Models of visual attention postulate the existence of a bottom-up saliency map that is formed early in the visual processing stream. Although studies have reported evidence of a saliency map in various cortical brain areas, determining the contribution of phylogenetically older pathways is crucial to understanding its origin. Here, we compared saliency coding from neurons in two early gateways into the visual system: the primary visual cortex (V1) and the evolutionarily older superior colliculus (SC). We found that, while the response latency to visual stimulus onset was earlier for V1 neurons than superior colliculus superficial visual-layer neurons (SCs), the saliency representation emerged earlier in SCs than in V1. Because the dominant input to the SCs arises from V1, these relative timings are consistent with the hypothesis that SCs neurons pool the inputs from multiple V1 neurons to form a feature-agnostic saliency map, which may then be relayed to other brain areas.

Models of visual attention postulate the existence of a bottom-up saliency map that is formed early in the visual processing stream. Although studies have reported evidence of a saliency map in various cortical brain areas, determining the contribution of phylogenetically older pathways is crucial to understanding its origin. Here, we compared saliency coding from neurons in two early gateways into the visual system: the primary visual cortex (V1) and the evolutionarily older superior colliculus (SC). We found that, while the response latency to visual stimulus onset was earlier for V1 neurons than superior colliculus superficial visual-layer neurons (SCs), the saliency representation emerged earlier in SCs than in V1. Because the dominant input to the SCs arises from V1, these relative timings are consistent with the hypothesis that SCs neurons pool the inputs from multiple V1 neurons to form a feature-agnostic saliency map, which may then be relayed to other brain areas.

Models of visual attention postulate the existence of a bottom-up saliency map that is formed early in the visual processing stream. Although studies have reported evidence of a saliency map in various cortical brain areas, determining the contribution of phylogenetically older pathways is crucial to understanding its origin. Here, we compared saliency coding from neurons in two early gateways into the visual system: the primary visual cortex (V1) and the evolutionarily older superior colliculus (SC). We found that, while the response latency to visual stimulus onset was earlier for V1 neurons than superior colliculus superficial visual-layer neurons (SCs), the saliency representation emerged earlier in SCs than in V1. Because the dominant input to the SCs arises from V1, these relative timings are consistent with the hypothesis that SCs neurons pool the inputs from multiple V1 neurons to form a feature-agnostic saliency map, which may then be relayed to other brain areas.

Significance

Theories of visual attention postulate the existence of a saliency map that guides attention/gaze toward the most visually conspicuous stimuli in complex scenes. This study compared saliency coding in the two dominant visual gateways: the primary visual cortex (V1) and the evolutionarily older visual system that exists in the midbrain superior colliculus. Our results show that neurons in the superficial visual layers of the superior colliculus (SCs) encoded saliency earlier and more robustly than V1 neurons. This was surprising, because the dominant input to the SCs arises from V1. This result is in line with models that place a feature processing stage (V1) before the feature-agnostic saliency map in SCs.

Author contributions: B.J.W., J.Y.K., L.I., and D.P.M. designed research; B.J.W. and J.Y.K. performed research; R.L. contributed new reagents/analytic tools; B.J.W. and J.Y.K. analyzed data; R.L. was the primary surgeon in the chronic implantation of a microelectrode array in the primary visual cortex; L.I. contributed significantly to the research idea and hypotheses and cowrote portions of the manuscript; and B.J.W., J.Y.K., and D.P.M. wrote the paper. The authors declare no conflict of interest.

This article is a PNAS Direct Submission. M.C. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. Email: brian.white@queensu.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1701003114/-/DCSupplemental.
computations (center-surround contrasts that make a stimulus stand out from its isoluminant neighbors in a complex scene). In addition, the goal-irrelevant oddball/item was centered in the RF in two ways: (i) the stimulus appeared abruptly with the oddball/single-item centered IN or OPP the RF; or (ii) the oddball/single item was brought IN or OPP the RF via a saccade (11). This allowed us to examine possible differences in saliency coding associated with passive viewing of an abrupt visual onset vs. active gaze-dependent shifting of stimuli over different parts of the visual field via saccades (11), which is characteristic of real-world viewing.

