patch (2 cycles per degree) centered on all recorded V1 RFs. Neuronal responses were fitted by a Gaussian function of the grating orientation to determine the preferred orientation. The goodness of Gaussian fit was estimated using R^2. Only electrodes yielding $R^2 > 0.7$ for the RF profiles were included in data analyses (for Fig. 2 and the analogous plots in *SI Appendix*, Figs. S1–S3, $R^2 > 0.7$ for the orientation tuning curve is additionally required). The contributing electrodes varied between days due to changes in signal quality. The RF size ($\sigma_x + \sigma_y$) was 0.57 ± 0.22° (mean ± SD) for MA and 0.39 ± 0.08° for MD; the orientation tuning width (full width at half height) was 56 ± 25° (MA) and 58 ± 24° (MD).

Analysis of V1 Responses. For each stimulus condition, the spike trains from each electrode across trials were binned into 1-ms intervals and smoothed by a 9-ms square window to construct a poststimulus time histogram (PSTH). The PSTHs were averaged across V1 sites, conditions, and days if applicable, as specified in Figs. 2 and 4 and *SI Appendix*, Figs. S1–S3, S5, S7, and S8. Considering that the saccadic reaction times toward the recorded RFs were >160 ms and that the mean response latency of V1 neurons was ∼40 ms, we used neuronal responses within 0–200 ms after stimulus onset for analysis. We treated V1 responses recorded by the same electrode in different days as coming from the same sample (a single V1 site) for the results presented in Figs. 2 and 4. Noticing that the distribution of the firing rates recorded by the same electrode usually changed significantly from one day to another (Kolmogorov–Smirnov test; *SI Appendix*, Fig. S2 *A–C*), we also presented in *SI Appendix* the comparable results from treating these data as independent samples (different V1 sites).

Decoding Analysis. The SVM classifier had a radial basis function kernel. Training of each SVM used randomly 90% of the trials from the corresponding conditions; the remaining trials were used to test the accuracy of the classifier. This process was repeated 1,000 times to calculate the average classification accuracy. A total of 72 SVMs (for 3 target locations × 12 target orientations × 2 time windows of V1 responses) were trained for each day.

The data, code, and materials used in the current study are available from the corresponding authors upon request.

1. Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J Neurophysiol* 67:961–980.
2. Nothdurft HC, Gallant JL, Van Essen DC (1999) Response modulation by texture surround in primate area V1: Correlates of "popout" under anesthesia. *Vis Neurosci* 16:15–34.
3. Kastner S, Nothdurft HC, Pigarev IN (1997) Neuronal correlates of pop-out in cat striate cortex. *Vision Res* 37:371–376.
4. Blakemore C, Tobin EA (1972) Lateral inhibition between orientation detectors in the cat's visual cortex. *Exp Brain Res* 15:439–440.
5. Hegdé J, Felleman DJ (2003) How selective are V1 cells for pop-out stimuli? *J Neurosci* 23:9968–9980.
6. Li Z (1999) Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc Natl Acad Sci USA* 96:10530–10535.
7. Zhaoping L (2008) Attention capture by eye of origin singletons even without awareness—A hallmark of a bottom-up saliency map in the primary visual cortex. *J Vis* 8:1.
8. Zhang X, Zhaoping L, Zhou T, Fang F (2012) Neural activities in v1 create a bottom-up saliency map. *Neuron* 73:183–192.
9. Merigan WH, Nealey TA, Maunsell JHR (1993) Visual effects of lesions of cortical area V2 in macaques. *J Neurosci* 13:3180–3191.
10. Merigan WH (2000) Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. *Vis Neurosci* 17:949–958.
11. White BJ, Kan JY, Levy R, Itti L, Munoz DP (2017) Superior colliculus encodes visual saliency before the primary visual cortex. *Proc Natl Acad Sci USA* 114:9451–9456.
12. Lee TS, Yang CF, Romero RD, Mumford D (2002) Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nat Neurosci* 5:589–597.
13. Li W, Piëch V, Gilbert CD (2006) Contour saliency in primary visual cortex. *Neuron* 50:951–962.
14. Poort J, et al. (2012) The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* 75:143–156.
15. Yan Y, et al. (2014) Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat Neurosci* 17:1380–1387.
16. Li W, Piëch V, Gilbert CD (2008) Learning to link visual contours. *Neuron* 57:442–451.
17. Li Z (2002) A saliency map in primary visual cortex. *Trends Cogn Sci* 6:9–16.
18. Braun J (1998) Vision and attention: The role of training. *Nature* 393:424–425.
19. Sigman M, et al. (2005) Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron* 46:823–835.
20. Sireteanu R, Rettenbach R (2000) Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vision Res* 40:2925–2949.
21. Wang Q, Cavanagh P, Green M (1994) Familiarity and pop-out in visual search. *Percept Psychophys* 56:495–500.
22. Roelfsema PR, Tolboom M, Khayat PS (2007) Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron* 56:785–792.
23. Chen M, et al. (2014) Incremental integration of global contours through interplay between visual cortical areas. *Neuron* 82:682–694.
24. Chen R, Wang F, Liang H, Li W (2017) Synergistic processing of visual contours across cortical layers in V1 and V2. *Neuron* 96:1388–1402.e4.
25. Li CY, Li W (1994) Extensive integration beyond the classical receptive field of cat's striate cortical neurons—Classification and tuning properties. *Vision Res* 34:2337–2355.
26. Wachtler T, Sejnowski TJ, Albright TD (2003) Representation of color stimuli in awake macaque primary visual cortex. *Neuron* 37:681–691.
27. Jones HE, Grieve KL, Wang W, Sillito AM (2001) Surround suppression in primate V1. *J Neurophysiol* 86:2011–2028.
28. DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. *J Neurophysiol* 71:347–374.
29. Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12:97–136.
30. Buffalo EA, Fries P, Landman R, Liang H, Desimone R (2010) A backward progression of attentional effects in the ventral stream. *Proc Natl Acad Sci USA* 107:361–365.
31. Gilbert CD, Li W (2013) Top-down influences on visual processing. *Nat Rev Neurosci* 14:350–363.
32. Li W (2016) Perceptual learning: Use-dependent cortical plasticity. *Annu Rev Vis Sci* 2:109–130.
33. Roelfsema PR, Holtmaat A (2018) Control of synaptic plasticity in deep cortical networks. *Nat Rev Neurosci* 19:166–180.
34. Bisley JW, Goldberg ME (2010) Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci* 33:1–21.
35. Sato TR, Schall JD (2003) Effects of stimulus-response compatibility on neural selection in frontal eye field. *Neuron* 38:637–648.
36. Thompson KG, Bichot NP (2005) A visual salience map in the primate frontal eye field. *Prog Brain Res* 147:251–262.
37. Finlay BL, Schiller PH, Volman SF (1976) Quantitative studies of single-cell properties in monkey striate cortex. IV. Corticotectal cells. *J Neurophysiol* 39:1352–1361.
38. Schiller PH (1998) The neural control of visually guided eye movements. *Cognitive Neuroscience of Attention: A Developmental Perspective*, ed Richards JE (Lawrence Erlbaum Associates, Mahwah, NJ), pp 3–50.
39. Schiller PH, Sandell JH, Maunsell JH (1987) The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. *J Neurophysiol* 57:1033–1049.
40. Isa T, Yoshida M (2009) Saccade control after V1 lesion revisited. *Curr Opin Neurobiol* 19:608–614.
41. Zhaoping L (2016) From the optic tectum to the primary visual cortex: Migration through evolution of the saliency map for exogenous attentional guidance. *Curr Opin Neurobiol* 40:94–102.
42. National Research Council (2011) *Guide for the Care and Use of Laboratory Animals* (National Academies Press, Washington, DC), 8th Ed.
43. Matsuda K, Nagami T, Kawano K, Yamane S (2000) A new system for measuring eye position on a personal computer. *Soc Neurosci Abstr* 744.2.

Yan et al.

www.manaraa.com