**Results**

**Saliency Coding Emerges Earlier in SCs than V1.** We compared visually evoked responses when a salient but goal-irrelevant oddball appeared IN vs. OPP the RF, representing high- vs. low-saliency...
regions of the display, the difference of which defined the saliency index (SI Materials and Methods). Fig. 2 A–D shows normalized population-averaged responses for V1 (Fig. 2, blue) and SCs (Fig. 2, red) neurons (note that these results were not affected by the normalization procedure as indicated by the identical statistical outcome using absolute firing rates) (Fig. S1). For a given neuron, we normalized to the averaged peak response of the single-item condition within the window from 0 to 500 ms (SI Materials and Methods). Because the single item always produced the greatest visual response, this ensured that the relative differences between conditions for a given neuron were retained. Although V1 and to a lesser degree, SCs (31) neurons showed some feature preference (Fig. S2), the response curves depicted in Fig. 2 represent the data collapsed across the feature combinations, effectively canceling out feature-specific differences. It should be noted that the V1 response curves represent combined single units plus multiunit sites (Materials and Methods), the results of which were qualitatively similar but less robust with the single units alone (Fig. S3). At the population level, V1 and SCs neurons showed a significant preference for the salient oddball as indicated by the difference between the oddball IN (Fig. 2, thick traces) and oddball OPP (Fig. 2, thin traces) conditions (P < 0.05, Wilcoxon signed rank test at 10-ms intervals, Bonferroni–Holm corrected). In addition, this oddball selectivity was qualitatively similar whether the oddball appeared abruptly in the RF (Fig. 2 A and C, array-aligned condition) or was brought into the RF via a saccade (Fig. 2 B and D, saccade end-aligned condition).

To explore these differences in detail, we derived two quantities from the response profiles of each neuron, the visual response onset latency (VROL; the earliest emergence of a visual response) and the saliency response onset latency (SROL; the earliest at which the neuron signaled a preference for the oddball) (Materials and Methods). Fig. 2 E and F shows the cumulative distributions of VROLS (Fig. 2, solid traces) and SROLS (Fig. 2, dotted traces) for V1 and SCs neurons. Only the subset of neurons which showed significant preference for the oddball was included in the cumulative distribution of SROLS. For the array-aligned condition (Fig. 2E), VROL was significantly earlier for V1 (median = 40 ms, n = 56) (Fig. 2, vertical blue line) than SCs (median = 49 ms, n = 24; P = 0.0008, rank sum test) (Fig. 2, vertical red line), whereas SROL was significantly earlier for SCs (median = 65 ms, n = 24) than V1 (median = 75 ms, n = 69; P = 0.0021, rank sum test). These results were qualitatively similar in the saccade end-aligned condition (Fig. 2F). There, VROL was significantly longer for V1 (median = 47 ms, n = 64) than SCs (median = 38 ms, n = 21; P = 0.013, rank sum test). More importantly, SROL was significantly earlier for SCs (median = 60 ms, n = 21) than V1 (median = 121 ms, n = 64; P = 0.0041, rank sum test). The results were qualitatively similar whether RFs were overlapping or non-overlapping between brain areas (Fig. S4).

These results indicate that, although V1 signaled the earliest visual arrival times in the stimulus-aligned condition (although not in the saccade end-aligned condition), on average, the saliency representation in V1 did not emerge until 60–75 ms after it had already appeared in the SCs. This delay was relatively shorter for V1 single units (14–37 ms) (Fig. S3 E and F) than V1 single units plus multiunits (Fig. 2 E and F), but in either case, the results were generally the same. Using the larger dataset that included single-unit plus multiunit V1 recordings led to more reliable and statistically powerful V1 results (compare, for example, the difference between the population-averaged spike density traces in Fig. 2 A and B and Fig. S3 A and B).

In addition, in V1, there was an average delay of 45–99 ms from visual onset to the emergence of the saliency representation (comparing the blue solid traces with the blue dotted traces in both Fig. 2 E and F and Fig. S3 E and F). This delay is explained by the fact that the saliency representation in V1 did not emerge until after the initial volley of visual activity, as indicated by the blue vertical dotted line in Fig. 24 (Fig. S34). In contrast, while SCs neurons showed overall later VROLS than V1 in the stimulus-aligned condition, the saliency representation emerged within the initial volley of visual activity around the earliest part of the visual response (~15–20 ms after visual response onset) as indicated by the red vertical dotted line in Fig. 2C. This relatively slow signaling of saliency in V1 suggests either a processing delay within V1 itself to generate a saliency representation or that the saliency representation in V1 emerged via feedback from other brain areas. Interestingly, VROL was the same or shorter for SCs than V1 in the saccade end-aligned case (Fig. 2F and Fig. S3F). This might indicate that, under real-world viewing conditions, where stimuli are more often brought into a neuron’s RF as a result of saccades, VROL may be earlier in SCs than V1. Most importantly, the saliency representation emerged reliably earlier in SCs than V1.

### Surround Modulation Emerges Earlier and Stronger in SCs than V1.

Most models of visual saliency depend critically on center-surround feature contrasts that extend widely across the visual field (1–3), and therefore, comparison of the surround suppression characteristics can provide a strong index of the degree to which each brain area represents saliency. To evaluate surround suppression, we compared visual responses evoked by the wide-field array (surround) (Fig. 1C with a single item (no surround) (Fig. 1E). Fig. 3 A–D shows population responses in the single-item IN condition (no surround) (Fig. 3, black trace) vs. the oddball IN condition (surround) (Fig. 3, color trace). From these averaged population traces, one can see noticeable differences in the timing of surround suppression between brain areas, with SCs (Fig. 3 C and D) showing markedly earlier suppression than V1 (Fig. 3 A and B) (P < 0.05, Wilcoxon signed rank test at 10-ms intervals, Bonferroni–Holm corrected; note that these results were not affected by the normalization procedure as indicated by the identical statistical outcome using absolute firing rates (Fig. S5)). In particular, the suppression in V1 did not emerge until well after the initial volley of visual activity as exemplified by the blue vertical dotted line in Fig. 34 (Fig. S6A). This is in contrast to SCs, in which the suppression emerged within or before the peak of the initial volley of visual activity as exemplified by the red vertical dotted line in Fig. 3C. To quantify the magnitude of the differences, we compared the percentage of suppression within two time periods: (i) during the peak of the visual response and (ii) during the sustained portion of the response (details are in SI Materials and Methods). Percentage suppression (Fig. 3 E–H) was significantly greater for SCs neurons (Fig. 3, red) than for V1 neurons (Fig. 3, blue) during both the peak (Fig. 3 E, array-aligned; P = 9.3e-07 and G, saccade end-aligned; P = 1.2e-05) and sustained (Fig. 3 F, array-aligned; P = 0.0089 and H, saccade end-aligned; P = 2.8e-04) portions of the response (rank sum test).

As with SROL above, we computed the surround suppression onset latency (SSOL) for each neuron that showed significant surround suppression, defined as the earliest time in which a neuron signaled a significantly greater response for the single-item IN condition (no surround) vs. the oddball IN condition (surround) (SI Materials and Methods). Fig. 3f shows that, for V1 neurons in the stimulus-aligned condition, SSOL (median = 75 ms, n = 69) was significantly delayed relative to VROL (median = 40 ms, n = 69; P = 6.8e-16). In contrast, for SCs neurons in the stimulus-aligned condition, SSOL (median = 51 ms, n = 25) was not significantly delayed relative to VROL (median VROL = 49 ms, n = 25; P = 0.18). Fig. 3j shows that, for V1 neurons in the saccade end-aligned condition, SSOL (median = 95 ms, n = 52) was again significantly delayed relative to VROL (median = 46 ms, n = 52; P = 2.2e-12). For SCs neurons in the saccade end-aligned condition, SSOL (median = 51 ms, n = 22) was moderately delayed relative to VROL (median VROL = 39 ms, n = 22; P = 0.0004). Importantly, SSOL emerged in V1, on average, 22–44 ms later.
than in SCs (note the difference between the red and blue dotted curves in Fig. 3 I and J) \( (P = 2.8e-05 \) for stimulus-aligned data; \( P = 6.1e-05 \) for saccade end-aligned data, rank sum test). These results were qualitatively similar using V1 single units alone (Fig. S6 A and I). The results were also qualitatively similar whether RFs were overlapping or nonoverlapping between brain areas (Fig. S7). This indicates that, similar to the SROL results described above, SSOL also showed relatively slow dynamics in V1. This suggests either a processing delay to generate surround suppression within V1 itself or that surround modulation in V1 arises, in part, via feedback from other brain areas (32–34). In either case, these results are in agreement with the hypothesis that the SCs, but not V1, have center-surround characteristics suited for computing saliency from wide-field visual inputs (35).

**Goal-Irrelevant Saliency Is Weakly Represented in SCi.** The input–output structure and response characteristics of SCi neurons differ in many important ways from SCs neurons (23–25). In particular, SCi neurons receive a diverse set of inputs predominantly from frontal and parietal cortices and the basal ganglia (Fig. 1B) and project directly to the brainstem saccade generator (23–25). Functionally, these neurons have multisensory and oculomotor responses and have been shown to play a central role in spatial attention (19, 20). This has led to the hypothesis that SCi is best described as a priority map for goal-relevant stimuli (36, 37). These results suggest a mechanism that restricts goal-irrelevant saliency signals in SCi from passing through to downstream gaze control circuitry.

**Discussion**

This study compared saliency coding in two early visual areas of the primate visual system, V1, and the hub of the evolutionarily older visual system, SCs. We found that, while both V1 and SCs encoded the presence of a salient but goal-irrelevant oddball, the saliency representation in V1 did not emerge until after it had already emerged in SCs (Fig. 2 E and F). This indicates that, although V1 is the dominant input to SCs (24, 25),
the saliency representation in SCs was not simply a readout of computations done upstream in V1. This is interesting, because V1 neurons registered the arrival of visual signals reliably earlier than SCs in the stimulus-aligned condition (although not the saccade end-aligned condition), but V1 was slower to signal both saliency (Fig. 2E and F) and surround modulation (Fig. 3), an essential component for computing saliency across the visual field (1–3). Moreover, the saliency representation was generally limited to the superficial visual SC layers. SCI visuomotor neurons, with dominant inputs that arise from frontal–parietal cortices and the basal ganglia (23–25), did not show a significant preference for the salient oddball (Fig. 4A and B). This finding is consistent with an earlier study that reported a weak saliency representation in SCI during free viewing of natural scenes (26). Taken together, the results provide evidence that the evolutionarily older SCs, not V1, embodies the role of a saliency map.

The relative timing of these visual representations offers insight into the flow of information processing between V1 and SC. Using the traditional index of visual latency timed from stimulus onset (Fig. 2A and C), V1 was the earliest to receive visual inputs (Fig. 2), which is consistent with the role of an early feature extraction stage (Fig. 1A, blue and B, blue). The evolutionarily older SCs then signaled the arrival of visual inputs, which may, in part, be inherited from pooled inputs of multiple V1 neurons (24, 25). This would effectively cancel out most feature-specific preferences, consistent with the broadly tuned characteristics of SCs neurons (31, 38, 39). Interestingly, VROL was the same or shorter for SCs than V1 in the saccade end-aligned case (Fig. 2F and Figs. S1, S3–S5, and S7). This might indicate that, under real-world viewing conditions, where stimuli are more often brought into a neuron’s RF as a result of saccades, retinotectal inputs might dominate the earliest part of the SCs visual response. In either case, SCs signaled the presence of the salient oddball before V1. Interestingly, the saliency representation in SCs was immediately preceded by robust surround suppression, which was markedly stronger and earlier than in V1. This is strong evidence that SCs has center-surround characteristics better suited for computing saliency from wide-field visual inputs (35). For SCI, visual inputs not associated with the saccade goal were strongly attenuated (Fig. 4D). This result suggests a mechanism that suppresses goal-irrelevant visual inputs in SCI to prioritize which salient signals gain access to the downstream gaze control system. Taken together, these results support a simple, intuitive framework for the formation of the saliency map.
midbrain seems to have solved this biological problem long before the elaboration of visual cortex (16–18). A potential functional advantage of a saliency mechanism in SC is that it would reside close to the gaze-orienting circuitry. The sensory and motor maps of the SCs and SCi are closely aligned (45) and ideally suited to integrate diverse inputs from cortex and engage the orienting system to act on such inputs. Thus, we postulate that, through evolution, V1 was integrated into an already existing saliency system, providing greater feature specificity and fine-tuned control over visually guided behavior.

Materials and Methods

Data were collected from four male rhesus monkeys (Macaca mulatta): two for the SC recordings and two for the V1 recordings (details are in SI Materials and Methods). All procedures were approved by the Queen’s University Animal Care Committee in accordance with the guidelines of the Canadian Council on Animal Care.

The stimuli consisted of a radial arrangement of equally spaced color bars (210 items), with the diameter of the entire display spanning 40° to 45° visual angle (Fig. 1C and D). The oddball, which could appear IN or OPP the RF, was always the opposite color and orientation as the remaining items, but all stimulus combinations were run interleaved. A single-item control condition served as a benchmark with no surround stimuli (Fig. 1E and F).

On a given trial (Fig. 1), the animal fixated a fixation point (FP), which appeared at the screen center (stimulus-aligned conditions) (Fig. 1C and D) or above or below center (saccade end-aligned conditions) (Fig. 1D and F) or to the right (Fig. 1E and F). The animals were required to fixate the FP for 0.5–0.7 s, after which the go irrelevant stimulus array appeared. The animals continued to fixate the FP for an additional 0.5–0.7 s, after which the FP stepped from center to one of the specified peripheral locations (array/item-aligned conditions) (Fig. 1C and E) or from the peripheral location to center (saccade end-aligned conditions) (Fig. 1D and F).

The animals were then required to launch a saccade to the new FP location and hold fixation for 0.5–0.7 s within a −3° × 3° computer-controlled window, after which a liquid reward was issued. Visually evoked responses were measured during the stimulus-aligned vs. saccade end-aligned time points, highlighted by the black outlines in key frames of Fig. 1 C–F.

ACKNOWLEDGMENTS

We thank Ann Lablans, Donald Brien, Sean Hickman, and Mike Lewis for technical assistance. This project was funded by Canadian Institutes of Health Research Grant FDN-148418 and National Science Foundation Grants CNS-0827764 and CCF-1317433. D.P.M. was supported by the Canada Research Chair Program.
Supporting Information

White et al. 10.1073/pnas.1701003114

SI Materials and Methods

Animal Preparation. Data were collected from four male Rhesus monkeys (Macaca mulatta) weighing between 10 and 12 kg: two for the SC recordings (monkeys I and U) and two for the V1 recordings (monkeys D and Y). For the SC recordings, surgical procedures and extracellular recording techniques have been detailed previously (45). For V1 recordings, one animal (monkey D) had a recording chamber implanted over V1, centered on the midline to allow access to both left and right lower visual field representations using single microelectrodes. In the second animal (monkey Y), a 96-channel microelectrode array (Blackrock Microsystems) was chronically implanted on the surface of the right V1 using surgical procedures outlined by Blackrock Microsystems (46). All animal care and experimental procedures were approved by the Queen’s University Animal Care Committee in accordance with the guidelines of the Canadian Council on Animal Care.

Stimuli and Equipment. Visual stimuli were presented on a high-definition LCD video monitor (Sony Bravia 55 inch; model KDL-46XBR6) at a screen resolution of 1,920 x 1,080 pixels (60-Hz noninterlaced, 16-bit color depth). Viewing distance was 70 cm, resulting in a viewing angle of 82° horizontally and 52° vertically. The viewing area that extended beyond the monitor was blackened using black nonreflective cloth.

The main stimuli consisted of a radial arrangement of equally spaced color bars (210 items), with the diameter of the entire display spanning 40° to 45° visual angle (Fig. 1 C and D). The items were horizontally or vertically oriented (typically 0.4° x 1.2° but modified slightly depending on RF size of V1 neurons) and were red or green derived from the red-green cardinal axis in Derrington-Krauskopf-Lennie (DKL) color space (47, with ~40% luminance contrast relative to the neutral gray background (65 cd/m²). The main condition consisted of a stimulus array with a single oddball that was always the opposite color and orientation as the remaining items [e.g., red horizontal against green vertical (depicted in Fig. 1 C and D), red vertical against green horizontal, green horizontal against red vertical, green vertical against red horizontal].

The oddball could appear IN or OPP the RF. This was compared with a single-item control condition, in which a single red or green, horizontal or vertical, stimulus appeared IN or OPP the RF (Fig. 1 E and F). All experimental conditions were interleaved.

The tasks were controlled by a Dell 8100 computer running a UNIX-based real-time data control system [real-time experimentation system (REX) 7.6] (48), which communicated with a second computer running in-house graphics software (written in C++) for presentation of stimuli. Stimulus timing was controlled using a photodiode placed at the left lower corner of the monitor and hidden by nonreflective tape. The photodiode measured the onset of a stimulus (20 x 20 pixels) that pulsed for one frame simultaneously with the onset of the main stimulus (i.e., the photodiode stimulus turned white for one frame and then returned to black). The REX was synchronized to the timing of the photodiode pulse by holding the current state until the pulse was detected.

Eye position was monitored using a 1,000-Hz video-based eye tracker (Eyelink 1000; SR Research). Saccades were detected based on a velocity (eye position > 50°/s) and amplitude (>1°) criteria and confirmed offline. Saccade end was defined as the point in time when the velocity of a detected saccade first fell below the saccade threshold defined above and was successfully inside the ~3° x 3° computer-controlled (imaginary) target window required for receiving a reward (only rewarded trials were analyzed). The data were recorded in a third computer running a multichannel data acquisition system (Plexon Inc.). Eye position, event data, and spike times were digitized at 1 kHz.

Procedure. Monkeys were seated with the head restrained in a primate chair (Cires Instruments) ~70 cm from the LCD video monitor. For SC recordings, single glass-insulated tungsten microelectrodes (2.0 MΩ; Alpha Omega) were lowered into the SC through a stainless steel guide tube. For V1 recordings in one animal (monkey D), we used more durable tungsten microelectrodes (250-μm shank diameter; 1.0 MΩ; FHC) that were able to pierce directly through the dura stabilized by a guide tube that was positioned against the dura. V1 recordings in the second animal (monkey Y) were obtained from a 96-channel chronically implanted microelectrode array (impedance range = 0.08–0.35 MΩ; Blackrock Microsystems). The animals viewed a dynamic video, which provided rich visual stimulation that facilitated the localization of the visually responsive dorsal SC surface or the uppermost layers in V1.

When a neuron was isolated, its visual RF was mapped using a rapid visual stimulation procedure described previously (21). Briefly, for SC neurons, single stimuli (white spot against a black background) were successively flashed (150-ms flashes with 150-ms interflash intervals) across the visual field, while the animal held fixation on a central fixation stimulus. For V1 neurons, with RFs that are significantly smaller than in SC (36, 37), the visual stimulation procedure was localized within a ~5° x 5° grid positioned over the approximate location of the neuron’s RF. The exact size and position of the stimulus grid could be manipulated online. The stimuli consisted of alternating black-white flashes against a neutral gray background (65 cd/m²). For neurons obtained from the V1 array, this stimulus grid was positioned in a manner that allowed us to coarsely map the spatial RFs across all electrode sites simultaneously. This online RF mapping was used to determine stimulus placement in the main experiment. Although it was not possible to determine the optimal stimulus parameters (orientation, size, direction of motion), the chosen stimulus parameters did evoke vigorous visual responses for most of the V1 neurons and multiunits that we sampled, and therefore, we assumed from this that our stimuli were effective for the purpose of this study. For SC neurons, a delayed saccade task was also used to characterize whether the neurons had visual and/or motor response properties using previously established methods (21). The location of the RFs for V1 neurons fell in the lower left hemisphere, a few degrees below the horizontal meridian, ±5° to 10° eccentric from the fovea. The RFs of SC neurons were generally left or right of the fovea and were above, below, or very near the horizontal meridian at eccentricities ranging from 6° to 24°, with the most frequent eccentricity at ~9° (mode = 8.8). The array was rotated and organized to place the oddball at the center or opposite the RF.

The animals then performed the main task. Specifically, on a given trial (Fig. 1), the animal fixated an FP (black Gaussian-windowed spot, SD = 0.3°), which appeared at the screen center (array-aligned or single item-aligned conditions) (Fig. 1 C and E) or above or below center (saccade end-aligned conditions) (Fig. 1 D and F) orthogonal to the RF at the same eccentricity (exactly ±90° polar angle from RF center). The animals were required to continue fixating on the FP for a 0.5- to 0.7-s random period, after which the goal-irrelevant stimulus/array appeared. The animals were required to continue fixating on the FP for an
additional 0.5–0.7 s, after which the FP stepped from center to one of the specified peripheral locations (array-aligned or item-aligned conditions) (Fig. 1 C and E) or from the peripheral location to center (saccade end-aligned conditions) (Fig. 1 D and F). The animals were then required to launch a saccade to the new location of the FP and to hold fixation on this new location for 0.5–0.7 s within a ~3° × 3° computer-controlled window, after which a liquid reward was issued. Thus, in the array/item-aligned case, the goal-irrelevant but salient stimulus (oddball or single item) abruptly appeared IN or OPP the RF, whereas in the saccade end-aligned case, the goal-irrelevant stimulus (oddball or single item) was brought into the RF via the saccade to screen center. Importantly, when the eyes were at center in either condition, the stimulus display was the same. Visually evoked responses were measured during the stimulus-aligned vs. saccade end-aligned time points, which are highlighted by the black outlines in the key frames of the illustration in Fig. 1 C–F.

**Data Analyses.**

**Single unit and multiunit recordings.** Single units were isolated online using a window discriminator and confirmed offline using spike-sorting software (Plexon Inc.). Spikes were convolved with a function that resembled an excitatory postsynaptic potential (49), with rise and decay values of 5 and 20 ms, respectively. A total of 94 neurons were isolated (38 V1, 31 SCs, 25 SCI). Using the microelectrode array in V1, several electrodes yielded reliable visually evoked multiunit spiking activity and showed qualitatively similar results to the V1 single units. Therefore, for the main analyses, we combined V1 single units (n = 38) and multiunit sites (n = 55) for a total of n = 93. Thus, although V1 was represented by a larger sample size, yielding potentially greater statistical power, this gave additional support for our hypothesis, because SCs neurons still showed a more robust and earlier saliency representation.

**SC neuron classification.** The SC is composed of two dominant functional layers (9, 10), a visual-only superficial layer (SCs) and a multisensory–cognitive–motor-related intermediate layer (SCI). SC neurons were functionally classified as visual SCs or visuomotor SCI based on their discharge characteristics using a visual RF mapping procedure (21) to determine the presence of a visual component and a delayed saccade task to determine the presence of a motor component using previously established methods (21, 31). Briefly, neurons were defined as having a visual component if the visual mapping procedure yielded the presence of a localized visual RF (21). Neurons were defined as having a motor component if the average firing rate around the time of the saccade (~25 to +25 ms relative to saccade onset) into the neuron’s preferred visual RF was significantly greater than a presaccade baseline period (~150 to ~50 ms relative to saccade onset).

**Data normalization.** The normalization procedure was similar to conventional normalization to the maximum firing rate, except that we scaled the firing rate of each neuron to its minimum and maximum evoked by the single item within the relevant time window (from stimulus onset for 500 ms). Specifically, for each neuron, neuronal discharge rates were normalized to the minimum and maximum values of the single-item condition (which always yielded the greatest response) using a zero to one rescaling of the spike density function with the following:

\[
\text{SDF normalized} = \frac{\text{SDF} - \text{min(single item SDF)}}{\text{max(single item SDF)} - \text{min(single item SDF)}}
\]

where SDF is the original spike density function (averaged across trials) for a given condition, and \(\text{min(single item SDF)}\) and \(\text{max(single item SDF)}\) are the minimum and maximum values, respectively, within the poststimulus period (0–500 ms) of the single-item condition. Thus, when normalizing the array condition, for example, its response can be seen as a proportion of the maximum response possible using a unitary item. This ensures that the relative differences between conditions for a given neuron are retained. This also makes the percentage surround suppression index very intuitive and easy to compute, because the response curves in the array condition will always be some proportion of the maximum response elicited by the single item.

**Latency and magnitude indices.** VROL, SROL, and SSOL were computed for each neuron when possible (i.e., when the neuron exhibited each of these properties). We used an established statistical method used by others in the attention selection literature (10, 50, 51), which involves a running Wilcoxon rank sum test at a fixed alpha level (P < 0.05). This statistical method has been shown to be comparable with conventional receiver operating characteristic analysis (51) for determining neuronal selection times and therefore, provides a reasonable estimate of the latencies and selection times comparable with other studies. The test was run on a temporally averaged moving window at every millisecond from 0 to 300 ms. Specifically, VROL was defined as the time in which activation (averaged over a moving 5-ms window) was significantly greater than a prestimulus baseline (single item-aligned condition: −100 ms to stimulus onset; saccade end-aligned condition: −200 to −100 ms relative saccade onset). SROL was defined as the time in which activation (averaged over a moving 50-ms window) was significantly greater for oddball IN vs. oddball OPP condition. We used a longer moving window here compared with VROL, because the dynamics of this process was slower and more variable, and the longer window yielded more reliable results. SSOL was defined similarly to SROL, but instead of comparing oddball IN with OPP conditions, the single-item IN condition (no surround) was compared with the oddball IN condition (surround). The latency indices were verified by observing the estimated VROL, SROL, and SSOL of each neuron plotted against its spike density function.

Because surround suppression emerged at noticeably different time periods across the brain areas (Fig. 3), we computed surround suppression magnitude indices during the peak response and later, during a sustained portion of the response. The surround suppression index was defined as the ratio of the response evoked by the single item relative to the response evoked by the array. The time of the peak response differed for each neuron and brain area but in general, ranged from ~40 to 100 ms post-stimulus onset. The sustained period was defined as the period from the VROL of a given neuron for 300 ms.
**Fig. S1.** The time course of saliency representations in V1 and SC (absolute firing rates). (A–D) Population average responses for V1 (blue; n = 93) and SCs (red; n = 31) across conditions. These results were derived from the same data as in Fig. 2, except here, they are nonnormalized absolute firing rates. The tick marks immediately above the x axis in A–D indicate the points where the oddball IN response was significantly greater than the oddball OPP response (P < 0.05, Wilcoxon signed rank test at 10-ms intervals, Bonferroni–Holm corrected). (E and F) Cumulative distributions of VROL (solid traces) and SROL (dotted traces), with the relative median latencies indicated by the vertical lines with arrows. *P < 0.05, rank sum test.

**Fig. S2.** Feature preferences. Averaged population spike density functions for preferred and nonpreferred stimulus features (orientation, color) for V1 (A and B) and SCs (C and D).
Fig. S3. The time course of saliency representations in V1 (units only) and SC. (A–D) Normalized population average responses for V1 (blue; $n = 38$; monkeys D and Y) and SCs (red; $n = 31$; monkeys I and U) when the salient oddball appeared abruptly in the RF (A, C, and E; array-aligned condition) or was brought there via a saccade (B, D, and F; saccade end-aligned condition). The tick marks immediately above the x axis in A–D indicate the points where the oddball IN response was significantly greater than the oddball OPP response ($P < 0.05$, Wilcoxon signed rank test at 10-ms intervals, Bonferroni–Holm corrected). (E and F) Cumulative distributions of VROL (solid traces) and SROL (dotted traces), with the relative median latencies indicated by the vertical lines with arrows. ns, Not significant. *$P < 0.05$, rank sum test.

Fig. S4. The time course of saliency representations in V1 and SC (overlapping vs. nonoverlapping RFs). Comparison of VROL and SROL when RFs were overlapping or nonoverlapping between brain areas. As described in Materials and Methods, V1 RFs were clustered in the lower left visual field at 5° to 10° eccentricity from the fovea, whereas SCs RFs were sampled widely across the visual field, with little direct RF overlap as our V1 sample. To examine potential differences between overlapping and nonoverlapping RFs between brain areas, we compared our V1 sample with a subset of SCs neurons with RFs that were within the same eccentricity and lower visual field as V1 (overlapping condition) and a subset of SCs neurons with RFs that were outside this range (nonoverlapping). (A–D) Cumulative distributions of VROLs (solid traces) and SROLs (dotted traces) for overlapping (A and C) and nonoverlapping (B and D) RFs. Median latencies indicated by the vertical lines with arrows. ns, Not significant. *$P < 0.05$, rank sum test.
Fig. S5. Surround suppression in V1 and SC (absolute firing rates). (A–D) Population average responses for V1 (blue; n = 93) and SCs (red; n = 31) across conditions. These results were derived from the same data as in Fig. 3, except here, they are nonnormalized absolute firing rates. The tick marks immediately above the x axis in A–D indicate the points where the single-item IN response was significantly greater than the oddball IN response (P < 0.05, Wilcoxon signed rank test at 10-ms intervals, Bonferroni–Holm corrected). (E–H) Averaged percentage suppression across V1 (n = 93) and SCs (n = 31) recordings. (I and J) Cumulative distributions showing a comparison of the SSOL (dotted traces) relative to VROL (solid traces). ns, Not significant. *P < 0.05, rank sum test.

Fig. S6. Surround suppression in V1 (units only) and SC. (A–D) Normalized population average responses for V1 (blue; n = 38; monkeys D and Y) and SCs (red; n = 31; monkeys I and U) comparing the single-item IN condition (black traces; no surround) with the oddball IN condition (colored traces; surround) aligned on array/item onset (A and C) and saccade end (B and D). The tick marks immediately above the x axis in A–D indicate the points where the single-item IN response was significantly greater than the oddball IN response (P < 0.05, Wilcoxon signed rank test at 10-ms intervals, Bonferroni–Holm corrected). (E–H) Averaged percentage suppression across V1 (n = 38) and SCs (n = 31) recordings. Percentage suppression was computed from the ratio of response evoked by the single item (no surround) vs. the array (surround) with the oddball IN the RF and was computed during the peak and during a sustained period of the response (Materials and Methods). (I and J) Cumulative distributions showing a comparison of the SSOL (dotted traces) relative to VROL (solid traces). ns, Not significant. *P < 0.05, rank sum test.
Fig. 57. Surround suppression in V1 and SC (overlapping vs. nonoverlapping RFs). This figure shows the surround suppression results with overlapping vs. nonoverlapping RFs between brain areas (as detailed in Fig. S4). (A–D) Cumulative distributions of VROLs (solid traces) and SSOLs (dotted traces) for overlapping (A and C) and nonoverlapping (B and D) RFs. Median latencies are indicated by the vertical lines with arrows. (E–L) Mean percentage suppression for overlapping (E–H) and nonoverlapping (I–L) RFs between brain areas. ns, Not significant. *P < 0.05, rank sum test.